Spatial Modelling and Analysis

Lecture 11

Point Pattern Analysis

Lecturer: Dr. Tingneyuc Sekac B.Sc., M.Phil., Ph.D.
PNG University of Technology, 2022

Lecture Outline

1. Introduction to point Pattern Analysis (PPA)
2. Centrography
3. Density Based Analysis
4. Distance Base Analysis
5. Application of PPA
6. References

1. Introduction to point Pattern Analysis (PPA)

Point pattern analysis (PPA) focuses on: analysis, modeling, visualization, and interpretation of point data.

With the increasing availability of big geo-data, such as mobile phone records and social media check-ins, more and more individual-level point data are generated daily.

PPA provides an effective approach to analyzing the distribution of such data.

Yuan, Y., et al,. (2020).

1.1. Point Distributions

Point pattern analysis, in its basic form, deals with the distribution of homogeneous points, that is, one type of points.

This does not imply that we cannot treat points that are not homogeneous in the real world.

In basic point pattern analysis, we focus only on the spatial aspect of point distributions, neglecting their attributes.

Yuan, Y., et al,. (2020).

1.2. PPA Properties

PPA studies the spatial distribution of points .

Properties of point patterns are analyze and modeled in different ways.

two categories of Point properties:

1. first-order properties

Focuses on the characteristics of individual locations and their variations across space,

2. second-order properties.

Focuses on properties that concern not only individual points, but also the interactions between points and their influences on one another.

Source: Gimond. M, 2021

2. Centrography

A very basic form of point pattern analysis involves summary statistics such as:

- **mean center,**
- **standard distance**
- **standard deviational ellipse.**

2.1. Centrography Vs More Power Approaches

The Centrography point pattern analysis techniques were popular before computers.

It is a summary statistics and is too concise and hide far more valuable information about the observed pattern.

The more powerful analysis methods can be used to explore point patterns:

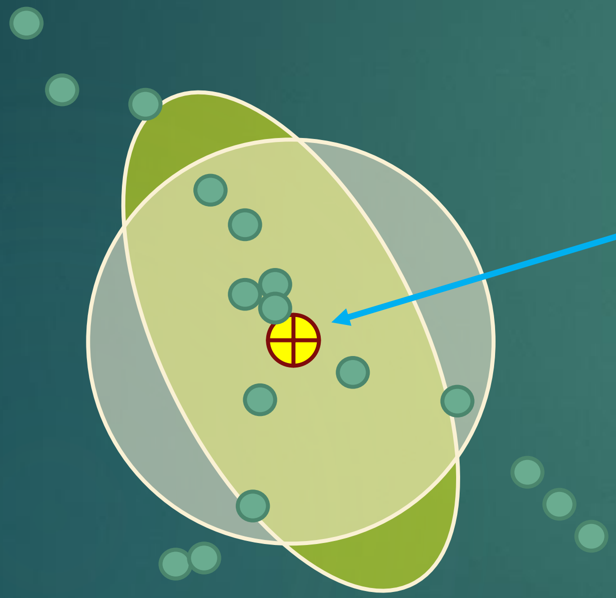
1. **Density based** approach

2. **Distance based** approach.

Source: Gimond. M, 2021

2.2. Mean Center

Computer average X and Y coordinate Values

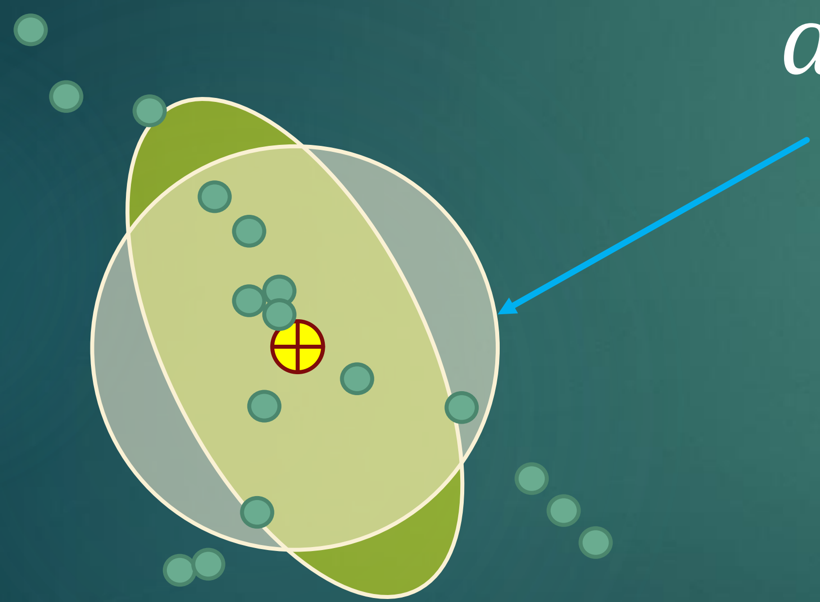


$$\bar{S} = \left(\frac{\sum_{i=1}^{n} X_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n}\right)$$

Source: Gimond. M, 2021

2.3. Standard Distance

Measure of the variance between the average distance of the features
To the mean center.

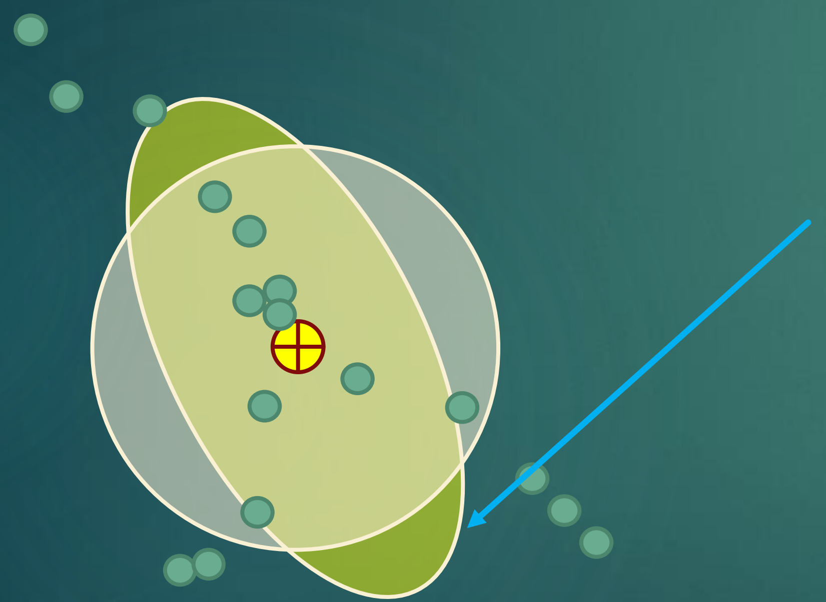


$$d = \frac{\sqrt{\sum_{i=1}^{n}(x_i - u_x)2 + (y_i - u_y)^2}}{n}$$

Source: Gimond. M, 2021

2.4. Standard Deviational Ellipse

Separate standard distance for each axis..



$$d_x = \frac{\sqrt{\sum_{i=1}^{n}(x_i - u_x)^2}}{n}$$

$$d_y = \frac{\sqrt{(y_i - u_y)^2}}{n}$$

where (μ_x, μ_y) are the coordinates of the mean center, (x_i, y_i) represent the coordinates of a given point i, and n is the total number of points.

