Spatial Methods for Heritage Preservation

A comparative report

CPLN 680: Advanced Topics in GIS | Spring 2021 | Ilil Feiglin

Spatial Methods for Heritage Preservation | Ilil Feiglin

CPLN 680: Advanced Topics in GIS | Professor Dana Tomlin

Weitzman School of Design | University of Pennsylvania

TABLE OF CONTENTS

Abstract

This study proposes preservation planners and advocates a spatial toolkit. The toolkit is based on several possible planning or inventory needs. The six methods are divided into three main groups, which are introduced at the beginning of the report. The methods are clearly explained and detailed so that beginner to intermediate level staff can repeat them to produce outputs relevant to their needs.

The toolkit aims to compare possible methods in terms of suitability to different needs while introducing their strengths and limitations. However, another goal that might surpass the first is to advocate and promote the usage of spatial analysis and monitoring within the preservation field. This study and toolkit promote innovative spatial uses of data, hoping that it will lead to new and innovative use cases that weren't anticipated within the study.
The toolkit is built from six main methods. It concludes with a table that summarizes and compares all methods and a comparison that completes the effort. The report refers to the glossary at its end for specific terms, or additional explanation is needed.

The study ranges from descriptive to prescriptive, without one specific usage that is its only question. However, in creating the toolkit, it became apparent that the most fundamental part is the user's intent. In other words, to choose the 'right' tool, planners should refine their questions and have a elementary understanding of available datasets. In this sense, the toolkit is similar to the Cheshire cat from Alice in Wonderland; it can only assist if you have some grasp of where you want to go.

Introduction

Historic listings in the US and around the world change (usually accumulated) over place and time. For example, the NPS publicly available dataset allows us to extract temporal and spatial data of listings throughout the years and across different states. These listings are affected (or correlated) with other parameters and affect other attributes in their environment.

While various indicators to the direct and indirect effect of historic listings are suggested within the literature, the spatial pattern of these listings, as well as its relation to outside attributes, is not often explored. When examined, it is usually aggregated to geographic units. While many cities maintain and publish their historic listings' locations, this data is not always used for assessment purposes (other than the traditional, physical state assessment that can or cannot be noted within the geographic data). So, while the initial data is available, it is rarely used to monitor or compare with external attributes. Even in the preferred case of using the data, as in the examples of the insightful reports by PlaceEconomics[1] , the results tend to stress economic features and the spatial aspects conclude in a large aggregation of the data (summing up by geographic units).

The book Preservation and the New Data Landscape (Avrami, 2019) is devoted to the different theoretical and practical aspects of incorporating data within the preservation field.[2] The book concludes with an action agenda that lists, among others, 'Think Beyond the Building' and 'Assess Outcomes.' The two, combined, stress the potential of data to 'support a new era in preservation' and provide us with 'a better understanding of the social-spatial relationships' (p.187).

1 https://www.placeeconomics.com/resources/

2 https://www.arch.columbia.edu/books/reader/370-preservation-and-the-new-data-landscape

Data: Sources & Collection

The data used for the examples within this report is from the National Register of Historic Places (NRHP). This data is available online at the NPS official website.[1] The data includes multiple listings, beginning from 1966. Since it covers the entire US, it was helpful to work within the context of this report. However, this is hardly the sole historic preservation dataset available. States and cities maintain local historical listings, with some of the datasets available online. The examples show the NPS data only, but in practice, these tools can be used on the local listings as well or on a combined file that includes both national and local listings. It is important to note that by choosing this data source, the report does not call to disregard other listings but encourages spatial analysis methods for both.

1 https://www.nps.gov/subjects/nationalregister/data-downloads.htm

Possible Software

While the software and outputs presented in this report are of ArcGIS pro, there are multiple platforms in which the presented tools are available. First, the ArcMap desktop version includes the same tools. In addition, there are coding-based interfaces such as RStudio (for R language) and Python-based platforms. Since this report aims to serve a broad planners' audience, coding is not required and is not covered.

Methods: The Three Parts

The methods presented in this report are organized into three parts: aggregation, grouping, and density. Each part varies in size and includes one or more tools, and varies in the main possible uses. This division helps explain the spatial tools. However, it is essential to note that aggregation and grouping are closer to one another than to density analysis. While aggregation and grouping revolve around the lists themselves, density analysis is about their influence on their surroundings. The difference between aggregation and grouping is that aggregation can be more subjective or arbitrary than grouping since it relies on pre-existing geographic units. Before presenting the tools in the preservation context, this part introduces the three in more general terms using the pedestrian example below.

AGGREGATION

I have given geographic units, and I want to count my listing through them.

Main uses: **Governing purposes & comparison**, both to self and other spatial characteristics

GROUPING (CLUSTERING)

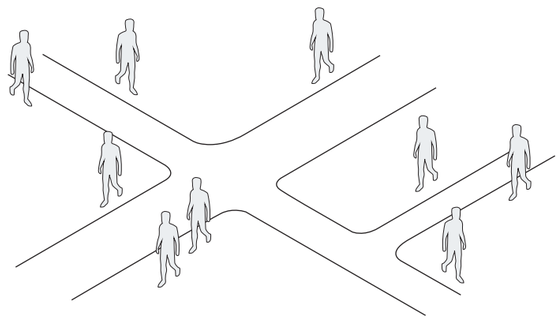
I want to assign my listings into groups, and give to some or all listings a group name.

Main uses: Cross-units collaboration, identification by attributes, smaller scaled groups, re-group listings not by pre-defined units

DENSITY ANALYSIS

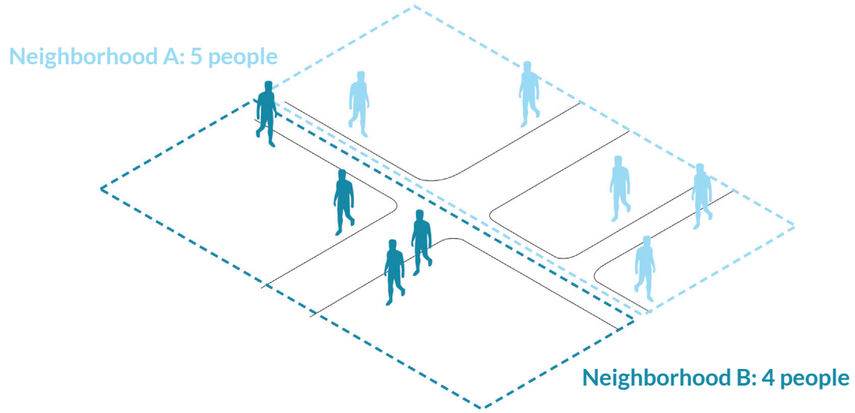
I want to identify high-density or low-density areas of the listings

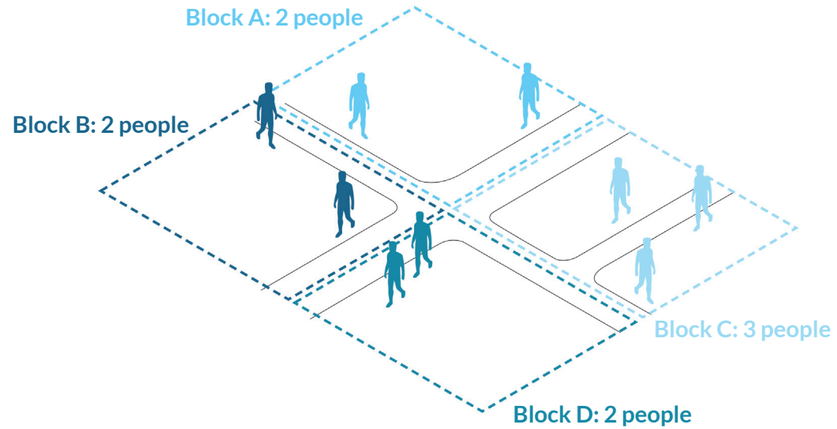
Main uses: Identify smaller-scaled areas for targeted intervention, review effect on surrounding



To present the differences between the three parts, we will use an example of pedestrians walking within a city.

