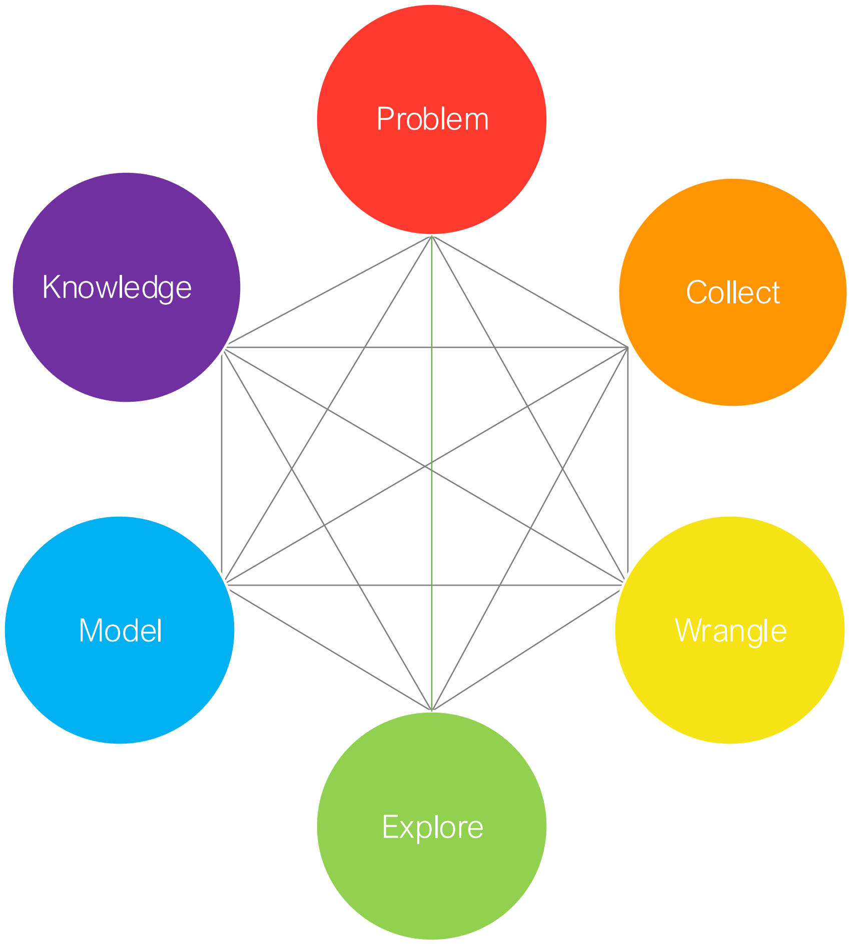
GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

WEEK 3: SPATIAL AUTOCORRELATION & DEPENDENCE

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

Contents

1. To understand the concepts of spatial dependence, and key characteristics of spatial data i.e., **spatial heterogeneity** and **spatial autocorrelation**
2. To understand the concept and importance of a **Spatial Weight Matrix (W)**
   - What are spatial weight matrices and how they are created
   - Contiguity-based weights
   - Distance-based weights
3. How do we quantify, measure and interpret the degree spatial dependence "globally" and "locally"?
   - Global Moran's I statistic
   - Local Indicators of Spatial Association (LISA)
4. Tips on making statistical inference & interpretation of the results churned from methods highlighted in Point 4



Quick Recap

We discussed the various types and ways of representing spatial data

In Week 2:

1. **Types of spatial data formats e.g., vector and raster**
2. **Spatial operations (aka Geoprocessing)**
    - Clipping
    - Aggregating
    - Union
    - Intersection
    - Buffering etc.,
3. **Thematic mapping and problems to consider**
    - Choropleths
    - Proportional mapping
    - Raster (our main focus for Week 4-6)

But in Week 1:

1. **We spoke of Tobler's first law of Geography**
    - Spatial dependence, Distance decay, Spatial spillovers

Let's rewind a bit to last week, and to the week before last

Formally, what is Spatial Dependance?

Definition [1]

Recall Tobler's 1st Law of Geography

Spatial dependence is the propensity for how nearby objects in geographic space tend to influence each other. It's a reflection of how values observed at one location (e.g., city, region, country) is dependent on the values of neighbouring observations from nearby locations.