Source: Gimond. M, 2021

3. Density Based Analysis

Density based techniques characterize the pattern in terms of its point distribution in the study area

The point properties considered is a **first-order** property of the pattern.

Ex:
the distribution of oaks will vary across a landscape based on underlying soil characteristics (resulting in areas having dense clusters of oaks and other areas not).

Density measurements can be broken down into two categories:

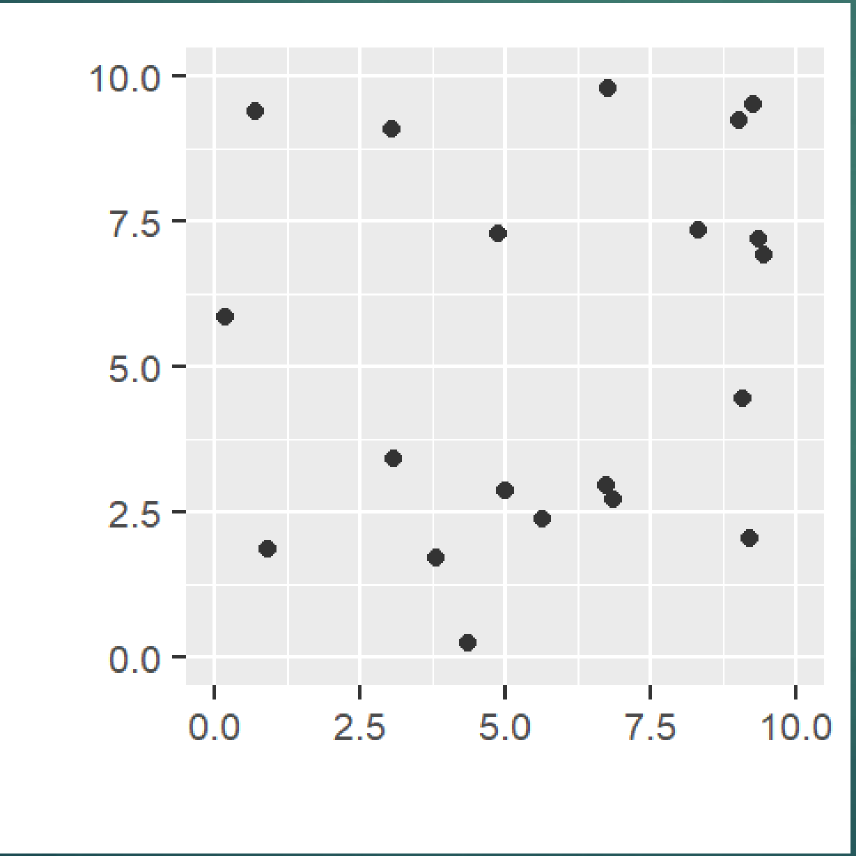
1. **Global**
2. **Local.**

3.1. Global Density Measurement

A basic measure of a pattern's density ('λ') is its overall, or **global,** density.

This is simply the ratio of observed number of points ('n'), to the study region's surface area ('a'),

$\lambda = n/a$



An example of a point pattern where n = 20 and the study area (defined by a square boundary) is 10 units squared. The point density is thus 20/100 = 0.2 points per unit area.

Image Source: Gimond. M, 2021

3.2. Local Density

A point pattern's density can be measured at different locations within the study area.

Such an approach helps us assess if the density–and, by extension, the underlying process' local (modeled) intensity 'λi' is constant across the study area.

This can be an important property of the data since it may need to be mitigated.

Several techniques for measuring local density are available, The two most common methods are:

1. *quadrat density*
2. *kernel density*.

Source: Gimond. M, 2021

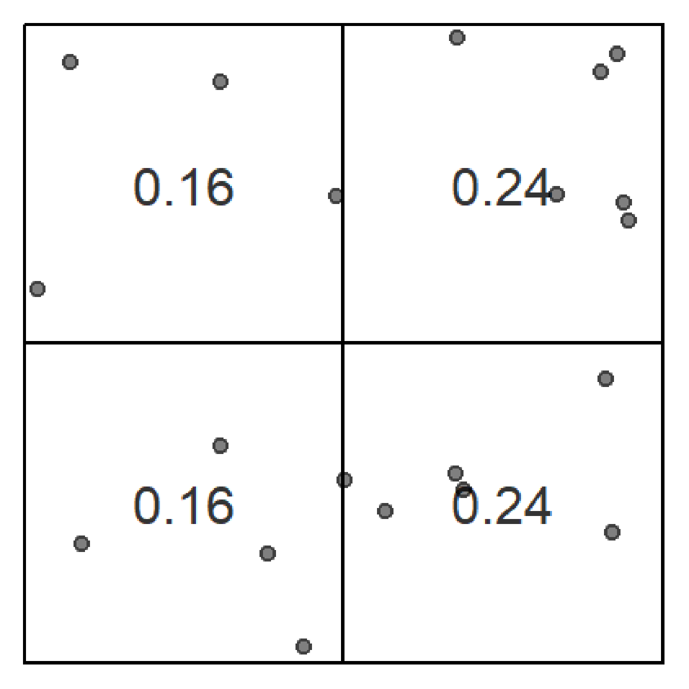
3.2.1. Quadrat Density

Technique requires that the study area be divided into sub-regions (*quadrats*).

The point density is computed for each quadrat by dividing the number of points in each quadrat by the quadrat's area.

Shapes such as hexagons and triangles are also use, here we use square shaped

Quadrant sizes matters



Quadrat count where the study area is divided into four equally sized quadrats whose area is 25 square units each.

