4i – A Guide to
Nearest Neighbour Analysis

Nearest Neighbour Analysis measures the spread or distribution of something over a geographical space. It provides a numerical value that describes the extent to which a set of points are clustered or uniformly spaced.

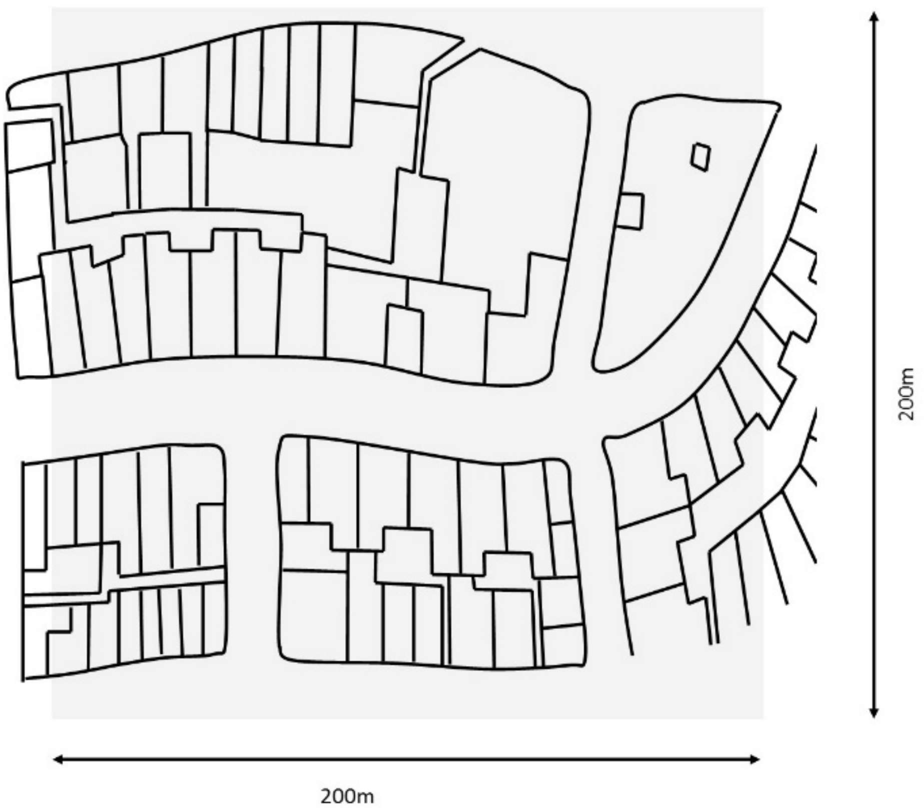
Why would we use nearest neighbour analysis?

Researchers use nearest neighbour analysis to determine whether the frequency with which something is observed spatially is comparable with other locations. It can provide researchers with a numerical value for the 'clustering' of a geographical phenomenon, allowing this value to be compared more accurately with other places.

Worked Example:

In this example, researchers have mapped the land use of each building in a 200m by 200m plot in a town centre, using a Goad map. A Goad map is a detailed street map that shows individual buildings and their plots, and is usually needed for land registry and insurance purposes. They can be purchased through online sources.