In this context, there are two overarching characteristics of spatial data:

- Spatial autocorrelation
- Spatial heterogeneity

Definition [2]

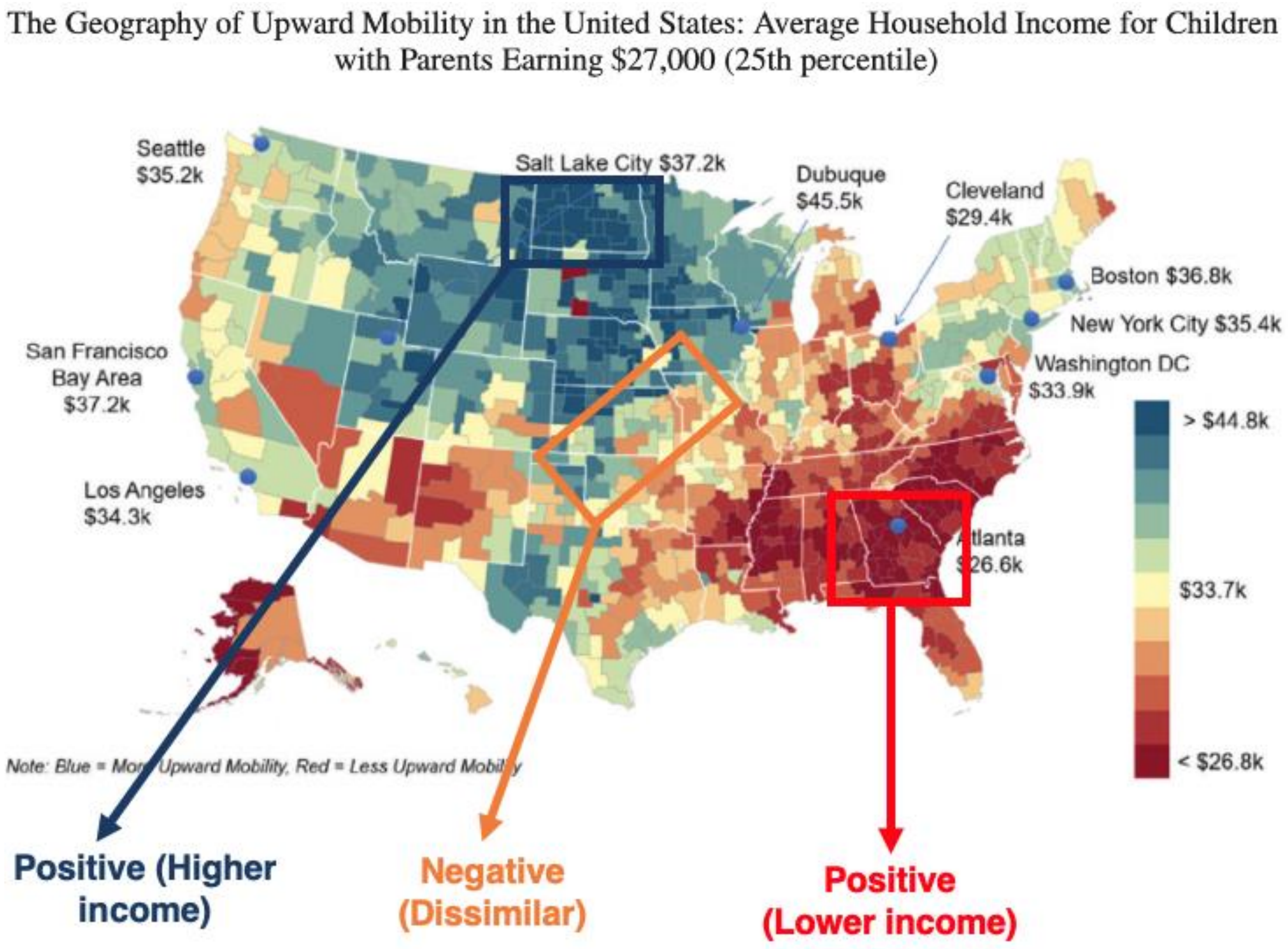
Recall Tobler's 1st Law of Geography

Spatial dependence is the propensity for how nearby objects in geographic space tend to influence each other. It's a reflection of how values observed at one location (e.g., city, region, country) is dependent on the values of neighbouring observations from nearby locations.

- **Spatial autocorrelation**

This describes the degree of how spatial locations (i.e., points, areas, or raster cells) **close** to each other share similar values (i.e., locations that are akin to each other).

An illustrative example of spatial autocorrelation



Notes: We concern ourselves with areas that are concentrated, or forms clusters of a specific value.

Here, we can see concentrations, or clusters of income in the SE region, high income in the Northern region etc.