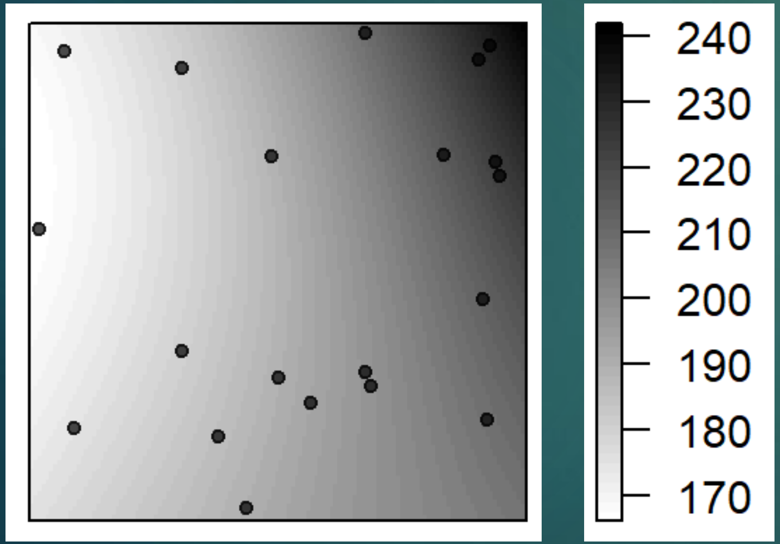
The density in each quadrat can be computed by dividing the number of points in each quadrat by that quadrat's area.

Image Source: Gimond. M, 2021

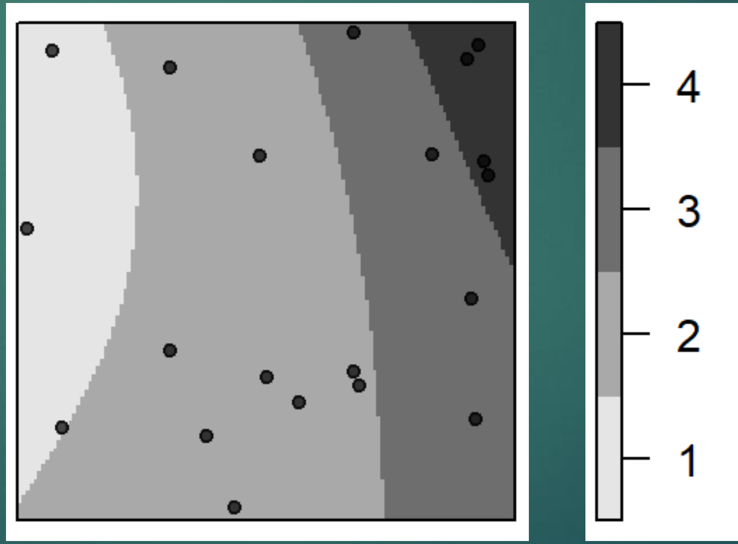
3.2.2. Quadrat regions and Covariate

Quadrat regions do not have to take on a uniform pattern across the study area, they can also be defined based on a **covariate**.

For example, if it’s believed that the underlying point pattern process is driven by elevation, quadrats can be defined by sub-regions such as different ranges of elevation values.



Elevation map



Tessellated Surfaces

Converting a continuous field into discretized areas is sometimes referred to as **tessellation**.

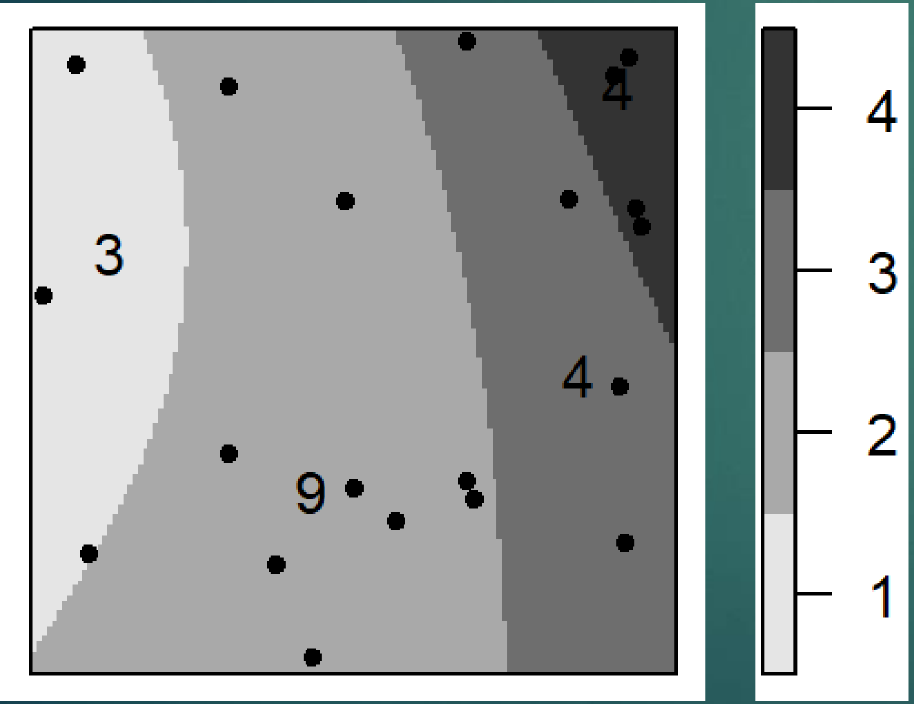
The end product is a **tessellated surface**.