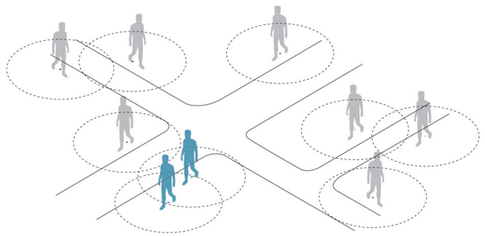
Pedestrians Example: Aggregation





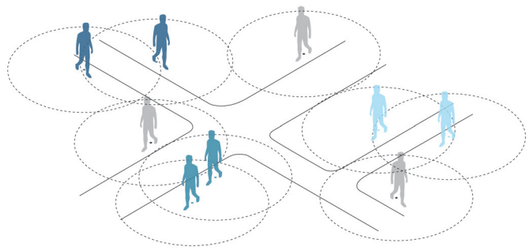
In these two examples, we can pedestrians are counted by the area they are currently in. First, they are counted by neighborhoods, then by city blocks. As seen in this example, the unit of choice has a great impact on the results. The chosen units might not always be relevant to the listed topic. For example, reviewing the number of pedestrians by municipal sanitation zones might make little sense.

Pedestrians Example: Grouping



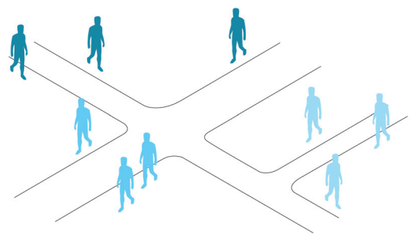
Example 1: Small radius (DBSCAN)

Cluster A, Non-clustered



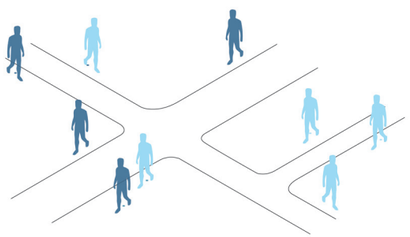
Example 2: Larger radius (DBSCAN)

Cluster A, **Cluster B**, Cluster C, Non-clustered



Example 3: 3 desired groups (K-MEANS)

Cluster A, Cluster B, Cluster C



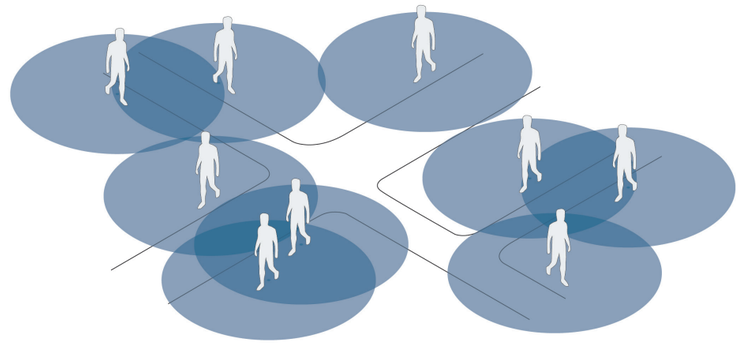
Example 4: By attribute (K-MEANS, NON-SPATIAL)

Female, Male

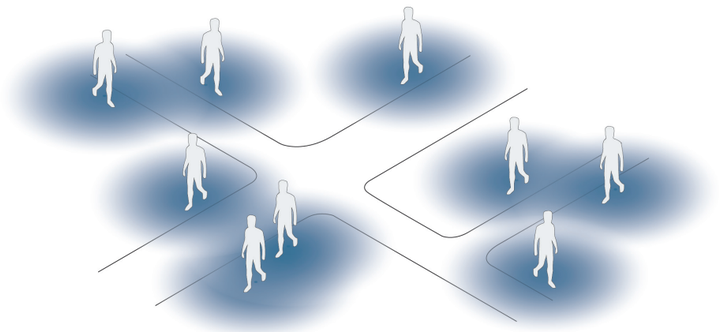
Grouping, or clustering, allows us several ways to look at the pedestrians. We can group them by the distance (radius) to others by setting the radius. In examples one and two, we can see how increasing the radius can affect our groups. In this option, which is based on the DBSCAN method, some people can be left non-grouped, in what ArcGIS automatically describes as 'noise'.

Another option is to group all listings while giving a desired number of groups. This way, seen in example 3, leaves no one ungrouped. To do so, we use the K-means method, selecting the coordinates or the people as inputs. Lastly, we can map non-spatial attributes, such as gender (example 4), as long as the attributes are given in a numeric term. In the report, I will present a fifth K-means based option that incorporates both spatial and non-spatial attributes.

Pedestrians Example: Density



Example 1: Density of fixed affect within radius X



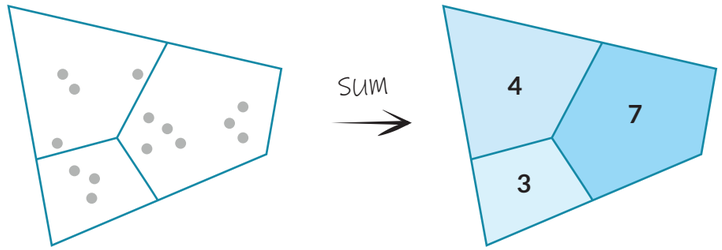
Example 2: Affect in radius X that reduces as the distance from the pedestrian is increased

As stated, density refers to the effect of the lists on the surroundings. Let's assume each pedestrian produces 50 decibels of noise in a radius of X around it. In the first example, the noise level is constant within this radius. To measure the noise, we will review the decibels' sum at each random point in the street. A checking point with a value of 100 Db will be X distance or less from two pedestrians. Now, let's assume the noise reduces within this radius as it moves away from the pedestrian (example 2). In this case, the value we can get at each point has a more nuanced scale, for example, 62 Db if you are in a range of two pedestrians or more. As I will cover density, I will discuss possible measurement tools in more depth.

AGGREGATION

Description

Aggregation of historic listings to a specific geographic unit (polygon) while selecting the desired calculation type, such as sum, mean or median that will be assigned to the geographic unit.



Pro

- SIMPLE- aggregating with polygons does not require high expertise
- USEFUL- municipalities often use artificial geographic units for planning purposes and the ability to quantify listings in each is critical

Con

- Geographic units can change over time (for example, census tracts)
- MAUP- Modifiable Areal Unit Problem (defined in glossary)

Possible Uses

Aggregation is frequently used in the planning context since it allows comparison between different areas (for example, between two districts) and/or between the current and past state of a given geographic unit, assuming it remains unchanged.