Definition [3]

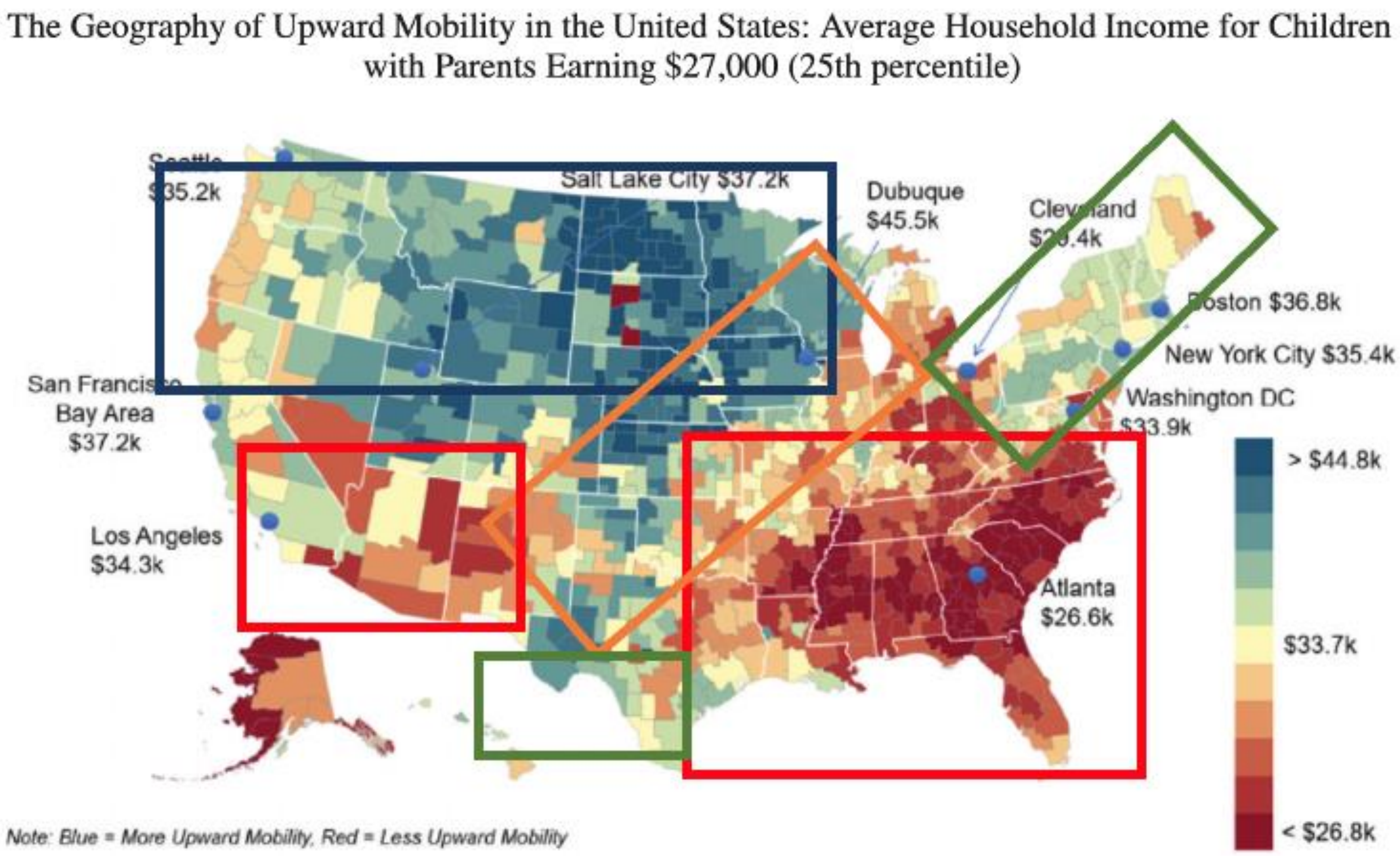
Recall Tobler's 1st Law of Geography

Spatial dependence is the propensity for how nearby objects in geographic space tend to influence each other. It's a reflection of how values observed at one location (e.g., city, region, country) is dependent on the values of neighbouring observations from nearby locations.

- **Spatial heterogeneity**

This essentially describes the uneven distribution of observed values **across the entire** geographical space.

An illustrative example of spatial heterogeneity



Notes: Variation in earnings across the US landscape is spatially heterogeneous – it changes across or "horizontally" US' landscape.