Image Source: Gimond. M, 2021

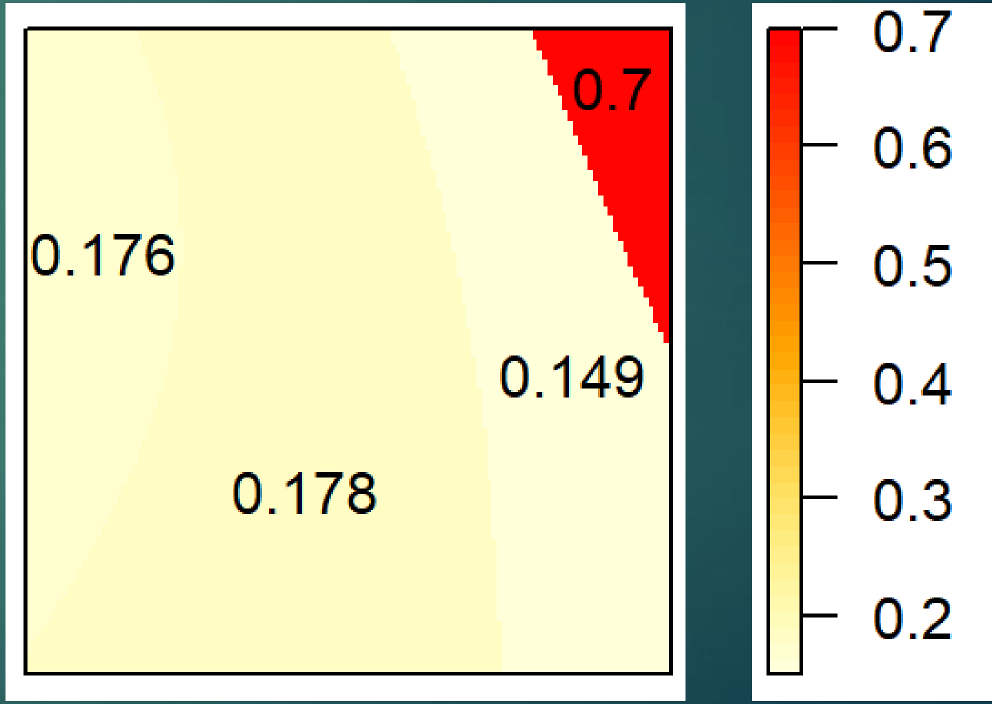
3.2.3. Quadrat Point Density

In our example, sub-regions 1 through 4 have surface areas of 17.08, 50.45, 26.76, 5.71 map units respectively.

To compute these regions' point densities, we simply divide the number of points by the respective area values.



Number of Points In Elevation Sub-Regions



Point Density Calculated

Image Source: Gimond. M, 2021

3.2.4. Surface Tessellate and its impact

It's important to note that how one *chooses* to tessellate a surface can have an influence on the resulting density distribution.

For example, dividing the elevation into *equal area* sub-regions produces the following density values.

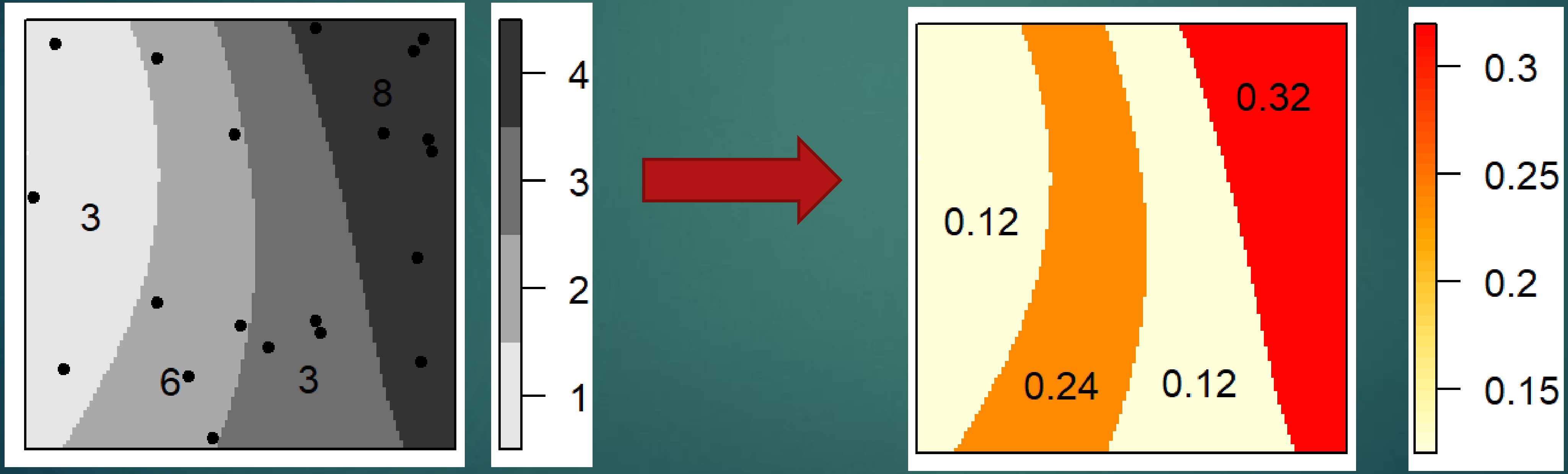


Image Source: Gimond. M, 2021

3.2.5 Kernel Density

The kernel density approach is an extension of the quadrat method:

Like the quadrat density, the kernel approach computes a localized density for subsets of the study area, but unlike its quadrat density counterpart, the sub-regions overlap one another providing a *moving* sub-region window.

The moving window is defined by a **kernel**.

The kernel density approach generates a grid of density values whose cell size is smaller than that of the kernel window.

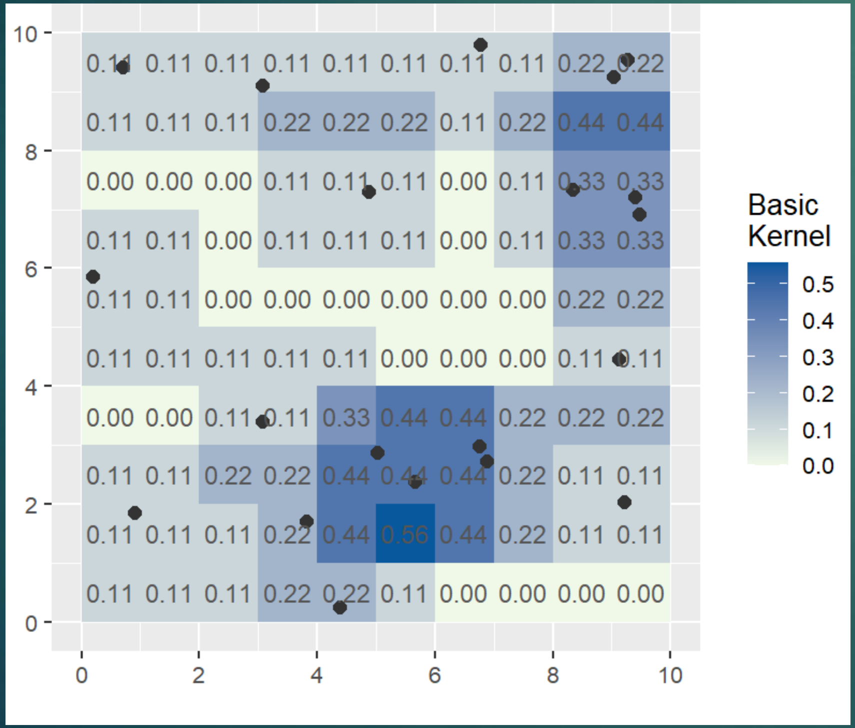
Each cell is assigned the density value computed for the kernel window *centered* on that cell.

