

SpAM

SpAM (Spatial Analysis and Methods) presents short articles on the use of spatial statistical techniques for housing or urban development research. Through this department of Cityscape, the Office of Policy Development and Research introduces readers to the use of emerging spatial data analysis methods or techniques for measuring geographic relationships in research data. Researchers increasingly use these new techniques to enhance their understanding of urban patterns but often do not have access to short demonstration articles for applied guidance. If you have an idea for an article of no more than 3,000 words presenting an applied spatial data analysis method or technique, please send a one-paragraph abstract to rwilson@umbc.edu for review.

Applying Spaghetti and Meatballs to Proximity Analysis

Alexander Din

U.S. Department of Housing and Urban Development

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development & Research, the U.S. Department of Housing and Urban Development, or the U.S. Government.

Abstract

The spaghetti and meatballs technique is a geoprocessing method used in a Geographic Information System (GIS) that counts the number of overlapping polygons that are of unequal size and shape. Often, this method is used to calculate densities of coverage areas including, but not limited to, the extent of an oil spill over a period of time or the extent of a burn during a wildfire, or to compare perceptions of a region. In this demonstration, I use the spaghetti and meatballs technique to measure the density of proximity to points of interest, or amenities, in Washington, DC. I calculate summary statistics to describe the densities of amenities by the District's eight city council wards.

Background

The spaghetti and meatballs technique counts overlapping polygons of varying shapes and sizes in a geographic information system (GIS). The spaghetti and meatballs technique appears in many academic papers and various blog posts (Honeycutt, 2012), but researchers have not described it with sufficient explanation as to how it works when being applied to additional geoprocessing tasks. This article demonstrates how to create overlapping buffer polygons of unequal shape and size, how to apply the spaghetti and meatballs technique to identify densities of overlapping polygons, and finally, how to summarize the results by geography. The inconsistency of the shape and size of the polygons is important to simulate because, in many situations, the extent of the shapes being analyzed will vary greatly, such as wildfire extents, flood extents, oil spills, soil composition, or other phenomena that do not fit neatly into clearly defined areas such as census tracts, hexagon tessellations, or other forms of grids. If all areas used the same grid, there would be no need to perform the spaghetti and meatballs technique.

The scenario behind this demonstration is to identify areas in Washington, D.C., that are proximate to amenities. For this demonstration, I chose amenities as points of interest from the Washington, D.C., Open Data page.¹ By creating buffers from these points, which vary in number and location, polygons of unequal shape and size can be generated. The method to generate the buffers is specific to this demonstration to create polygons of unequal shape and size, but the need to count overlapping polygons is not limited to this example. Measuring the extent of wildfires, oil spills, or floods over periods of days are common examples of why an analyst may need to perform this task. Site selection is another reason why an analyst may perform this task. For example, guidance may advise that assisted or supportive housing be located within a certain distance from amenities; the spaghetti and meatballs method could be used to identify suitable areas by the density of overlap areas that meet the proximity criteria.

To create polygons of varying shapes and sizes, I performed a distance analysis to find the 25th percentile of distance from points of interest to a dense grid of points used to simulate any location within the District of Columbia. Once I identified the 25th percentile of each point of interest, I created buffers using that distance with each associated point of interest. The process of creating these buffered locations, and then counting the overlaps, is similar to counting the overlaps of an informal region (Bowlick et al., 2015). This article demonstrates how to apply the spaghetti and meatballs technique to a proximity analysis of points of interest in Washington, DC. This demonstration primarily uses ArcGIS for its graphical user interface (GUI) to help make the process clearer, but it can be performed with any GIS software, such as QGIS or another open-source platform.

The end of this demonstration includes how to create summary statistics describing the land area. Although I used the political boundaries of the nation's capital for this analysis, I used only the land area to create the statistics because people are typically interested in site selection of a land area. The reason I used the entire political boundary, however, including creating buffers that may go outside of the city limits, was to perform geoprocessing on the entire dataset, then reduce later where I deemed appropriate.

¹ <http://opendata.dc.gov/>

Using Points of Interest to Create Overlapping Polygons

To create a set of overlapping polygons of varying shapes and sizes, I chose 12 sets of points of interest (PoI) in Washington, DC, from the Open DC GIS Catalog. Each data set contained points representing location data, such as entrances to Metro stations, Capital Bike Share stations, or other places of interest. The reason for choosing 12 different location data sets is that they have different spatial distributions across the District. The difference in the spatial distributions indicates that the data sets will likely cover different regions of the city to varying degrees. Performing a Near analysis, and creating buffers based on the results of the Near analysis, will enable creating overlapping polygons of varying degrees and sizes across the city that will be unique and complex shapes. The goal here is to use a set of gridded points, representing all potential locations within the city, to determine locations relative to the PoI, allowing for adjustments to the potential size of the buffers. All datasets were projected into the Maryland State Plane Coordinate System² for this analysis to orient the geography to preserve shape and area to be consistent with data standards for the District of Columbia.³