Why is Spatial Dependence important? [1]

1. Spatial dependence exists for both incident (event) and attribute data that are spatially-referenced (i.e., point, area or raster format)
2. It is a concept that allows the user to determine if event or attribute spatial data (i.e., some spatial phenomenon) exhibit spatial patterns such as **clustering**, **random distribution** or **dispersion**.
3. From point 2 concerning patterns of **random distribution**. Note that in most cases, the distribution of attribute values will seldomly show evidence of **Complete Spatial Randomness (CSR)**

Complete Spatial Randomness (CSR) essentially means that a pattern completely made up by chance. This is an important for understanding **spatial dependence** and significance in a spatial statistical context, especially when gauging evidence of spatial dependence, as it will always form our “null hypothesis”, while the ”alternative hypothesis” is patterns (i.e., clustering or dispersion) are not by chance.

4. This concept is a core element incorporated into several spatial analytical techniques

Examples of Spatial Analytical techniques that needs to incorporate this aspect of spatial dependence

For point data (use the spatial configuration of point locations that make-up the study area)

1. Geostatistical models such as **Kriging** through usage of **semivariogram analysis,** and others such as **Inverse Distance Weight (IDW).** These analyses are a function of distance.
2. Point process analysis with spatial **Bayesian models such as INLA-SPDE** through a **MESH / projection matrix (**which is grid-based and uses distance as a function).

For areal data (use the spatial configuration of areas that make-up the study area)

1. Spatial autocorrelation, we use Moran's I Statistics
2. Spatial clustering
3. Almost all spatial regression models (i.e., Spatial lags & errors, GWRs, Spatial Bayesian risk models)

What do we mean by spatial configuration of study area?

There has to be some formal mathematical way to explicitly represent or translate geography into some kind of numerical language, or device. We called this device a **Spatial Weight Matrix.**

Spatial Weight Matrices

Definition

A **Spatial Weight Matrix** is a mathematical structure or device used to represent geographical space. It can also be used to indicate the spatial relationship among special features within a defined geographic space.

W looks something like this

$$W = \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix}$$

The key characteristics of a spatial weight matrix:

- It is a square matrix of $n \times n$ size
- The entire square matrix is represented as $\boldsymbol{W}$
- The spatial relationship between location are represented as $\boldsymbol{w_{ij}}$ which indicates the degree of spatial relations between the observation seen in i-location and other observations seen j-locations

Note: The construction and usage of W in a spatial context is quite situational

Maths 101: Matrices

A **Matrix** is a set of number arranged in rows and columns to a rectangular array

Data structure with demographic and health information

	Age	BMI	Cancer
Row 1	26	30	0
Row 2	34	26	1
Row 3	74	34	1

Cancer: Yes = 1
Cancer: No = 0

It contains
- 3 columns (j)
- 3 rows (i)

$$\boldsymbol{A} = \begin{pmatrix} 26 & 30 & 0 \\ 34 & 26 & 1 \\ 74 & 34 & 1 \end{pmatrix}$$

- The information forms a matrix of $n \times n$ size, i.e., same number of columns and rows $[3 \times 3]$
- This information is represented as a square matrix with notation defined as $\boldsymbol{A}$
- $\boldsymbol{a_{ij}}$ a number from A.

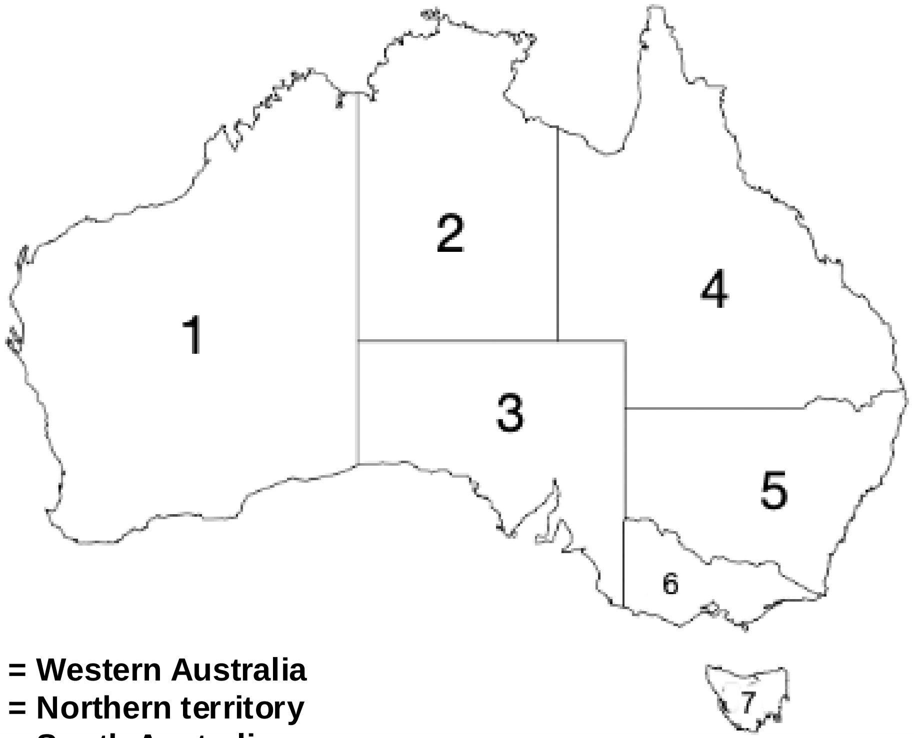
  Example 1: $\boldsymbol{a_{12}}$ (number on row 1 & column 2) = 30