Possible uses can be (but not limited to)-
A comparison between different areas within the city
A comparison between different cities, regions or states
Monitoring change over a period of time (an increase in hazardous buildings)

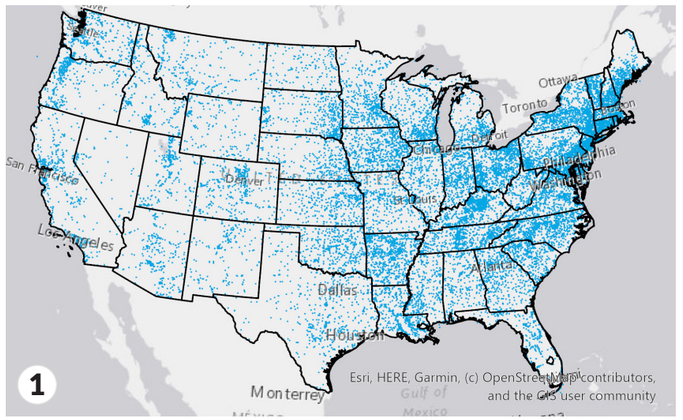
Needed Input Layers Points layer | Minimum cluster size | Search radius

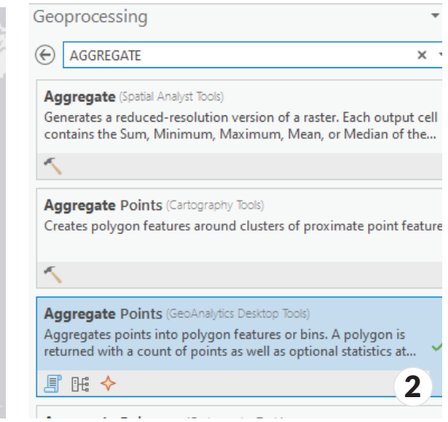
Overall complexity level

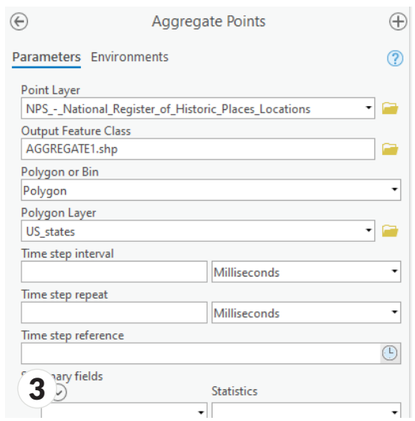
AGGREGATION

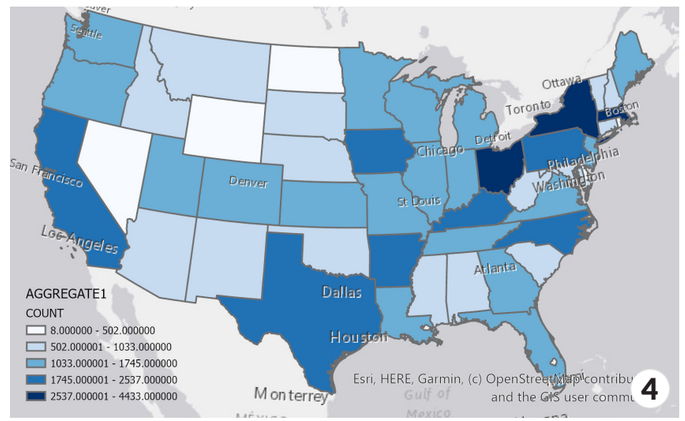
Example

Using NPS historic listing open data, we wish to count the listings per state. Steps:
1. Dragging the two input layers into ArcGIS
2 + 3. Searching and choosing the 'agregate points' tool and filling the desired parameters
4. After a new layer is added, adjusting the appearance according to your needs









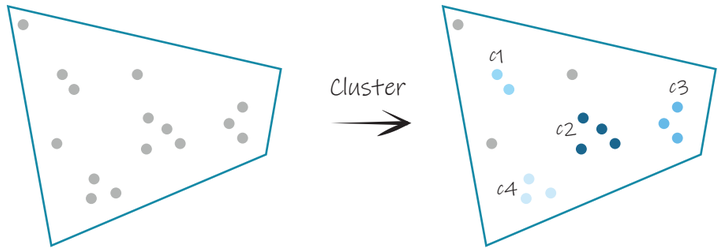
Notes

While using this method, one should pay attention to the polygon shapefile. It needs to be clean, and without overlapping areas, or unintentional gaps. Otherwise, the results will be less accurate.

CLUSTERING: Density-Based Only

Description

Defining groups (clusters) of nearby points within a listings map according to their distance from one another, using the DBSCAN* method (*in glossary).
The groups' amount is not preset. It depends on minimum listings per group and radius size, and the user provides both.



Pro

- Not limited to artificial boundaries
- Size and scale are adjustable according to needs

Con

- A planner must have a preliminary notion of the size and radius to create meaningful clusters

Possible Uses

Review possible groups of listings based on their location regardless of municipal or any artificial boundary an area might contain or be divided to.

Possible example uses:
- Reorganizing existing inventory in a way that is preferable for on-site surveyors
- Prioritizing intervention areas in a way that promotes cross-country or municipality collaborations
- Issuing a map for the local or non-local population to tour a historic area by foot.

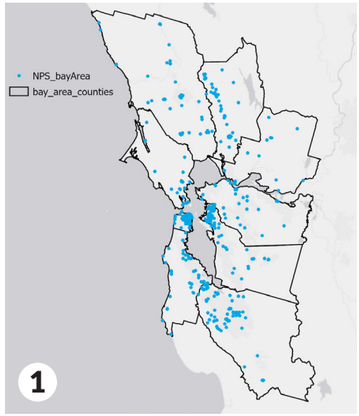
Needed Input Layers Points layer | Minimum cluster size | Search radius

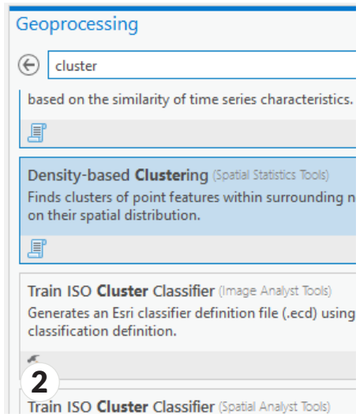
Overall complexity level

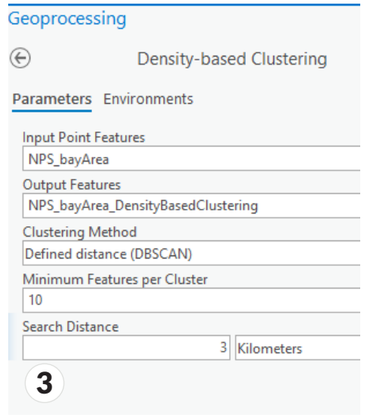
CLUSTERING: Density-Based Only

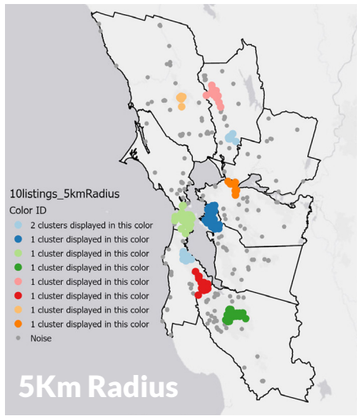
Example

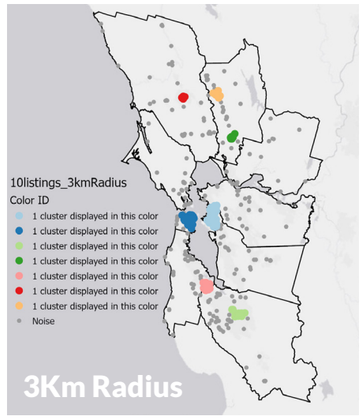
Clustering NPS listings across CA bay area's eight counties
1. Dragging the listing layer into ArcGIS (after choosing the points using 'select by location')
2. Searching & choosing 'aggregate points' tool, under 'clustering methods' choose DBSCAN
3. Testing different clusters minimum and radius sizes before selecting the most suitable size

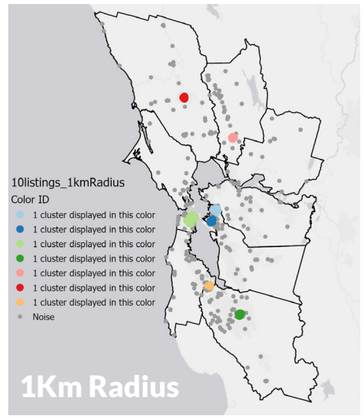












Notes

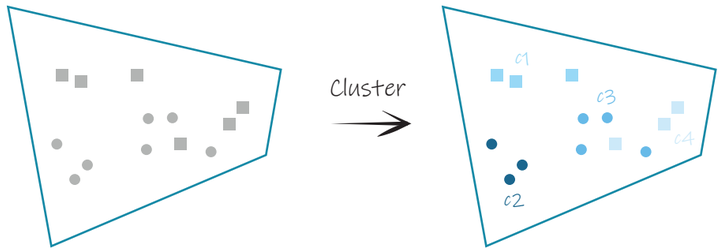
We can see that different radius size alters the size, amount and location of the groups. It is essential to choose the most suitable radius for each need.