To create the buffers, which are polygons of varying shapes and sizes, the first step was to create a fishnet grid that covered the District to measure distance at a more local (granular) level to assess any distance within the geography to a PoI. I created a fishnet grid using the extents of the District with 1,000 rows and 1,000 columns. I chose this number of rows and columns for the grid because it thoroughly covered the city, but other parameters may also be chosen to modify the distance at which the points are located from each other. Because the extents of Washington, DC, go further than much of the city's actual area, I removed points outside the city boundary using the Clip tool to keep the analysis within the city limits. The result was that 428,281 points (42.8 percent) remained of the original 1,000,000 points in the fishnet grid. Using the 1,000 by 1,000 points dimensions, the points were spaced 59.9 feet apart from each other to create a layer that measures local proximity in a systematic way across the city.⁴ This layer allows an analyst to measure the proximity of any PoI to any other location within the coverage area.

I used the Near tool to measure the distance from each PoI to the grid points, appending the distance to the PoI layer. For each PoI, I calculated the minimum, 25th percentile, median, 75th percentile, maximum, and average distance from the grid network to each PoI. Exhibit 1 presents a set of distance statistics to assess the proximity in the city to one of the PoI types. Each percentile distance indicates that within that number of feet from the PoI is a percentage of the area within the political boundary of Washington, DC. For example, approximately 25 percent of the area of the city is within 856.5 feet of a Capital Bikeshare station, and one-half of the city is within 1,658 feet of a station, roughly double the distance that 25 percent of the district is within. I repeated this process for each set of amenity locations to create a table of distance statistics for each PoI.

² The District of Columbia uses the Maryland State Plane Coordinate System <https://octo.dc.gov/page/coordinate-system-standards>

³ Maps in this paper are shown in North American Albers Equal Area Conic consistent with other work.

⁴ A map of the fishnet grid is not shown because at a Districtwide scale, the points cover the entire geography and are not discernible.