  Example 1: $\boldsymbol{a_{31}}$ (number on row 3 & column 1) = 74

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [1]

Remember: The construction and usage of W in a spatial context is quite situational

Situation 1: W represents a neighbouring matrix with binary values (0, 1) only, w_{ij} = 1 if j-location is a neighbour of location i



1 = Western Australia
2 = Northern territory
3 = South Australia
4 = Queensland
5 = New South Wales & ACT
6 = Victoria
7 = Tasmania

Table showing areal adjacency

Index location (i)	Neighbours to index location (j)
1	2 and 3
2	1, 3, and 4
3	1, 2, 4, 5, and 6
4	2, 3, and 5
5	3, 4, and 6
6	3, 5, and 7
7	6

Converting this areal adjacency table to a spatial weight matrix

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [2]

Table showing areal adjacency

Index location (i)	Neighbours to index location (j)
1	2 and 3
2	1, 3, and 4
3	1, 2, 4, 5, and 6
4	2, 3, and 5
5	3, 4, and 6
6	3, 5, and 7
7	6

1 = Western Australia (WA)
2 = Northern territory (NT)
3 = South Australia (SA)
4 = Queensland (Q)
5 = New South Wales & ACT (NSA-ACT)
6 = Victoria (V)
7 = Tasmania (T)

W is a [7x7] square matrix

- $w_{11} = 0$: Cannot neighbour WA to itselt
- $w_{12} = 1$: WA is adjacent to NT
- $w_{13} = 1$: WA is adjacent to SA
- $w_{14} = 0$: WA not adjacent to Q
- $w_{15} = 0$: WA not adjacent to NSA
- $w_{16} = 0$: WA not adjacent to V
- $w_{17} = 0$: WA not adjacent to T

etc.,

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
W = & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\end{array}
$$

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [2]

Situation 1: This type of spatial matrix is **Contiguity-Based**

Table showing areal adjacency

Index location (i)	Neighbours to index location (j)
1	2 and 3
2	1, 3, and 4
3	1, 2, 4, 5, and 6
4	2, 3, and 5
5	3, 4, and 6
6	3, 5, and 7
7	6

1 = Western Australia (WA)
2 = Northern territory (NT)
3 = South Australia (SA)
4 = Queensland (Q)
5 = New South Wales & ACT (NSA-ACT)
6 = Victoria (V)
7 = Tasmania (T)

W is a [7x7] square matrix

- $w_{11} = 0$: Cannot neighbour WA to itself
- $w_{12} = 1$: WA is adjacent to NT
- $w_{13} = 1$: WA is adjacent to SA
- $w_{14} = 0$: WA not adjacent to Q
- $w_{15} = 0$: WA not adjacent to NSA
- $w_{16} = 0$: WA not adjacent to V
- $w_{17} = 0$: WA not adjacent to T

etc.,

$$
\begin{array}{cc}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
W = & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\end{array}
$$

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [3]

Remember: The construction and usage of W in a spatial context is quite situational

Situation 2: W represents a matrix with elements w_{ij} when its are based on distance d_{ij} between some location i and j. Here, we use point locations, or the centroids of such given area, to compute the distances whereby the coordinates for i is (x_i, y_i) and j is (x_j, y_j) are used. The goal here is to integrate distance decay.

Euclidean Distance:

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \quad [1]$$

Inverse Distance:

$$w_{ij} = \begin{cases} \dfrac{1}{d_{ij}^{\beta}} & \text{If } i \neq j \text{ (i.e., dealing with two different locations)} \\ 0 & \text{If } i = j \text{ (i.e., this basically the same point location)} \end{cases} \quad [2]$$

Note that β = 1 or β = 2 (it's up to you which ever value you pick!)