Source: Gimond. M, 2021

3.3.6 Kernel Weight Function

A kernel not only defines the shape and size of the window, but it can also *weight* the points following a well defined kernel function.

The simplest function is a **basic kernel** where each point in the kernel window is assigned equal weight.



3x3 kernel density map

Each point is assigned an equal weight.

For example, the second cell from the top and left (i.e. centered at location x=1.5 and y =8.5) has one point within a 3x3 unit (pixel) region and thus has a local density of 1/9 = 0.11.

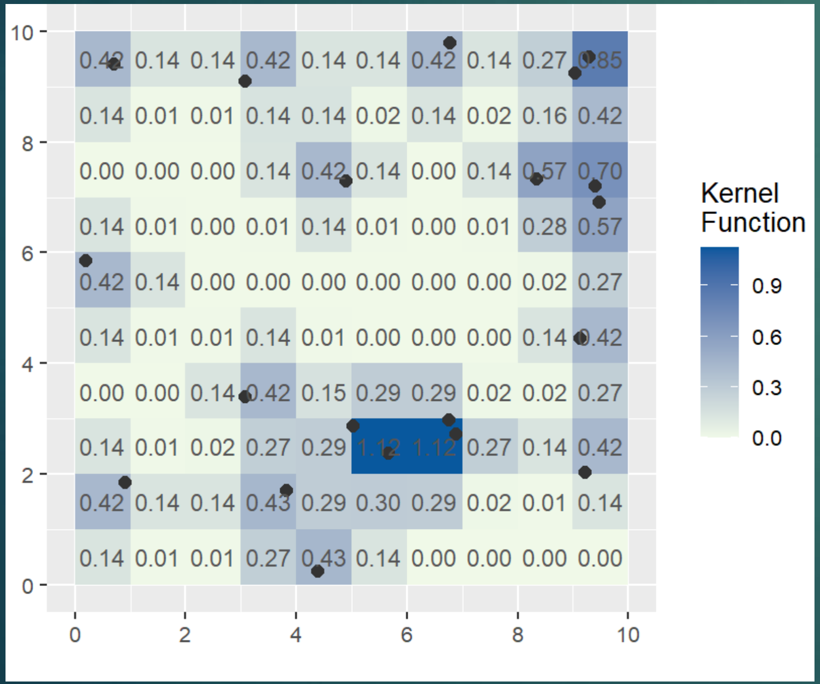
Image Source: Gimond. M, 2021

3.2.7. Kernel Quatic Distribution Function

Some of the most popular **kernel functions** assign weights to points that are inversely proportional to their distances to the kernel window center.

A few such kernel functions follow a *gaussian* or *quartic* like distribution function.

These functions tend to produce a smoother density map.



3x3 quartic kernel function

where each point in the kernel window is weighted based on its proximity to the kernel's center cell.

(typically, closer points are weighted more heavily).

Kernel functions, like the quartic, tend to generate smoother surfaces.

Image Source: Gimond. M, 2021

4. Distance Base Analysis

An alternative to the density based methods are the **distance based methods** for pattern analysis.

The interest lies in how the points are distributed relative to one another

This depends on a second-order property of the point pattern as opposed to how the points are distributed relative to the study extent.

Three distance based approaches are discussed here:

1. The average nearest neighbor (ANN),
2. the K and L functions,
3. the pair correlation function.

Source: Gimond. M, 2021

4.1. Second order Point Property

A *second order* property of a pattern concerns itself with the observations' influence on one another.

For example,

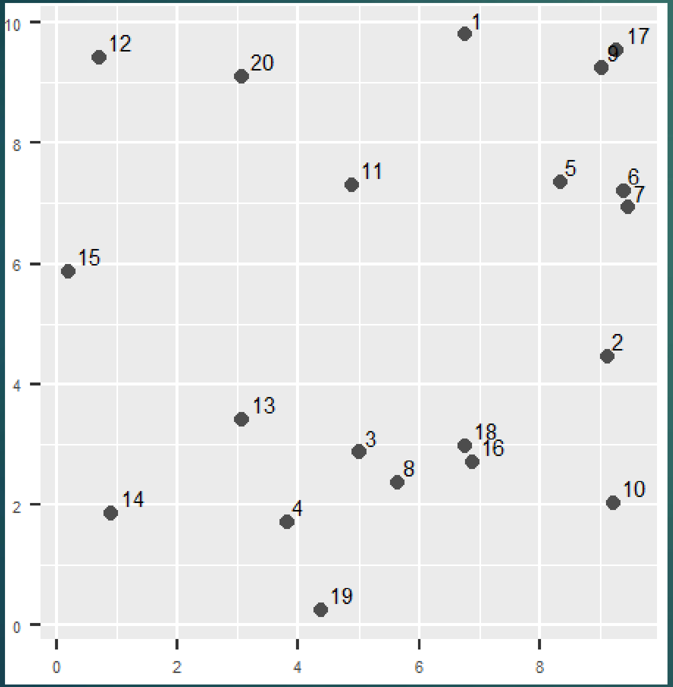
the distribution of oaks will be influenced by the location of parent trees– where parent oaks are present we would expect dense clusters of oaks to emerge.

Source: Gimond. M, 2021

4.2. Average Nearest Neighbor

An average nearest neighbor (ANN) analysis measures the average distance from each point in the study area to its nearest point.

In the following example, the average nearest neighbor for all points is 1.52 units.