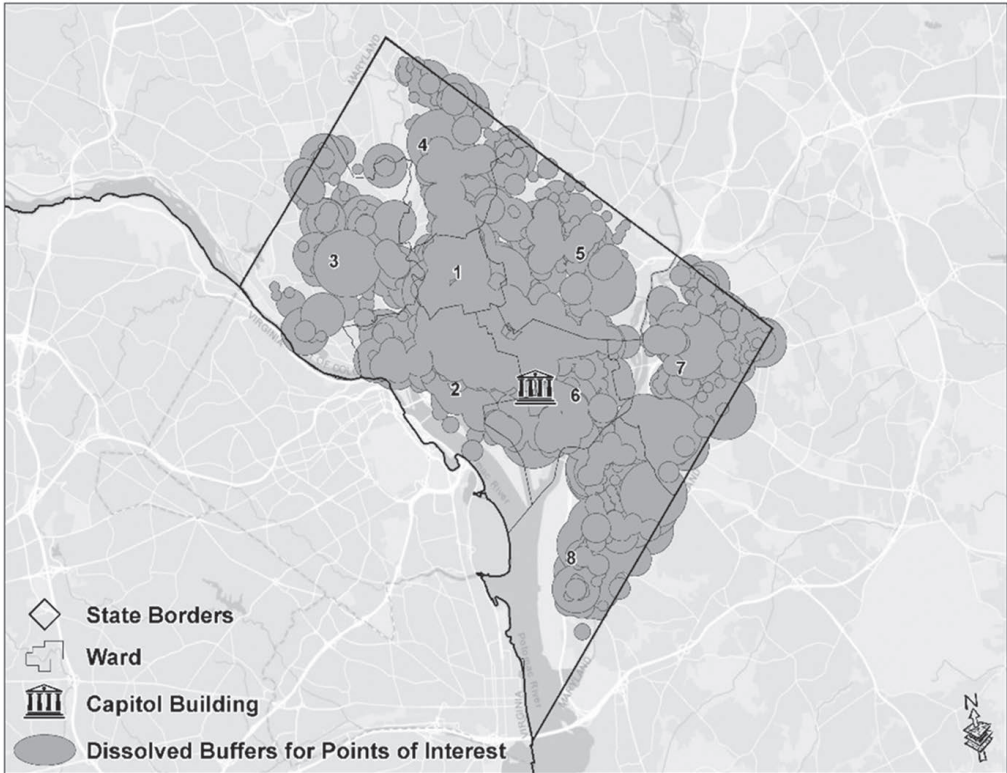
Exhibit 1

Distance from Fishnet Grid to Nearest Point of Interest (Feet)						
Points of Interest	Minimum	25th Percentile	Median	75th Percentile	Maximum	Average
Books and Barber Shops	6.5	3,316.1	6,074.4	10,331.8	27,103.5	7,397.7
Capital Bikeshare	1.0	856.5	1,658.0	3,031.7	12,867.3	2,269.0
Elementary Schools	5.6	1,514.9	2,476.6	3,922.8	15,207.6	2,973.4
Farmers Markets	6.3	1,745.9	2,974.2	5,091.3	16,695.6	3,673.4
Libraries	4.1	2,483.7	3,796.9	5,256.5	16,698.7	4,026.1
Metro Station Entrances	5.6	2,169.5	3,977.9	6,686.1	24,184.2	4,897.4
Museums	6.2	2,294.1	4,317.5	7,550.8	29,284.8	5,689.6
Places of Worship	2.6	554.3	1,131.1	2,286.7	15,754.2	1,802.2
Police Stations	16.5	3,488.3	5,555.4	8,221.6	29,906.8	6,490.4
Public Art	1.0	1,091.6	2,056.1	3,504.4	16,695.6	2,615.5
Taverns	2.1	1,517.2	3,464.7	6,369.6	23,560.5	4,411.1
Urban Agriculture Sites	2.7	2,553.5	4,384.9	7,371.9	19,480.3	5,568.6

I then created buffers based on each PoI's distance at the 25th percentile to represent activity space around that point. For each buffer created, the Dissolve Type parameter must be set to all to create single, continuous polygons where possible. This parameter eliminates overlapping polygons for the same PoI. For each dissolved buffer layer created for each PoI, I exploded the buffers into multipart features to ensure that each feature is a separate record. I appended the resulting buffers for all 12 PoIs into a single map layer. All buffers become individual features within a single layer, rather than many features—including potential multipart features—in 12 layers. Ensuring that all features are in a single layer is important for the spaghetti and meatballs portion of the analysis. The resulting layer contains a series of overlapping buffers for all PoI, as shown in exhibit 2.

Exhibit 2

Buffered Polygons of All Points of Interest



Spaghetti and Meatballs

After the overlapping buffers are created, the next step is to transform the polygons into spaghetti shapes using the “Feature-to-Polygon” tool, which divides the buffer polygons into smaller component polygons from each overlapping layer at the intersections of each overlapping buffer. That is, the tool creates buffer pieces that represent the intersecting areas of the overlapping buffers. In this example, the number of features grows from 359 in the layer of buffered polygons to 5,405 in the layer generated by the Feature to Polygon tool. The resulting spaghetti of polygons is shown in exhibit 3.

Exhibit 3

Spaghetti in and Around Ward Six



Using the spaghetti polygons, points (also called meatballs in this method) are made that represent the center of each polygon, that is, the centroid. These meatball center points are generated by converting each feature in the spaghetti layer to a point using the Feature-to-Point tool that identifies a polygon's center and creates a point at that location. When using this tool, it is important to define the center by using the Inside parameter. Because many of the shapes in the spaghetti layer are complex shapes, it is possible for the central point to be located outside the bounds of the corresponding polygon. Using the Inside parameter will ensure that the centroid point will remain within the shape of the polygon. A meatball is created for each spaghetti section, creating more meatballs than POIs. The meatballs are the centroids for each section of overlapping POI buffers. See exhibit 4.

Exhibit 4**Meatballs in and Around Ward Six**

The number of overlapping polygons can be counted using the meatball points and the original set of polygons. To do this, I used the “Spatial Join” tool⁵ to count the number of original polygons that intersect with each meatball, or central point, representing one section of each original polygon broken down into its smallest part. The Spatial Join is a tool available in many GIS software applications that computes the intersection between two layers. This computation can be a count of one layer intersecting with a second layer, like in this demonstration, or it can perform a more sophisticated calculation like that of an attribute of the first layer relative to the second layer.⁶ In this process, the Spatial Join tool counts the number of overlaps that occur between the two layers, thus providing a frequency of the overlapping intersections. In this example, the sections of overlapping polygons range from 1 (indicating no overlap) to 12 (indicating the most overlap), with the buffers from all 12 PoIs overlapping in a given area. When configuring the Spatial Join, the “Target Feature” (input) is the meatballs, and the “Join Feature” is the original set of polygons before the spaghetti was created. The Spatial Join joins the number of original polygons to each point representing an overlap section. The original polygons are filled with the number of overlap parts, represented as points (meatballs). Exhibits 5 through 7 demonstrate the output of the Spatial Join tool.

⁵ In ArcPro, the Summarize Within tool simplifies this process. Additionally, using Python and a combination of the Definition Query, Select Features by Location, SearchCursor, and UpdateCursor functions could be used instead in ArcGIS or ArcPro. Other alternatives are also available in open source and other platforms.

⁶ An example of such a computation would be to use a Spatial Join between a point layer of apartment buildings and a polygon of census tracts to calculate the number of units in the apartment layer by census tract.