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [4]

Remember: The construction and usage of W in a spatial context is quite situational

Situation 2: W represents a matrix with elements w_{ij} when its are based on distance d_{ij} between some location i and j. Here, we use point locations, or the centroids of such given area, to compute the distances whereby the coordinates for i is (x_i, y_i) and j is (x_j, y_j) are used. The goal here is to integrate distance decay.

Euclidean Distance:

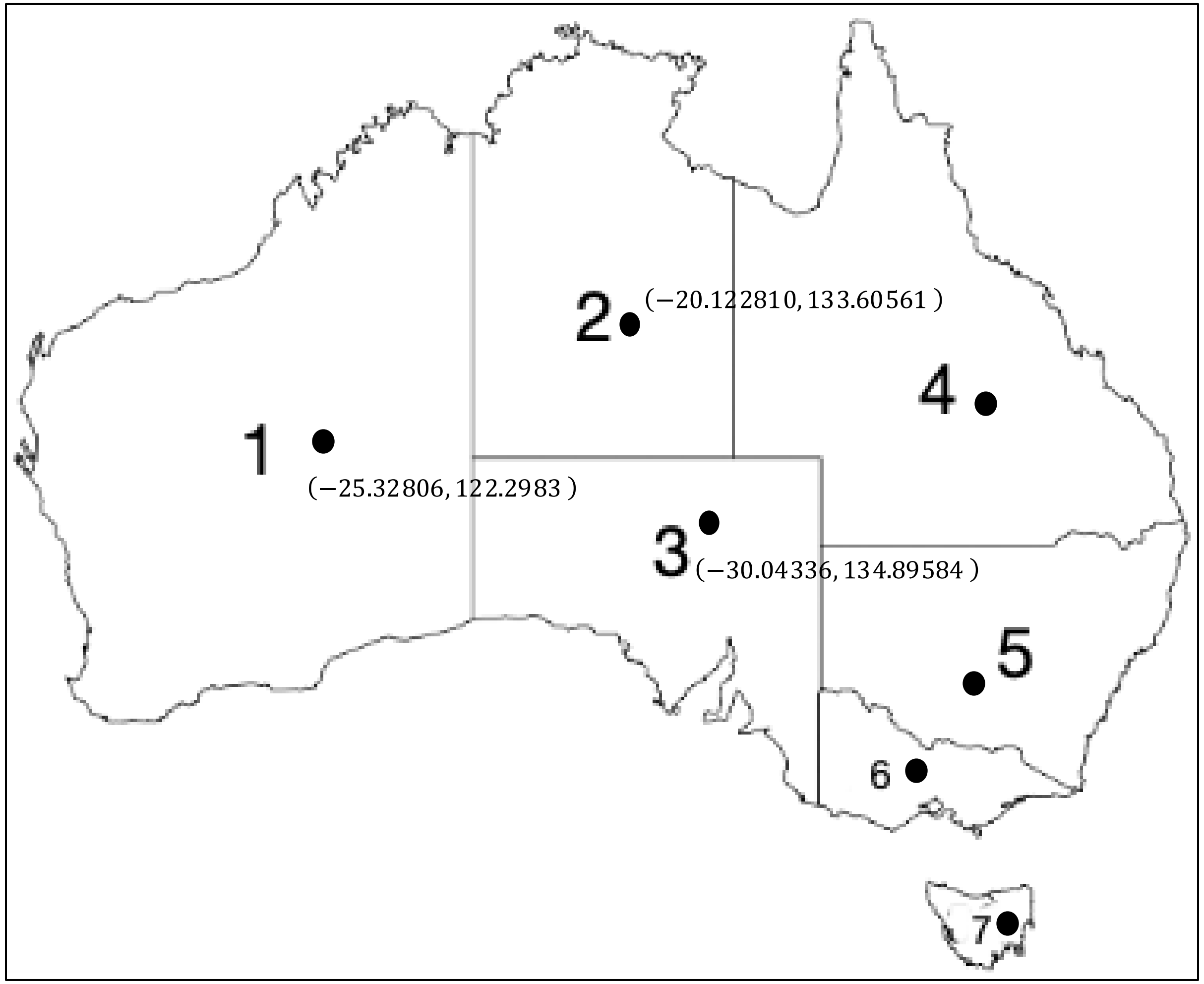
$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \textbf{[1]}$$

Negative exponential:

$$w_{ij} = \exp\left(-\frac{d_{ij}}{\beta}\right) \quad \textbf{[2]}$$

Note that β = 1 or β = 2 (it's up to you which ever value you pick!)