Also, we can see that clusters don't necessarily fall within one county boundary. They are also not distributed equally between the eight counties. The results underscore the strength of clusters versus the aggregation method covered before.

CLUSTERING: Multivariate

Description

Creating clusters based on both spatial (coordinates) and non-spatial attributes of listings. The number of factored attributes may vary, but they have to be numerical ones. Also, the coordinates must be chosen, or else it's a non-spatial test (covered next, , in page 16). This method is of K-means clustering, detailed in glossary.



Pro

- Allows nuanced clustering that incorporates specific details the data might already contain

Con

- The planner needs to decide which non-spatial attributes are relevant to the test
- The tool can only use numeric attributes as inputs

Possible Uses

- Grouping buildings according to their year of construction to create a specified or periodic touristic clusters/maps
- Mapping levels of hazard, to allow surveyors efficient route or assigned area
- Testing if a specific feature or a combination of such (for example, amount of broken windows) is linked to particular areas to prioritize areas or neighborhood for additional funding or intervention.

Needed Input Layers Points layer | Polygons layer (geographic units)

Overall complexity level ○ ○ ● ● ●

CLUSTERING: Multivariate

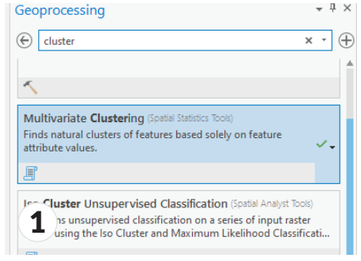
Example

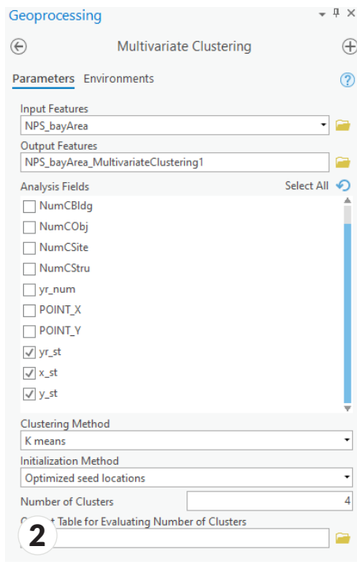
Clustering NPS listings across the CA Bay Area's eight counties according to location and year listed in the NPS list. Steps:

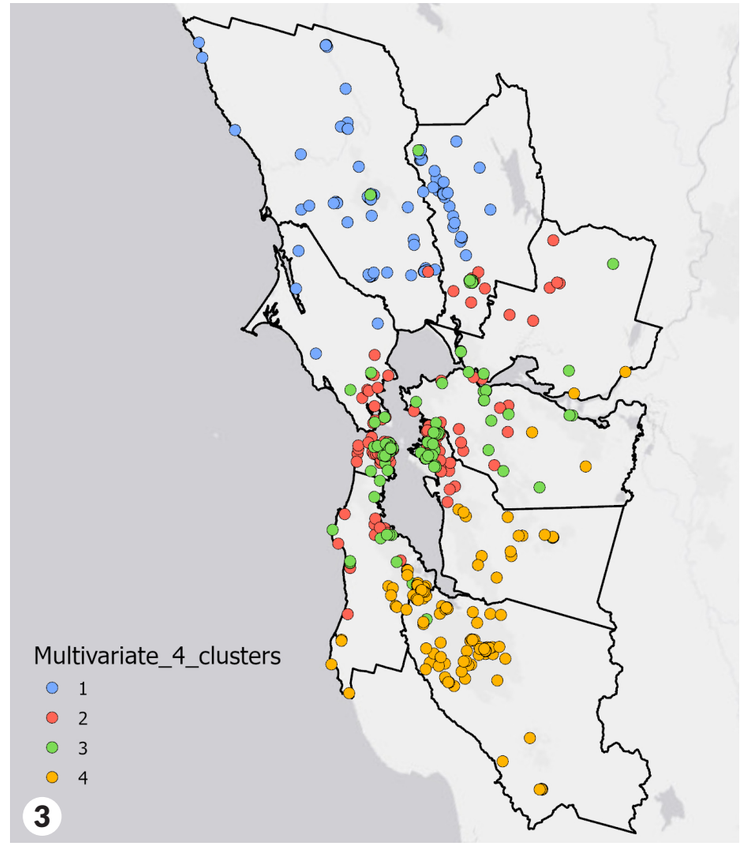
1. If the points don't have x and y columns, add these using the 'Add Geometry Attributes'* tool (*in glossary)
2. Normalize each column using this formula:

   1 + (!old_column! – minimumValue) * (desired scale-1) / (maxValue – minValue)

   (further instructions for this process are in glossary) **Make sure to normalize the x and y columns as well!**
3. Searching and choosing 'Multivariate clustering' tool, under 'clustering methods' choose K means
4. selecting the points file -> desired attributes in their normalized version (including the normalized X and Y) and desired number of clusters.