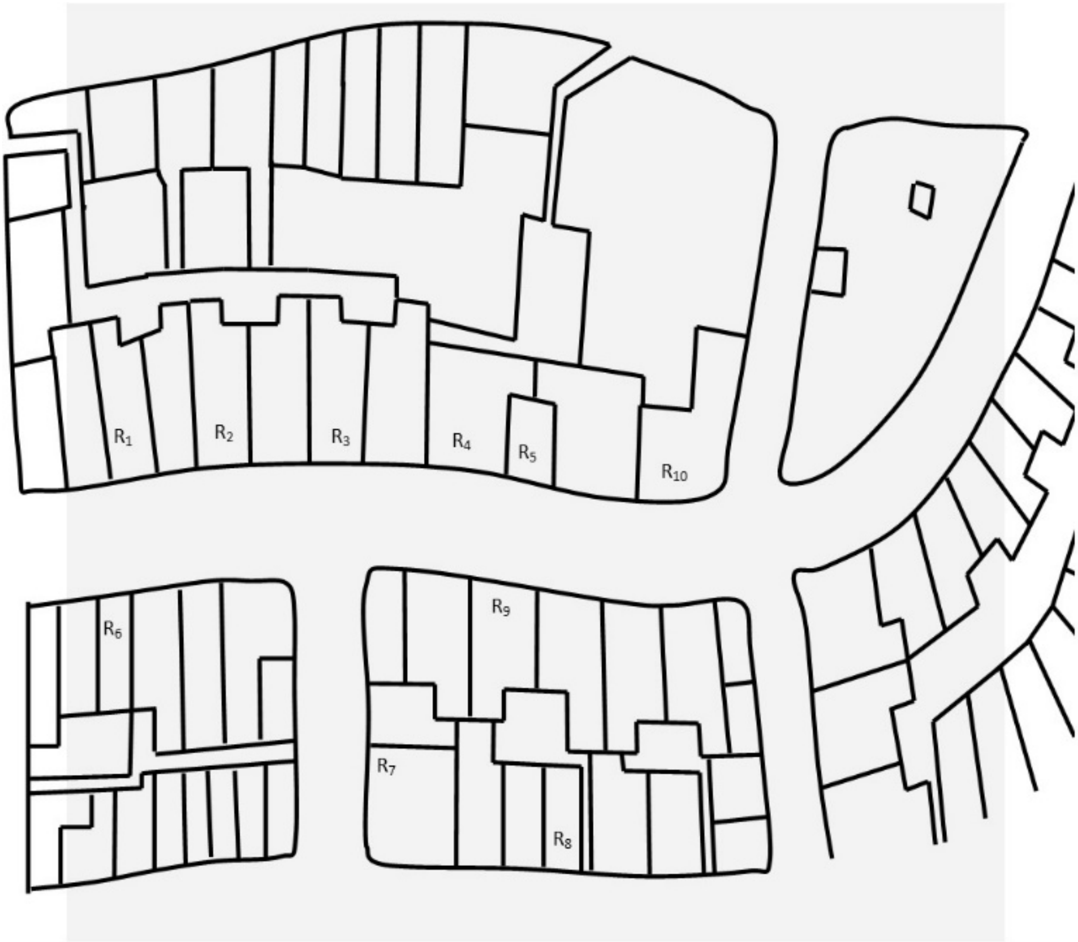
The nearest neighbour analysis can be used to identify whether there are clustered land use zones within that section of the town.



This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

The buildings have been categorised by their primary use: Retail; Financial; Municipal; Private Other; Residential (Housing), and Green Space.

1. In each category, the buildings are coded and numbered. For example, in the retail category the buildings are coded R_1, R_2, R_3 etc.



2. For each building in each category, the distance to the next nearest building in that same category is measured. Where measurements are taken between different points of the map, as in this case, the researcher must ensure there is a high degree of consistency, by always measuring from the centre of a feature or in the case of retail outlets, from the centre of the main entrance doorway. This ensures that large buildings or areas are subjected to exactly the same treatment as smaller ones. For ease in this example, only the first ten buildings of the retail category will be analysed, but to cover the full land use of the town, every category should be done in full.

Building	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
Distance to nearest neighbour (m)	8.2	4.5	6.9	9.6	9.5	12.3	23.4	18.5	5.0	12.9

3. The following calculation is used to work out the nearest neighbour value (R_n)

$$R_n = \frac{D(Obs)}{0.5\sqrt{\frac{a}{n}}}$$

where

$D(Obs)$ = the mean value of the nearest neighbour distances
a = the area sampled
n = the number of points (in this case retail buildings)

$$R_n = \frac{11.08}{0.5\sqrt{\frac{40000}{10}}} \qquad R_n = \frac{11.08}{0.5\sqrt{4000}} \qquad R_n = \frac{11.08}{0.5 \times 63.25}$$

$$R_n = \frac{11.08}{31.62} \qquad \boxed{R_n = 0.35}$$

4. Using the following scale, the researcher can then identify the extent of clustering exhibited by the category in question.

0 ——— 1 ——— 2.15

Clustered — Random — Uniform distribution

Therefore, the retail sector appears to be very clustered in the example town. However, the ten points used in the example are too few to really show an accurate picture: the researcher should aim for at least thirty to be able to say they have a significant result.