Exhibit 5

Spatial Join Tool Parameters

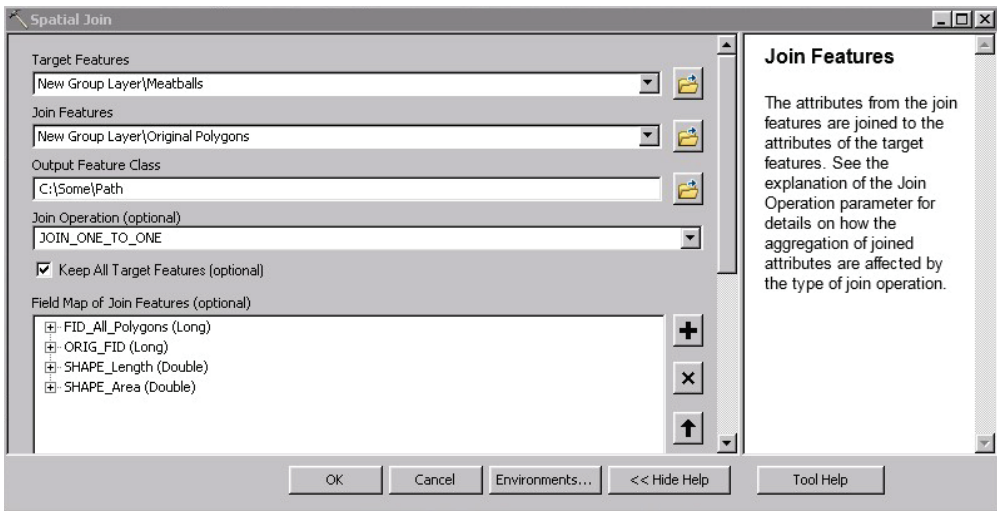


Exhibit 6

Point Layer Resulting from Spatial Join Overlaid on Original Polygons Layer in and Around Ward Six



Exhibit 7

Point Layer Resulting from Spatial Join Attribute Table

OBJECTID *	SHAPE *	Join_Count	TARGET_FID	FID_All_Polygons	ORIG_FID
1	Point	1	1	-1	1
2	Point	1	2	-1	2
3	Point	2	3	-1	3
4	Point	2	4	-1	4
5	Point	3	5	-1	5
6	Point	1	6	-1	6
7	Point	2	7	-1	7
8	Point	4	8	-1	8
9	Point	2	9	-1	9
10	Point	3	10	-1	10
11	Point	3	11	-1	11
12	Point	1	12	-1	12
13	Point	3	13	-1	13
14	Point	4	14	-1	14
15	Point	3	15	-1	15
16	Point	1	16	-1	16
17	Point	4	17	-1	17
18	Point	6	18	-1	18
19	Point	5	19	-1	19
20	Point	1	20	-1	20
21	Point	6	21	-1	21
22	Point	4	22	-1	22
23	Point	6	23	-1	23
24	Point	5	24	-1	24
25	Point	6	25	-1	25
26	Point	3	26	-1	26
27	Point	7	27	-1	27
28	Point	5	28	-1	28
29	Point	4	29	-1	29
30	Point	4	30	-1	30
31	Point	5	31	-1	31
32	Point	1	32	-1	32
33	Point	6	33	-1	33

The point layer resulting from the spatial join operation contains the number of overlaps in the Join_Count field, which is the result of the intersection of each buffer layer with a centroid. The ORIG_FID is the ID of the meatball centroids that are created from the spaghetti polygons. This field links the meatball points back to the spaghetti polygons for the purpose of visualizing the number of overlaps, described in the Join_Count field, in the spaghetti polygons. In the spaghetti

polygons, a numeric Frequency field is created that contains the number of points (meatballs) that represent the number of overlapping polygons. When the point layer from the spatial join is overlaid with the spaghetti layer, the OBJECT_ID from the point layer is transferred into the ORIG_FID field to link them if needed. These links enable rejoining the two layers by a common ID so that any additional data from each of the layers can be used together.

Once the values from the Join_Count field in the spatial join layer have been joined to the Frequency field in the spaghetti layer, there are records in the spaghetti layer that have a value of zero. Those records are the result of areas that were not within a polygon in the original polygon layer but were continuously bounded on all sides by polygons. Those records are deleted from the spaghetti layer, dropping the number of features from 5,405 to 5,392. The results are thematically mapped in the exhibits 8 and 9.