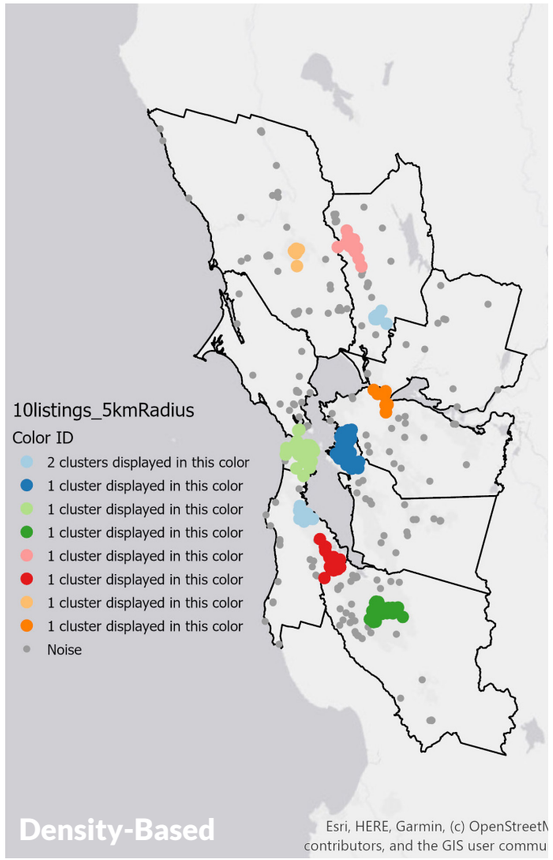
Notes

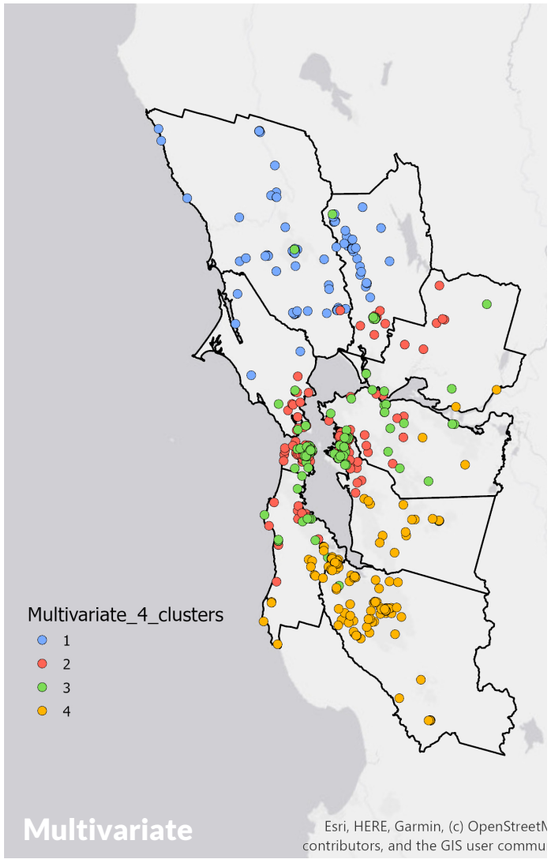
All points receive a cluster value in this clustering option, without any 'noise' points as seen on page 12

CLUSTERING: Comparison

Density-Based vs. Multivariate

Some points are left outside of clusters (noise) setting minimum points per cluster AND setting specific search distance	All points are assigned a cluster value setting specific clusters number choosing additional numeric attributes to factor





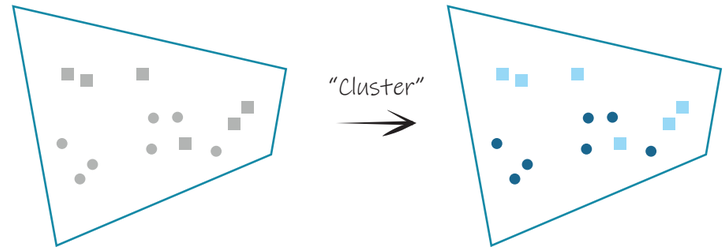
Notes

The selection of a clustering method depends on the need. If all we care about is the location, and we are willing to accept points that are outside of the clusters, we should use the DBSCAN. If some attributes matter, or all points need to be sorted into groups, we should create the more elaborate model of K-means.

CLUSTERING: Non-Spatial

Description + Steps

This is essentially a mapping option. If all you want is to show ONE attribute, you should use the appearance & legend options instead. If you wish to show several non-spatial attributes combined, you can:
1. Normalize each attribute in the way that is covered on page 14
2. Choose the 'multivariate clustering' and select only these attributes from the list + desired clusters



Pro

-

Con

The non-spatial clustering tool is a rather complex and not very simple way if your goal is to map attributes. Consider using a series of side-by-side maps instead.

Possible Uses

If two attributes are inherently linked, you might consider grouping and displaying them together. This is a mapping option that is not related to the spatial attributes of the listings.

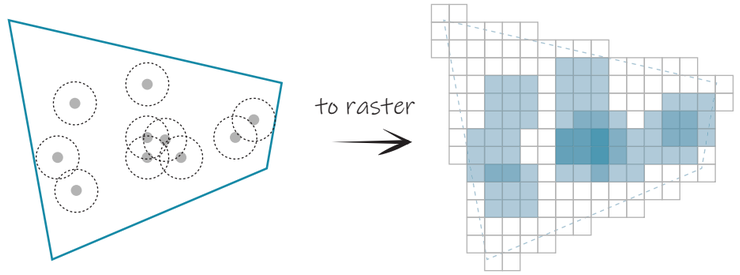
Needed Input Layers Points layer | Selected numeric attributes | # of clusters

Overall complexity level

DENSITY: Point Density Analysis

Description

This method, and the one follows, are both raster-based. In other words, they turn the initial vector layer input (in this case, of points) to a raster (pixels) output. This method takes defined cell size (for the raster layer) and an 'influence' radius. It then turns the points into circles of weights and sums the values in each cell. The point density method can give different weights to points by a specific numeric attribute, but this option is not utilized in this example.



Pro

- Testing the influence of preservation listings on their surroundings.
- Identifying areas where a listing has the highest density.
- Clear, easy to read, and to grasp

Con

- The radius can create coarse boundaries within the map, especially if an overly large radius for the purpose used.
- Radius' size is a trial and error process
- The reviewed influence might not affect the surrounding in a linear way

Possible Uses

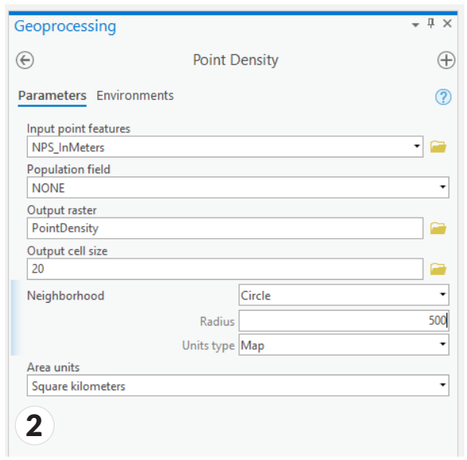
- Assessing the areas that will benefit from publicly used preserved buildings
- Construct a survey of a sample of buildings in an area with highly 'historic' character
- Test possible gentrification phenomena in said area
- Creating a new tourism-walk map

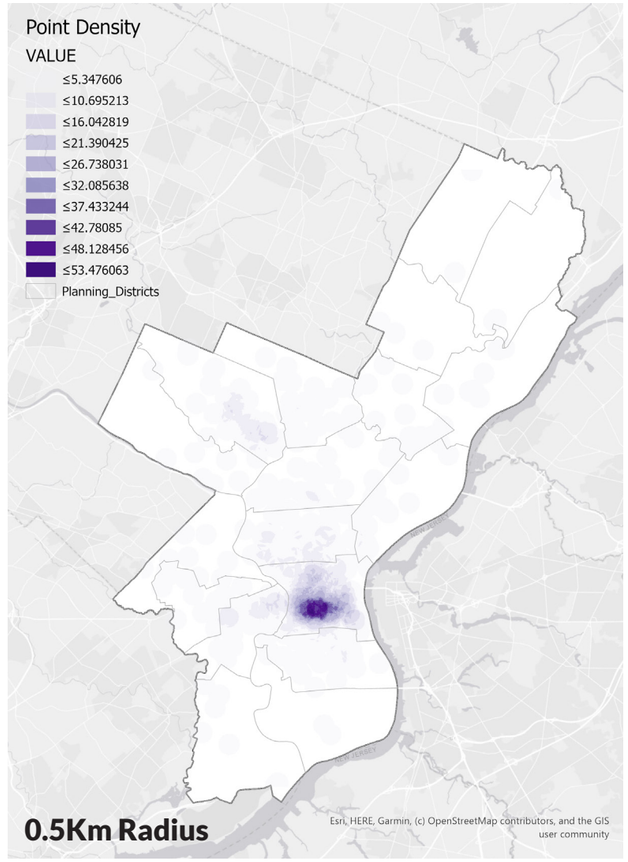
Needed Input Layers Points layer | Output cell size | Radius

Overall complexity level

Example

NPS listings in the Philadelphia area are used for the density examples. Steps:
1. Search the 'point density' tool in the geoprocessing catalog
2. Provide the three inputs. Make sure you choose the units type you wish to use
3. Review the output raster and test other inputs of cells or radius size if needed





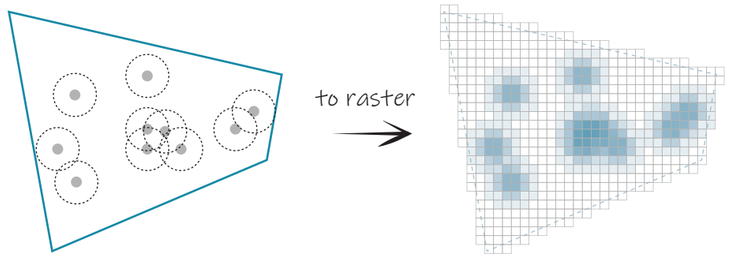
Notes

In Philadelphia's case, the densest area is the Central planning district (center city).
To display the output raster's values, you can use multiple breaks. Click the 'appearance' tab, choose 'symbology,' and in the options opened on the right, switch 'method' and the number of classes to your preferred display.
In this example, the planning districts are used as a background. Though the raster layer can be read alone, adding a layer such as the planning districts helps simplify the reader's navigation.