From	To	Distance	From	To	Distance
1	9	2.32	11	20	2.55
2	10	2.43	12	20	2.39
3	8	0.81	13	4	1.85
4	19	1.56	14	13	2.67
5	6	1.05	15	12	3.58
6	7	0.3	16	18	0.29
7	6	0.3	17	9	0.37
8	3	0.81	18	16	0.29
9	17	0.37	19	4	1.56
10	2	2.43	20	12	2.39

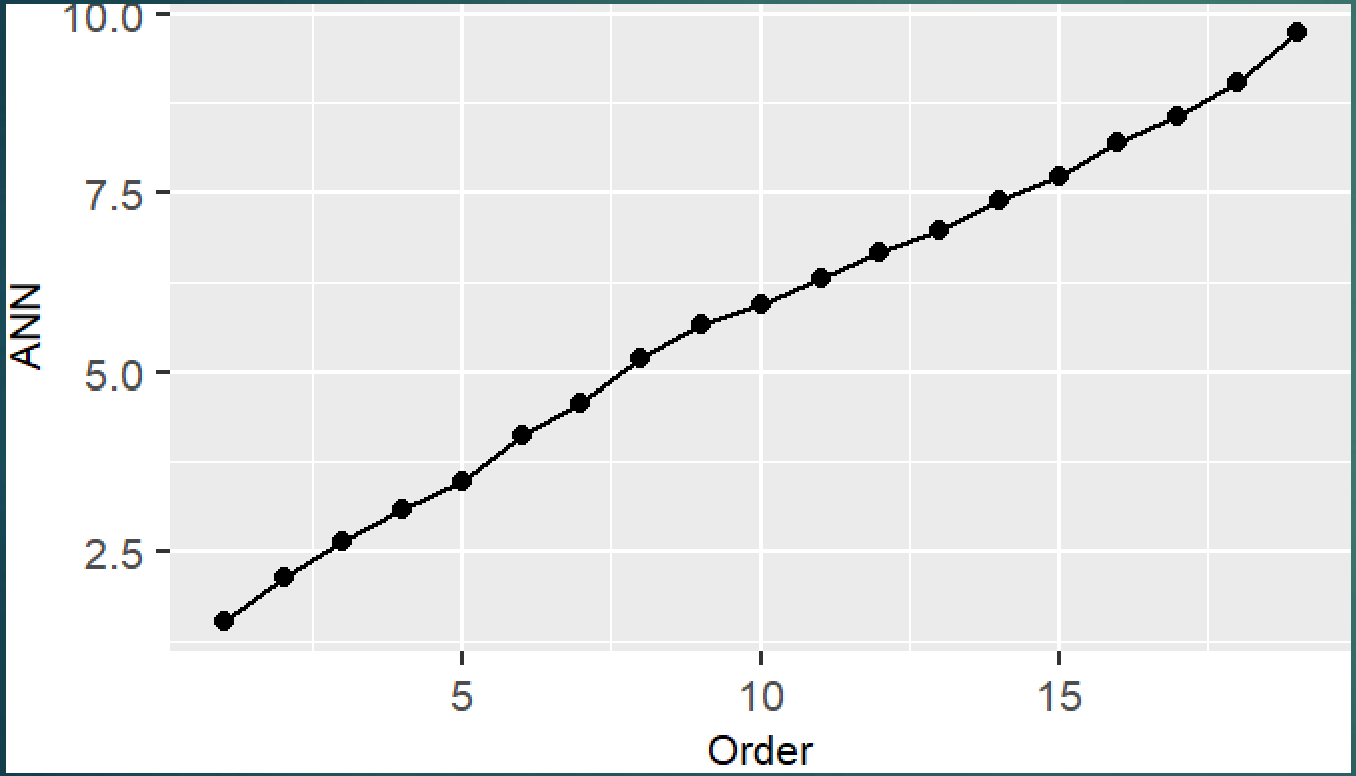
Distance between each point and its closest point.

For example, the point closest to point 1 is point 9 which is 2.32 map units away.

Image Source: Gimond. M, 2021

4.3. ANN Graph Plotting

Plotting the ANN values for different order neighbors, starting from first closest point, then the second closest point, and so forth.



The ANN for the first closest neighbor is 1.52 units; the ANN for the 2nd closest neighbor is 2.14 map units; and so forth.

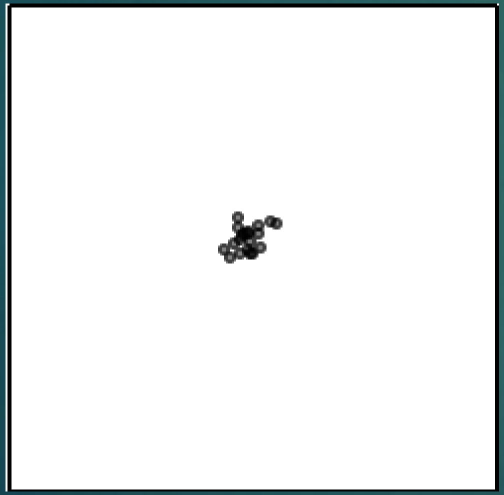
Figure Source: Gimond. M, 2021

4.4. Shape and Size of Study region

The shape of the ANN curve as a function of neighbor order can provide insight into the spatial arrangement of points relative to one another.

In the following example, three different point patterns of 20 points are presented.