Exhibit 8

Overlapping Buffers of Points of Interest

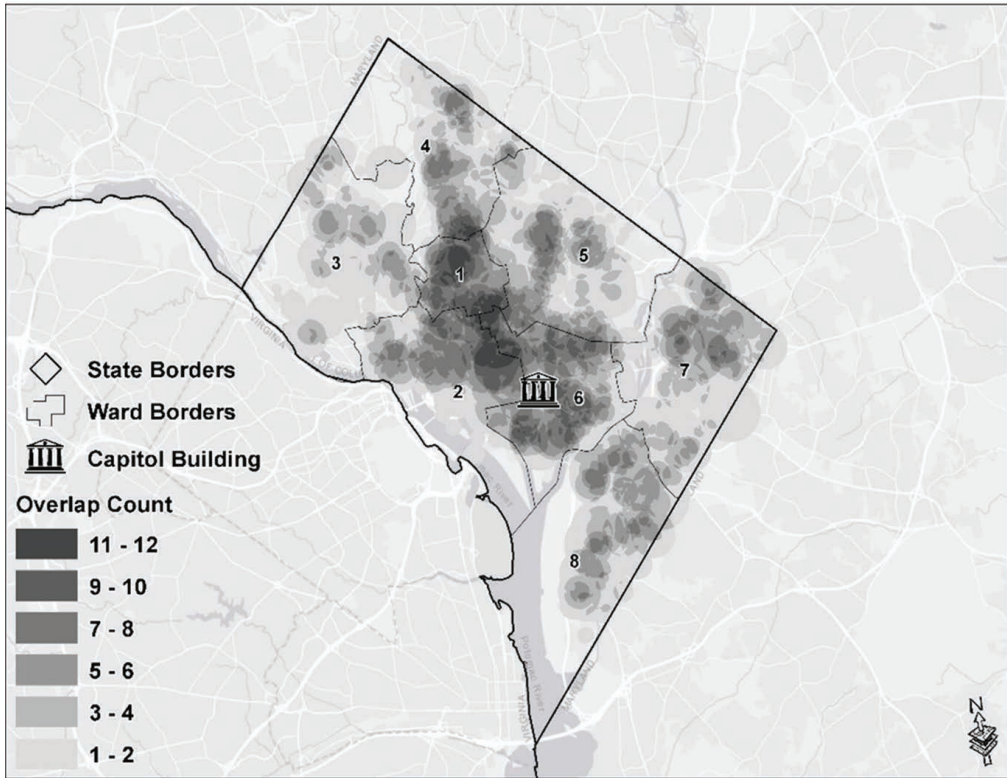
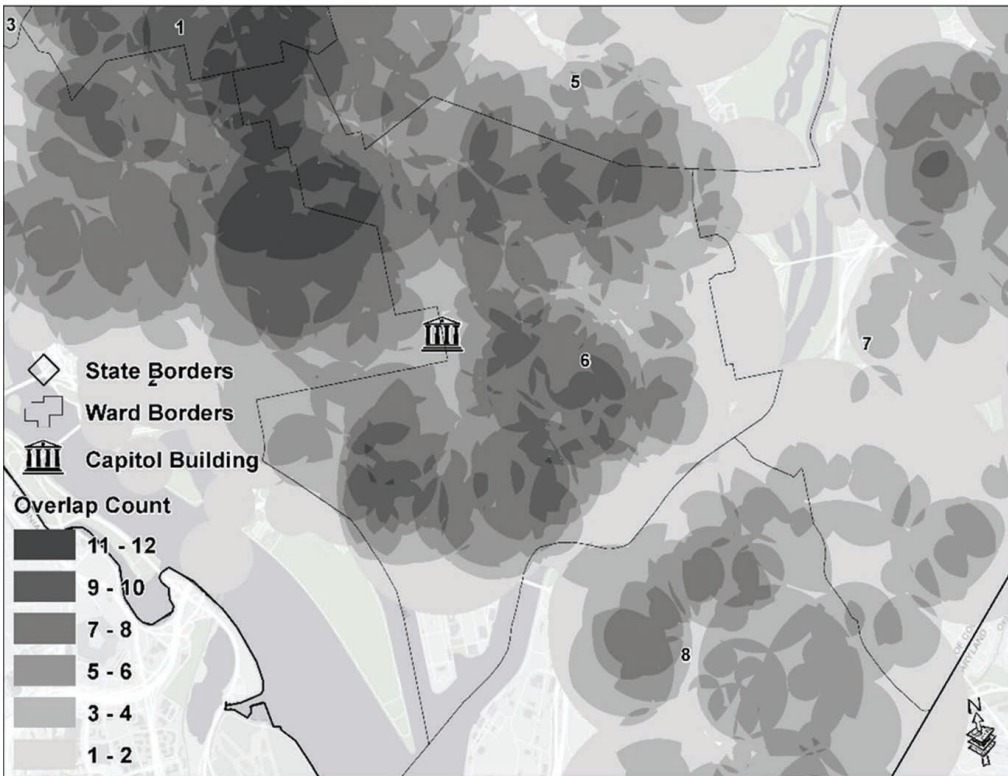


Exhibit 9**Overlapping Buffers of Points of Interest Around Ward Six****At a Glance—Visual Analysis and Histogram**