DENSITY: Kernel

Description

The Kernel density method places a quadric distribution on each point in the given sample, then summing up each cell's value in the observed area and dividing by the overall points' number. The formula also includes the input radius. In a kernel density analysis, the effect of a given point decreases as its distance increases from the point under analysis. For more on the Kernel method, see glossary.



Pro

- Smooth surface output
- Assumes affect changes with distance

Con

- The planner needs to decide on the scale of cells and search radius in a way that fits the specific purpose. An incorrect choice can lead to a meaningless outcome
- It might be difficult to interpret without setting clear divisions

Possible Uses

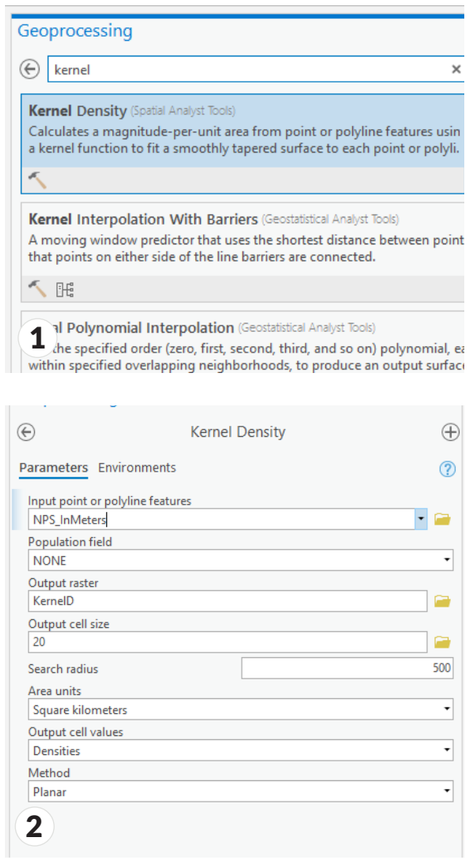
- Identify areas with very high or very low density of listings
- Creating polygons (targeted areas) of the highest percentile/quantiles/other units of listings density (as seen on page 25)

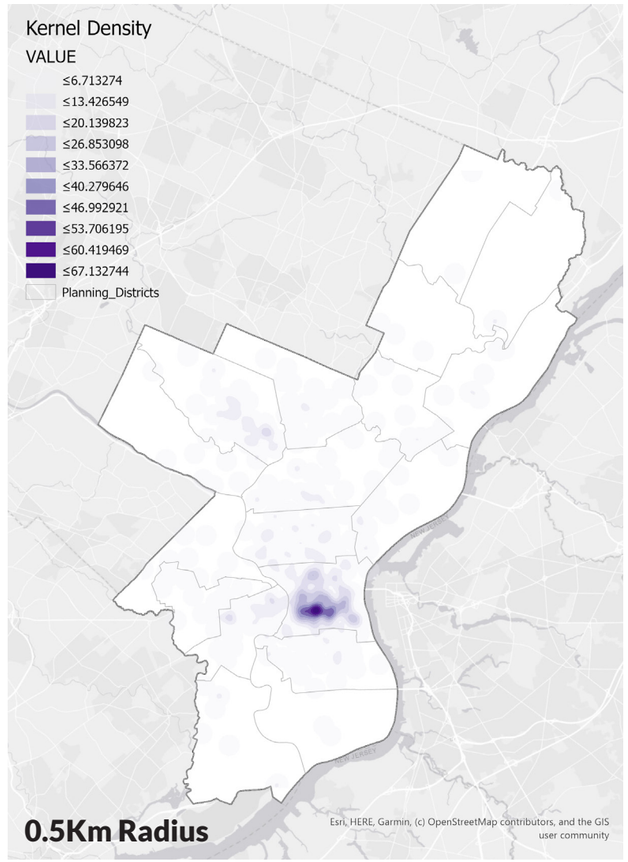
Needed Input Layers Points layer | Output cell size | Radius

Overall complexity level

Example

Steps:
1. Search the 'Kernel density' tool in the geoprocessing catalog
2. Provide the three inputs. Make sure you choose the units type you wish to use
3. Review the output raster and test other inputs of cells or radius size if needed





Notes

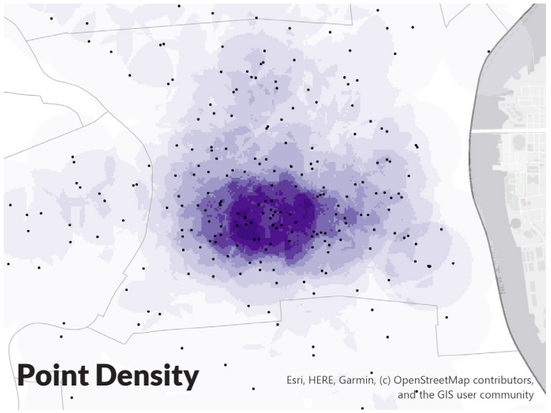
Similar to the point analysis, this tool transforms the layers from vectors (points) into raster (pixel). While doing so, we need to set the cell definition and pay attention to our project's raster environment definitions.

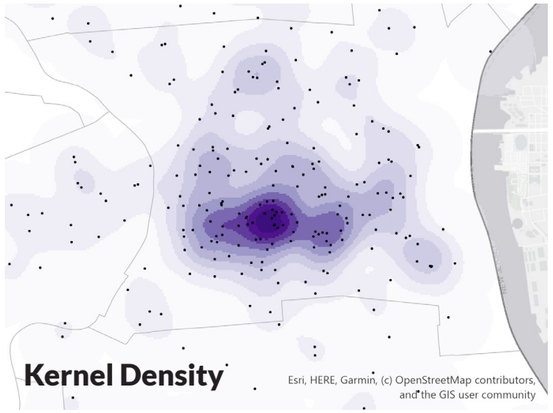
DENSITY: Comparison

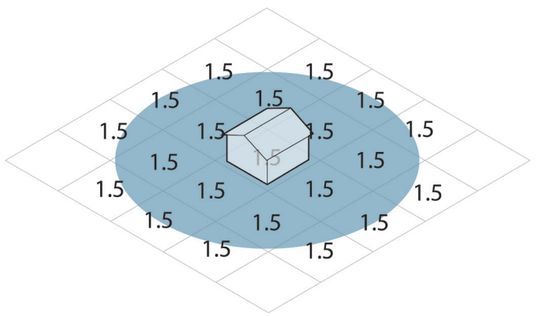
Point Density vs. Kernel Density

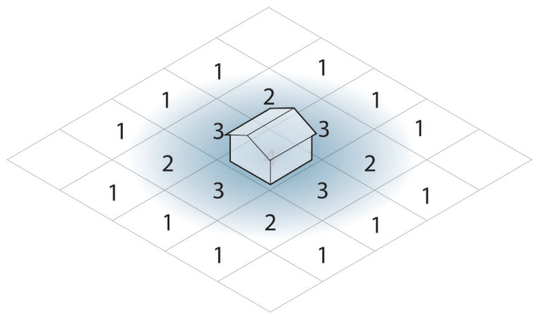
Point density calculates the effect of each point as fixed across the given radius.
Different weights can be given to the points if needed, by selecting a specific attribute column

Kernel Density is calculated differently- the value in the given radius decreases as we move away from the point. An example can be seeing a historic property- even if no other buildings around, if we are far away, we are less able to see its details.