Example: Calculation of Euclidean Distance and $\boldsymbol{w_{ij}}$



1 = Western Australia (WA)
2 = Northern territory (NT)
3 = South Australia (SA)
4 = Queensland (Q)
5 = New South Wales & ACT (NSA-ACT)
6 = Victoria (V)
7 = Tasmania (T)

You must use formula [1] or [2] to compute distance between **ALL** possible pairs of points (or centroids) to each estimate for $\boldsymbol{d_{ij}}$, and then use $\boldsymbol{d_{ij}}$ to obtain each estimate for $\boldsymbol{w_{ij}}$ to build matrix $\boldsymbol{W}$

Here is an incredibly crude example using points WA and NT

Step 1: compute d_{ij}

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

$$\text{WA} = (x_i, y_i) = (x_1, y_1) = (-25.32806, 122.2983)$$
$$\text{NT} = (x_j, y_j) = (x_2, y_2) = (-20.122810, 133.60561)$$

$$d_{12} = \sqrt{(-25.32806 - (-20.122810))^2 + (122.2983 - 133.60561)^2}$$
$$d_{12} = 12.44789$$

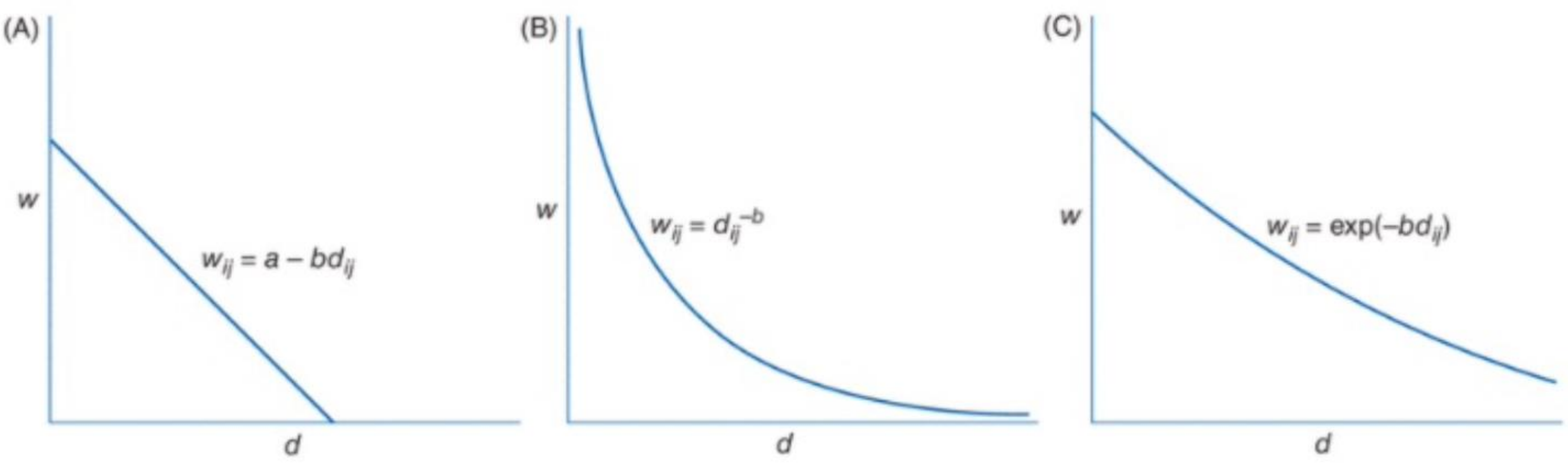
Step 2: compute w_{ij} using inverse distance

$$w_{12} = \frac{1}{d_{ij}^{\beta}} = \frac{1}{12.44789} = 0.0803349$$

Step 3: construct matrix $\boldsymbol{W}$

The matrix is a [7x7] row-column array. Insert value in 1st row 2nd column position