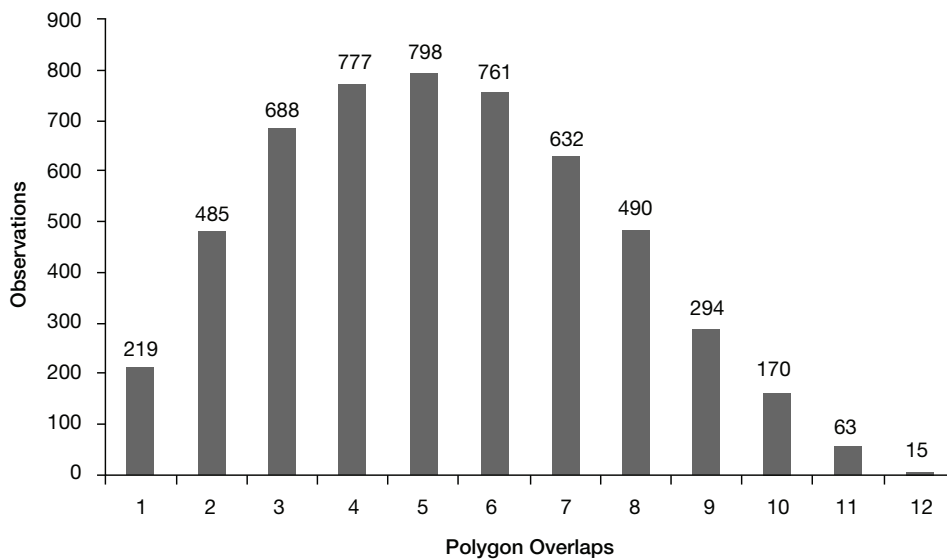
On visual examination of exhibit 8, Ward 1 appears to have the highest level densities of overlaps as a share of total area than any other ward. Ward 1 contains large portions of Howard University and Shaw in the southern area, Adams Morgan and Mt. Pleasant in the west, and Columbia Heights and Park View in the central area. These neighborhoods have all experienced a great deal of change over the past two decades, and, unsurprisingly, they are closer to more points of interest than other neighborhoods. Ward 2 also appears to have high densities of greater overlaps, continuing from Ward 1. Ward 6 appears to have a large concentration of moderate-to-high overlaps spread across the ward with few areas that have low numbers of overlaps—indicating that there are not many areas proximate to a low number of amenities in this ward.

To the east of Ward 6 and Ward 5 is a gap of overlapping polygons before reaching Wards 7 and 8. This gap is likely the Anacostia River because no PoIs were in the river and few were likely to be particularly close to the river. These wards have historically been isolated from the rest of Washington, DC, and contain a large portion of the city's poorest and most socially vulnerable populations. Even as the District has made progress on many metrics, the East-of-the-River sections of the city lag behind other area of the city.

The histogram in exhibit 10 shows the number of observations for each type of resulting polygons. The histogram appears to have a relatively normal distribution with a slight right-skew. Most polygons are in areas of five or fewer overlaps (54.9 percent). This finding is in keeping with the map in exhibit 6 and later maps that show relatively few areas of densities of proximity to greater numbers of sets of points of interest. The histogram does not show areas within the District that are not proximate to any of the 12 sets of PoIs. The histogram also does not demonstrate the size of each area. For example, more areas may overlap being proximate to three sets of PoIs than to two sets of PoIs, but the polygons that were made from only two overlaps comprise a larger geographic area. Lastly, neither the histogram nor the maps indicate to which PoIs the overlapping polygons are proximate.