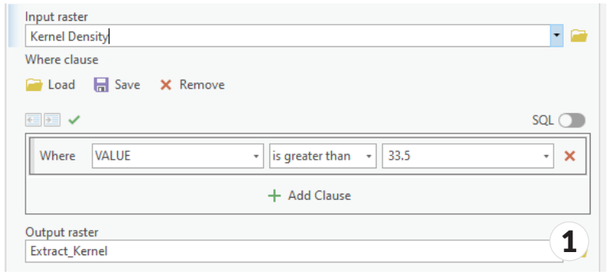
Notes

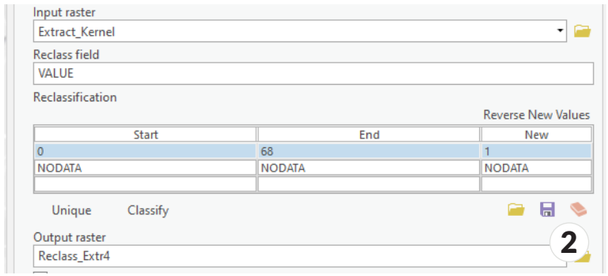
For the sake of comparison, we will zoom in to Philadelphia's CBD area. In the diagrams above, the difference in the calculation is shown. Because of the way value is calculated within the radius, the output of the Kernel density is smoother across space and with fewer edges. However, selecting the method should be according to the relationship you are testing and assumptions regarding it: is the effect of the preserved buildings decreasing with distance, or are you assuming a fixed effect within X distance from each listing?

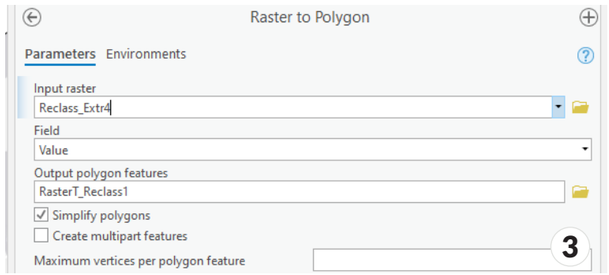
DENSITY: Possible Use & Presentation

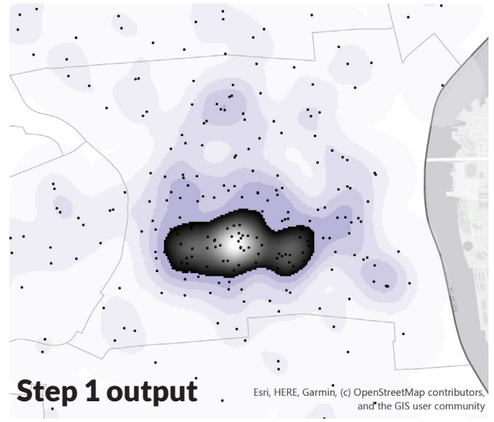
Highest Densities (subset values) Polygons

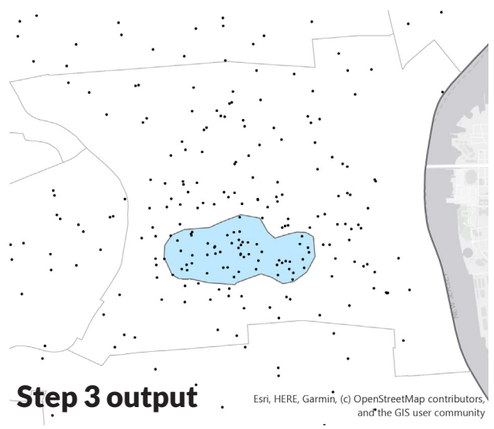
Steps:
1. Using the 'extract by attributes' tool, select the relevant part of the value and extract to a new raster
2. Search and use the 'reclassify' tool to create an additional raster with the value of 1 (or other fixed value)
3. Turn the raster into polygons using the 'raster to polygon' tool











Notes

- You can subset the highest percentile, quantile, half, or any other value that is meaningful to the project.
- When reclassifying, make sure to include the entire range
- In this example, there is only one polygon, but this process and tools can result in multiple polygons output (if the original raster has multiple peaks).
- This process can be implemented on point density output as well

METHODS SUMMARY

METHODS SUMMARY

	METHOD	DESCRIPTION	INPUTS	PRO	CON	USES EXAMPLE	COMPLEXITY
Aggregation	**Aggregation**	Aggregation to a specific geographic unit (polygon) while selecting a summing numeric action, that will be assigned to the geographic unit.	• points • polygons	• Simple • Highly Useful	• MAUP (in glossary) • Units can change	comparison with other phenomenon or over time	1/5
Grouping	**Density-Based Clustering**	Defining groups of nearby points within a listings map according to their distance from one another.	• points • Min. Clusters Size • Search Radius	• Not limited to artificial boundaries • Adjustable	• must have a preliminary notion of size and radius	Re-organizing existing inventory, Prioritizing intervention areas	2/5
Grouping	**Multivariate Clustering**	Clusters based on both spatial (coordinates) and non-spatial numeric attributes of listings.	• points • Selected numeric attributes • # of clusters	• Allows nuanced clustering that incorporates specific details	• choosing relevant attributes • only numeric	Grouping by year, mapping hazardous levels, testing specific feature	3/5
Grouping	**Non-Spatial Clustering**	Clusters based on non spatial numeric attributes of listings.	• points • Selected numeric attributes • # of clusters		• A complex way to map attributes	Grouping by multiple attributes, such as year and style	1/5
Density	**Point Density**	Assigning points circles of fixed weights and summing the values for each cell of an output raster	• points • output cell size • radius	• Testing influence • Identifying high density areas	• Coarse boundaries • Trial and error of radius & cell	Assesing effect, test gentrification, create a walk-in tourist map	3.5/5
Density	**Kernel Dnsity**	Assigning points circles of decreasing weights and summing the values for each cell of an output raster	• points • output cell size • radius	• Smooth surface • Assumes changing effect	• Trial and error of radius & cell • might be hard to interpret	Identify the areas with the highest or smallest density of listings	4/5

CONCLUSION & TAKEAWAYS

There are no good and bad spatial tools, neither are there better or worse ones. As seen in the last spread, the virtues and possible uses of the tools vary. Tool suitability changes according to what the planner, or advocator, wishes to accomplish.

Starting the process of this report included identifying three main needs group:

1. Comparison
between different areas, assessment of change over time (temporal), and between categories

2. Exploring Relationship
between historical listings and an external feature or phenomenon

3. Prioritization & Identification
of needed intervention areas and 'desert' areas

While these are not all-encompassing, they sum up the main possibilities or challenges planners and advocates might face (in addition to inventory keeping that is excluded from this list). There is not one specific tool that matches each group. Generally speaking, the first group will usually benefit from aggregating, although tracking change patterns using density is also possible. Grouping will be less useful unless one uses non-spatial clustering to compare different categories.