Single Cluster



Dual Cluster

Randomly

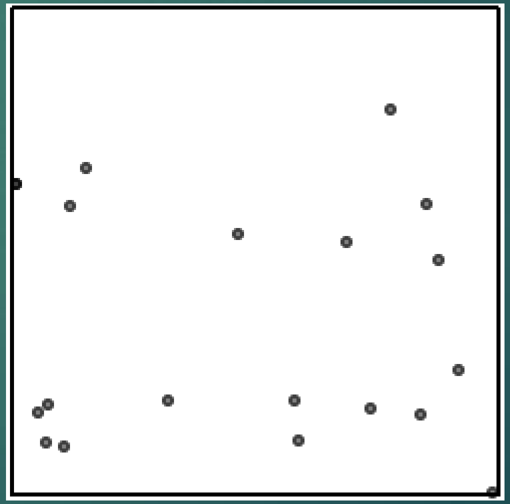
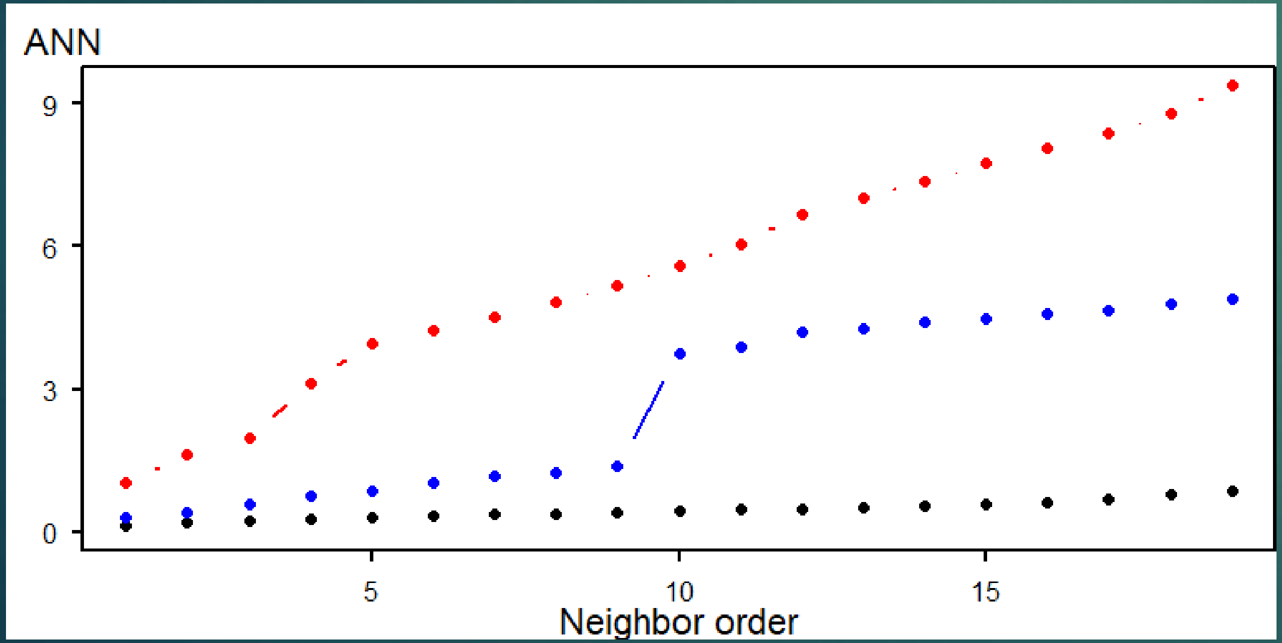


Image Source: Gimond. M, 2021

4.5. Cluster Patterns

ANN vs. neighbor order plots:

- The black ANN line is for single cluster
- Blue line is for double cluster
- Red line is for randomly scattered point pattern.



The way the patterns are describe is heavily influenced by the size and shape of the study region.

If the region was defined as the smallest rectangle encompassing the cluster of points, the cluster of points would no longer look clustered.

Figure Source: Gimond. M, 2021

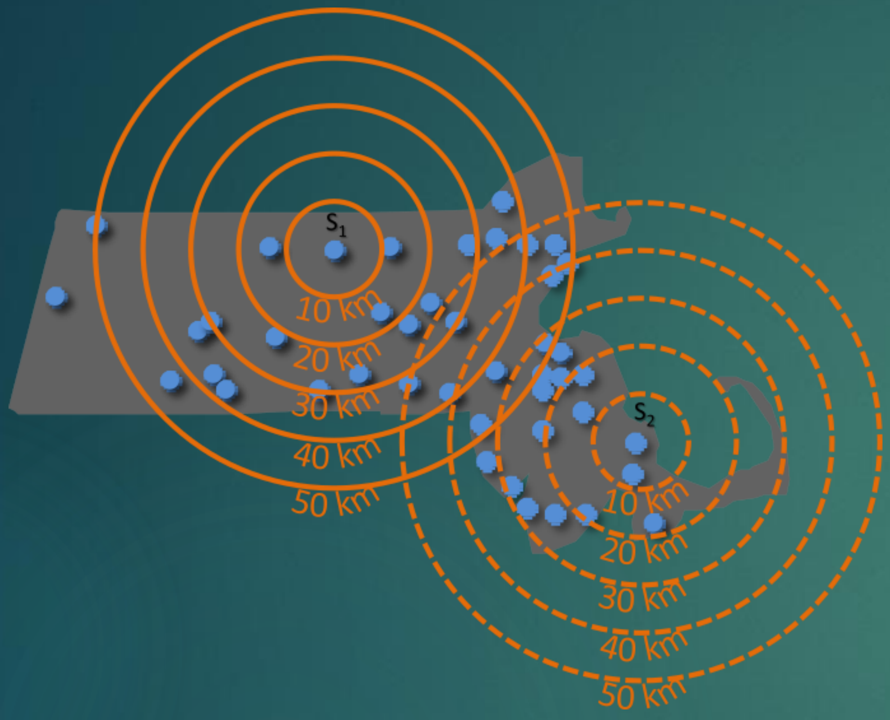
4.6. K Function

K-function summarizes the distance between points for *all* distances.

The calculation of K is fairly simple:

dividing the mean of the sum of the number of points at different distance lags for each point by the area event density.

4.7. K- Function Calculation



For point '*S1*' the circle is drawn with each of varying radius '*d*', centered on that point.

Then count the number of points (events) inside each circle.

Repeat this for point '*S2*' and all other points '*Si*'.

Next, compute the average number of points in each circle then divide that number by the overall point density 'λ' (i.e. total number of events per study area).