Exhibit 10

Histogram of Polygon Overlaps



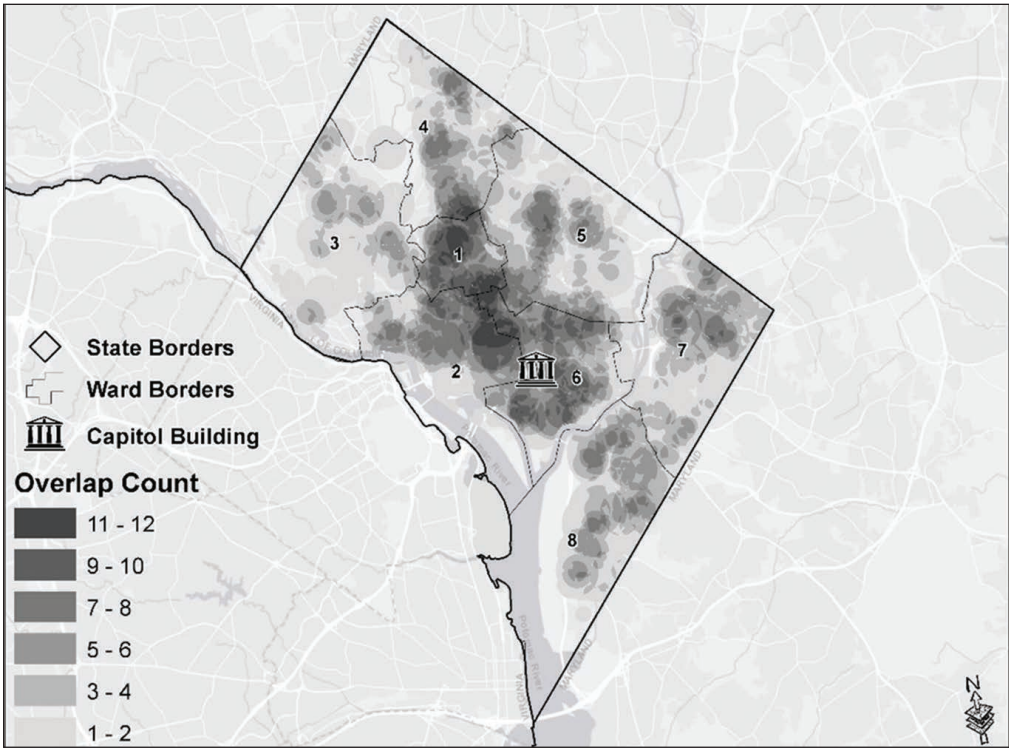
Intersection with Wards

Because this section is to provide summary statistics of the area within each number of overlaps, it is important to provide this information as *land* area rather than total area within the political boundaries of Washington, DC. To remove areas covered by water within the boundaries of DC, I erased areas that were defined as Freshwater Pond, Lake, or Riverine by the National Wetlands Inventory⁷ from the political boundary layer. Next, I clipped the polygons with overlap information by political boundary of the District of Columbia to remove areas outside of the District. The result is shown in exhibit 11. This action primarily removed areas like the Anacostia and Potomac rivers, but it also took out areas like Lake McMillan in Ward 1 and created islands such as Roosevelt Island in Ward 2.

⁷ <https://www.fws.gov/wetlands/data/State-Downloads.html>

Exhibit 11

Overlapping Polygons Clipped to Washington, D.C., Political and Land Boundaries



To accurately calculate the percentage of each Ward that had between 0 and 12 overlaps, the first step was to dissolve the spaghetti polygons by the overlap count. Next, I integrated the dissolved spaghetti polygons into the clipped Ward layer with the “Union” tool so that it contained only the land area within each ward. This step produced a layer of 131 polygons that described which Ward the polygon was a part of and the number of overlaps, or density, of that particular polygon. I created a new field to calculate the number of acres per polygon. A crosstab of areal percentages for each overlap count by Ward is shown in exhibit 12.

Exhibit 12

Percentage of Area by Ward by Overlap Count

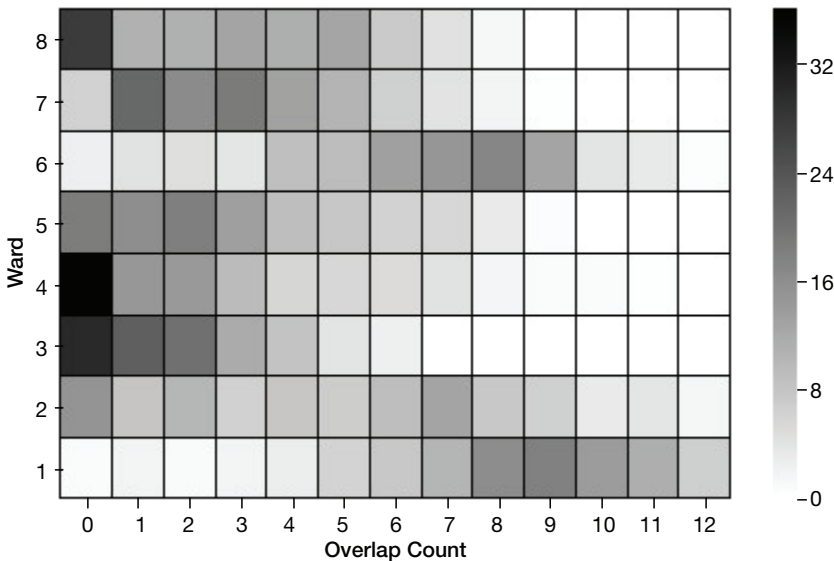
Ward	Overlap 1	Overlap 2	Overlap 3	Overlap 4	Overlap 5	Overlap 6	Overlap 7	Overlap 8	Overlap 9	Overlap 10	Overlap 11	Overlap 12
Ward 1	0.9%	2.0%	1.1%	2.1%	2.8%	6.2%	10.5%	16.3%	18.0%	14.0%	11.5%	6.9%
Ward 2	15.3%	8.1%	10.4%	6.5%	7.9%	7.1%	12.7%	7.7%	6.7%	3.1%	3.9%	1.6%
Ward 3	30.1%	22.6%	20.3%	12.0%	8.4%	4.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
Ward 4	36.0%	14.7%	14.5%	9.4%	5.8%	5.7%	4.3%	1.8%	0.9%	0.9%	0.6%	0.2%
Ward 5	18.5%	16.1%	18.3%	13.6%	9.1%	7.8%	5.7%	3.3%	0.8%	0.3%	0.0%	0.0%
Ward 6	2.6%	4.3%	4.7%	3.9%	9.0%	9.4%	14.8%	17.1%	12.8%	4.0%	3.4%	0.6%
Ward 7	6.4%	21.3%	16.4%	18.8%	13.1%	10.6%	4.2%	2.0%	0.5%	0.0%	0.0%	0.0%
Ward 8	27.6%	11.1%	11.2%	12.7%	11.7%	12.6%	4.4%	0.0%	0.0%	0.0%	0.0%	0.0%