The second group, exploring relationships, can benefit from density analysis coupled with another phenomenon or attributes. For example, what is the density of new businesses in relation to the density of historic listings? Seeing the two side by side or exploring the numeric ratio can provide insights. But, it can also be carried by aggregating, in case the additional attributes are given within defined units. An example of that might be the reviewing the relationship between population size and listings. If we look at a tract level, we can create a ratio of listings per population by aggregating our listings.

Lastly, prioritization can be carried out in various ways, but clustering is an effective tool with which to do so. If we want to prioritize buildings that were built before the 70s, are badly maintained, and in high proximity to one another, we can give these attributes as inputs to cluster analysis (as long as the maintenance characteristics are translated to a numeric value, like a ranking on a scale or binary attribute). The output can be used for a pilot program or a small-scale intervention. Of course, we can also prioritize by density analysis or aggregation, depending on the specific needs.

In addition to the desired use, there are a few main factors that can lead to choosing the appropriate tool and distinguishing one tool from another.

Pre-existing Geographic Units

In many cases, pre-existing geographic units are used. For a municipality that has already standardized its collection to particular units, continuing to use the same units for preservation might be the reasonable thing to do. A planner might want to assess the number of listings per unit for neighborhoods, districts, or other pre-defined areas. Creating an alternative method for preservation might only be time and resource-consuming and not feasible. That said, stand-alone projects, such as creating a new tourist map or selecting a smaller sample of historic buildings for an in-depth survey by experts, can happen separately from these units.

The scale on which you work

In relation to the previous factor, one of the most important things that a planner needs to decide before addressing the datasets is the preferred or suitable size of her/his project. In all tools presented here, the correct scale needs to be decided. If we are reviewing preservation locally, using oversized units or radiuses will have little meaning and vice versa.

The Planner's Audience

When choosing a method, it is important to pay attention to the output appearance and clarity. The outputs are clearer when a legend is used. In the case of the density analyses, the row output is of a raster layer that might require more clarifications. A way to simplify this output, if needed, is presented on page 22.

Level of Expertise & Additional Learning

The six tools covered in the report range from easy to medium-hard level of complexity. However, these are subjective observations. While some might see the overall list as pretty straightforward operations that are easy to comprehend, others might struggle with the platform used or the interoperation of the results.

Users with beginners' level should not be persuaded to give up these tools for two main reasons. First, the amount of user-friendly free guides, supports, tutorials, and Q&A available

online is ever-growing. ArcGIS has a designated web page for every tool that is proposed in the software. Second, while it is preferable and useful to fully understand the mathematical mechanisms at the core of each tool, it is not always essential for the purposes suggested here. In other words, since these tools are assigned on a list of points where the output is immediately given, simply by looking at the result (and potentially tweaking the used outputs), one can get the output they desire. There might be mistakes or better, finer outputs, but many of the proposed uses are for internal day-to-day work and can always be adjusted later.

Other tools

Lastly, there are other and more complex tools available on the different platforms. This sample is of frequently used and relatively simple tools. If a user feels or suspects that the tool they need is missing, they are encouraged to look for an additional alternative on the official ArcGIS site online, R platform guides, or any other reliable source.

Glossary

• **'Add Geometry Attributes' tool**
This operation adds X and Y coordinate values to the record of each feature in the attribute table of a point feature class. The user can define the desired units and the coordinate system of the output. The added columns are named 'POINT_X' and 'POINT_Y'.

• **Density-Based Spatial Clustering of Applications with Noise (DBSCAN)**
The DBSCAN model uses "a simple minimum density level estimation, based on a threshold for the number of neighbors, minPts, within the radius (with an arbitrary distance measure). Objects with more than minPts neighbors within this radius (including the query point) are considered to be a core point. The intuition of DBSCAN is to find those areas, which satisfy this minimum density, and which are separated by areas of lower density."[1] The DBSCAN clustering method creates an output of clusters and 'noise'- points that are within the lower density areas and are not considered within clusters.

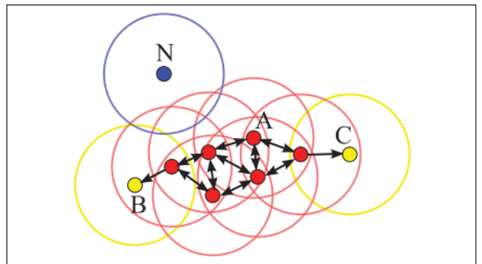


Fig. 1. Illustration of the DBSCAN cluster model.

Figure: DBSCAN Illustration.
Source: Erich Schubert et al., "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN," ACM Trans. Database Syst. 42, no. 3 (July 2017): 19:3

• **K-means**
K-means is a nonhierarchical clustering method. According to Salkind, "The k-means algorithm was originally developed as a method of computing an optimal partitioning in the sense of within-class variance of an n-dimensional population on the basis of a sample."[2] While the method will not always provide the optimal output solution, it will produce clusters that are efficient. The level of efficiency depends, to some extent, on the research choice of suitable clusters' amount.[3]

The algorithm repeatedly 1) associates each observation with whichever of a specified number of (initially arbitrary) points lies closest in an N-dimensional space whose axes correspond to observed attributes, and 2) repositions each of those points to the

1 Erich Schubert et al., "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN," ACM Trans. Database Syst. 42, no. 3 (July 2017): 19:2, https://doi.org/10.1145/3068335.
2 Neil Salkind, "Encyclopedia of Measurement and Statistics," May 4, 2021, 2, https://doi.org/10.4135/9781412952644.
3 Salkind.

N-dimensional centroid of all of its currently associated observations. The K-means algorithm can calculate the distance in various ways, with Euclidean distance squared being the most popular.

• Kernel Density Equation

"Kernel estimation (locally weighted averaging) is an extension of local averaging."[4] The Kernel Density method process is essentially equivalent to the repeated averaging of a value over a defined pixel neighborhood, as seen in the house illustration on page 21. The value for each cell in the output raster is calculated as the sum of values of all the kernel radiuses (surfaces) that overlay the specific cell. ArcGIS pro uses quadric kernel function (which equivalent to adding together many SE kernel functions, that use Gaussian distribution).

• Modifiable Areal Unit Problem (MAUP)

"The essence of the MAUP is that there are many ways to draw boundaries to demarcate space into discrete units to form multiple spatial partitioning systems."[5] A change in the dividing geographical units of an area might lead to a change in the analysis results when using aggregation, as seen on page 6 in the pedestrians' example. The MAUP is not a "problem" but, rather, an acknowledgment of the fact that statistics summarizing values associated with a set of observations are always dependent on the manner in which those observations are grouped. As such, it does not call for a "solution" so much as it calls for an understanding of the relationship between groups of observations and the inferences to be drawn from their values.
The use of Kernel Density is a way to address the MAUP by processing point data in terms of continuously varying densities rather than predefined boundaries.

• 'Normalize' columns formula and steps

To normalize column values, we will use the 'calculate field' option. Either search this tool or right-click on a new field to open the calculate field option. Within it, choose the column you wish to normalize and the new field name. Then, adjust this formula to the equation :
*1 + (!old_column! – minimumValue) * (desired scale-1) / (maxValue – minValue)*
Lastly, run the operation and verify that the new column was added to your layer.

4 "Nonparametric Simple Regression" (Thousand Oaks, California: SAGE Publications, Inc., 2021), 2, https://doi.org/10.4135/9781412985307.
5 David Wong, "The Modifiable Areal Unit Problem (Maup).," in The SAGE Handbook of Spatial Analysis, by pages 105-123 (London: SAGE Publications, Ltd, 2009), 2–3, https://doi.org/10.4135/9780857020130.