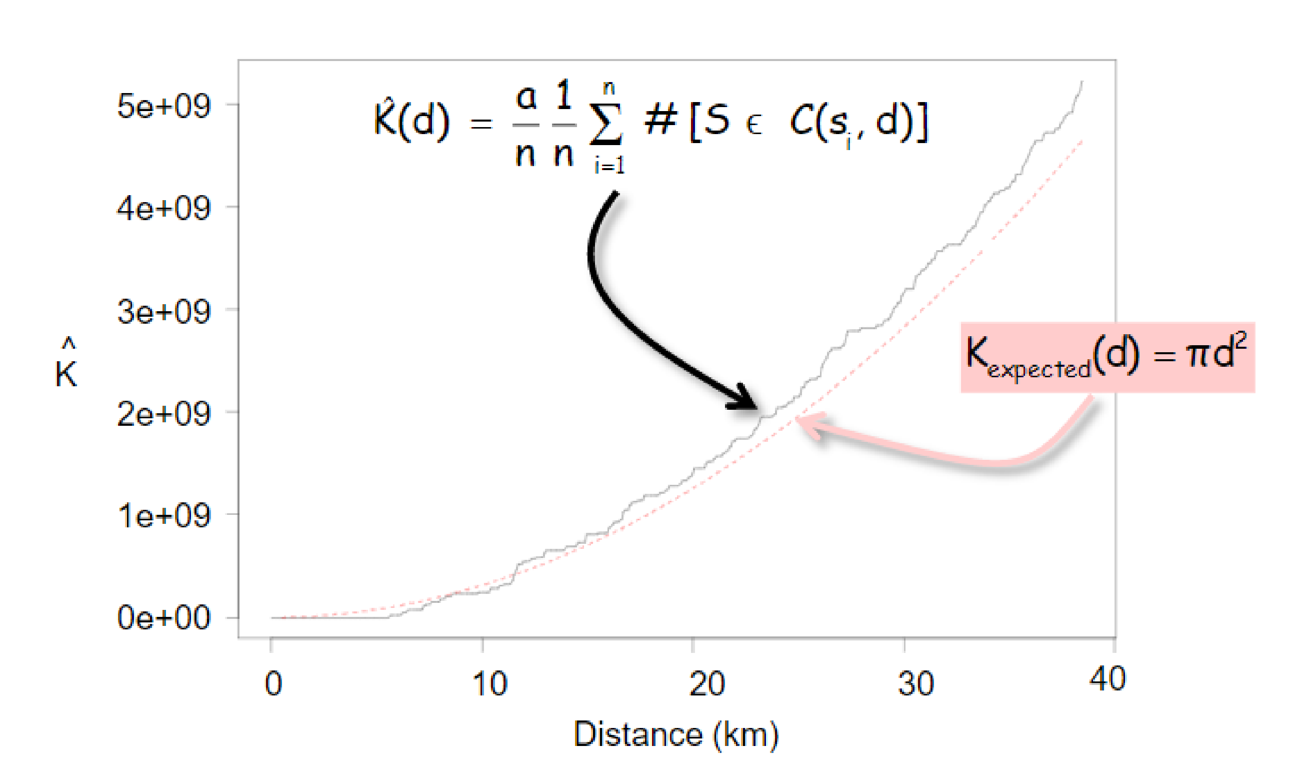
Distance Band (km)	No Events from S1	No Events from S2	No Events from Si	K
10	0	1	……………	0.012
20	3	5	…………..	0.067
30	9	14	………	0.153
40	17	17	………..	0.269
50	25	23	…………	0.419

map Source: Gimond. M, 2021

4.8. Calculated K vs Expected K

Calculated K is plotted and compare to a plot we would expect to get if an IRP/CSR process was at play ($K_{expected}$).

IRP - independent random process
CSR - complete spatial randomness



K values greater than *Kexpected* indicate clustering of points at a given distance band;

K values less than *Kexpected* indicate dispersion of points at a given distance band.

In our example, the stores appear to be more clustered than expected at distances greater than 12 km.

Figure Source: Gimond. M, 2021

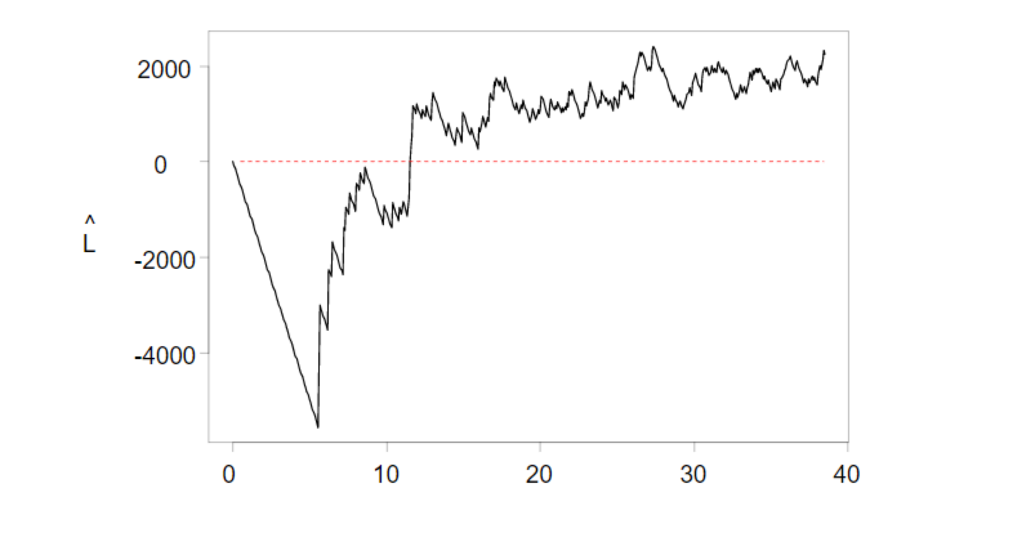
4.9. L - Function

L-function - a simple transformation of the K-function.

One problem with the *K* function is that the shape of the function tends to curve upward making it difficult to see small differences between *K* and $K_{expected}$. A workaround is to transform the values in such a way that the expected values, $K_{expected}$, lie horizontal.

The transformation is calculated as follows:

$$L = \sqrt{\frac{K(d)}{\pi}} - d$$



The *K* computed earlier is transformed to the following plot.

The $K_{expected}$ red line is now perfectly horizontal

Figure Source: Gimond. M, 2021

4.10. Simplified K-Function by L-Function

The graph makes it easier to compare K with $K_{expected}$ at lower distance values.

Values greater than '0' indicate clustering, while values less than '0' indicate dispersion.

It appears that store locations are more dispersed than expected under CSR/IRP up to a distance of 12 km but more clustered at distances greater than 12 km.

Source: Gimond. M, 2021

5. Application of PPA

PPA has applications in a wide range of areas:

- astronomy,
- archaeology,
- geography,
- ecology,
- biology,
- epidemiology.
- Crime analysis

6. References

Gimond, M (2021). Intro to GIS and Spatial Analysis (Book). Retrieve from: https://mgimond.github.io/Spatial/chp11_0.html

Pebesma, E., and Bivand, R., (2022). Spatial Data Science with Application in R (Book). Retrieve from: https://keen-swartz-3146c4.netlify.app/pointpatterns.html

Yuan, Y., Qiang, Y., Bin Asad, K., and Chow, T. E. (2020). Point Pattern Analysis. *The Geographic Information Science & Technology Body of Knowledge* (1st Quarter 2020 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2020.1.13.(link is external)

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Point pattern analysis. *The Geographic Information Science & Technology Body of Knowledge*. Washington, DC: Association of American Geographers. (2nd Quarter 2016, first digital).