The output table in exhibit 12 also shows the percentage of area in the overlap count, or level of density of proximity to points of interest, by Ward. This tabular representation aligns with the map in exhibit 9. Reading this table, it is easier to make comparisons of overlap counts between Wards or comparisons of Wards between different overlap counts. The table also shows that Wards 4 and 5 have higher portions of land area that are within areas of higher densities of proximity to PoIs. For example, the table shows that Wards 3, 4, and 8 range between 27.6 percent and 36.0 percent with no overlapping buffers of PoIs. The reader can also see that many Wards have values of 0.0 for higher overlap counts, indicating that these Wards have no areas or a limited number of areas with high proximity to points of interest.

It may be difficult to read the table because it has 96 values. To overcome this difficulty, I created a heatmap to better visualize the data (exhibit 13). The heatmap is a shaded matrix and, in this example, visualizes the values by using darker colors for greater values (Din, 2019). In the visualization, the table comes to life. It becomes clear that many Wards have few, if any, areas that have areas of high density of proximity to PoIs. The visualization enables the reader to better understand how much of the land area is within each overlap count and the density of overlaps.

Exhibit 13

Heatmap of Percentage of Area by Ward by Overlap Count



Conclusion

The spaghetti and meatballs technique is a way to count the number of overlapping polygons in an area. This method can be used for proximity analysis, site selection, cost-surface modeling, or other applications yet to be considered. Ultimately, because this data processing used a wide range of datasets for points of interest, what this particular analysis describes is that the central portion of Washington, DC, has the highest density of being most proximate to selected amenities. The

real power and versatility of the analysis are derived from how the polygons are created to generate meaningful graphic representations that help to identify something of value. The fishnet grid to create the nth percentile distances in a given area is not required, and the buffers can be created from a predetermined distance that is relevant to the analyst. Rather than being close to a PoI such as within the 25th percentile, the area to form the polygon could be the distance from the PoIs to the 75th percentile or greater. Static, fixed PoIs are not always required; PoIs that are in different locations may be used. There are numerous possibilities to create areas of varying shapes and sizes that the spaghetti and meatballs method can be applied to for proximity and density analysis.

Author

Alexander Din is a social science analyst in the Office of Policy Development & Research at the U.S. Department of Housing and Urban Development.

Notes

While this article was in review, Esri released a tool to count overlapping polygons in their release of ArcGIS 2.5. As of the submission of this paper, the author has not yet had the opportunity to make the leap from ArcGIS 2.4 to test the new tool.

Appendix A: Points of Interest

<http://opendata.dc.gov/datasets/capital-bike-share-locations>

<http://opendata.dc.gov/datasets/metro-station-entrances-in-dc>

<http://opendata.dc.gov/datasets/museums-in-dc>

<http://opendata.dc.gov/datasets/liquor-licenses>

<http://opendata.dc.gov/datasets/police-stations>

<http://opendata.dc.gov/datasets/public-schools>

<http://opendata.dc.gov/datasets/libraries>

<http://opendata.dc.gov/datasets/places-of-worship>

<http://opendata.dc.gov/datasets/books-and-barber-shops>

<http://opendata.dc.gov/datasets/washington-dc-public-art>

[Page removed from the Open Data DC website at the time of publication]

<https://opendata.dc.gov/datasets/urban-agriculture-sites-points>

<http://opendata.dc.gov/datasets/farmers-market-locations>

References

Bowlick, Forrest, Koehler Andermatt, Cathy Besmar, and Josh Erbe. 2015. "Defining the Palouse: Using Overlay Analysis to Delineate an Informal Region." *Yearbook of the Association of Pacific Coast Geographers*, 77: 40–51.

Din, Alexander. 2019. "Using Heatmaps to Explore Capital Bikeshare Data," *Cityscape* 21(1): 239–241.

Honeycutt, Dale. 2012. "More Adventures in Overlay: Counting Overlapping Polygons with Spaghetti and Meatballs." ArcGIS Blog. <https://www.esri.com/arcgis-blog/products/arcgis-desktop/analytics/more-adventures-in-overlay-counting-overlapping-polygons-with-spaghetti-and-meatballs/>