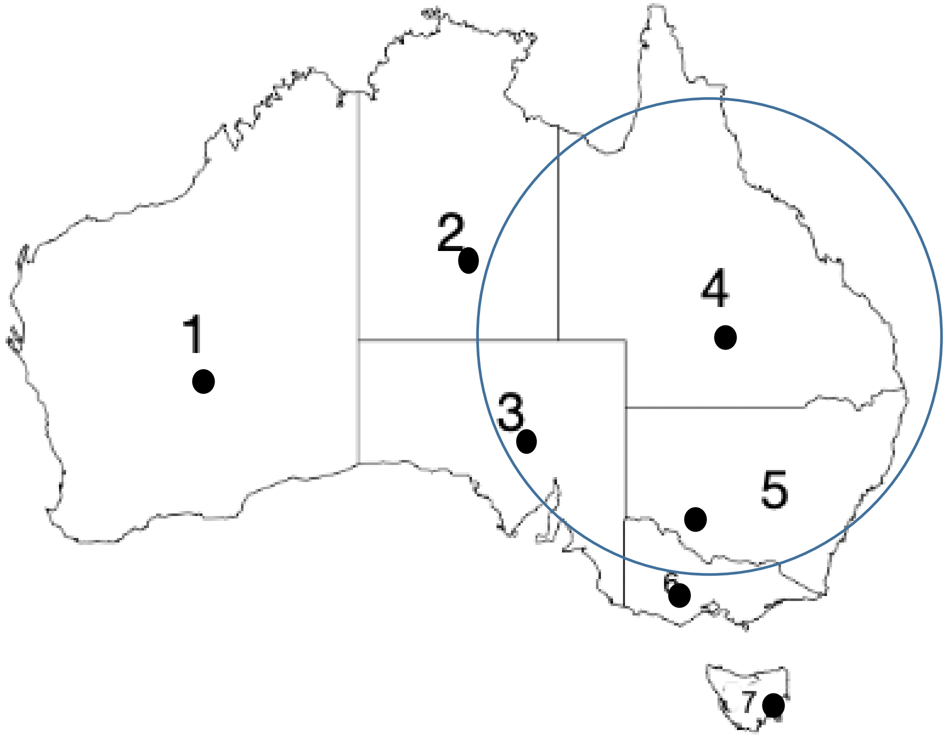
w_{ij} vs d_{ij} is a distance decay function



Linear **Inverse** **Negative exponential**

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [5]

- K-nearest neighbours based on distance: We explicitly limit the number of neighbours by specifying a threshold distance (or distance band weights) to also construct $\boldsymbol{W}$

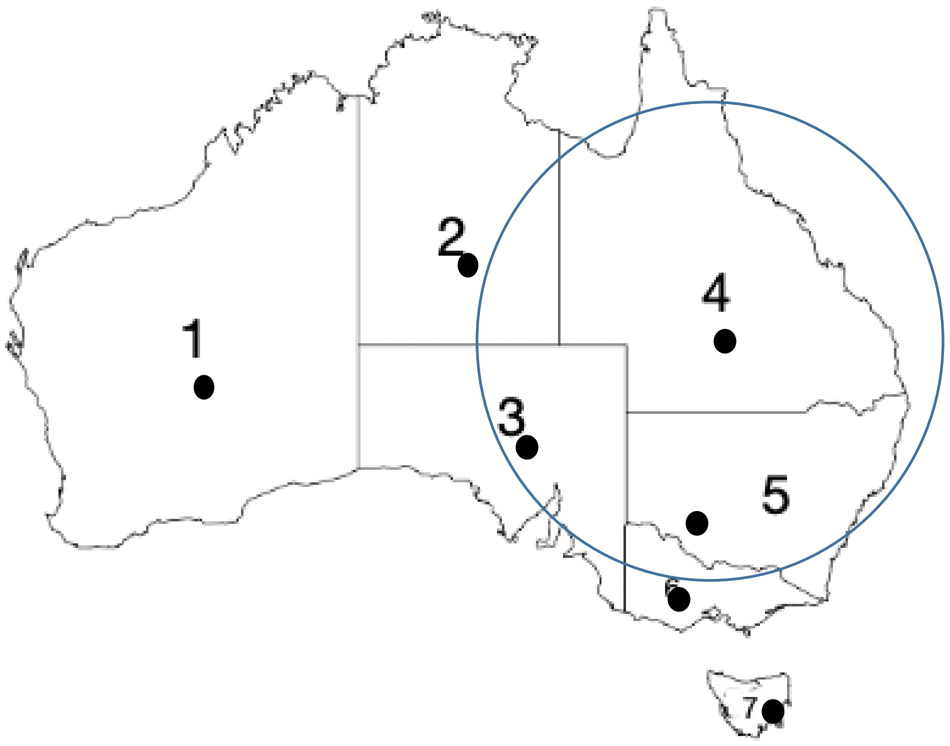


Index location (i)	Neighbours to index location (j) within distance band weight
1	none
2	3
3	2, 4, and 5
4	3 and 5
5	3, 4, 6, and 7
6	3, 5, and 7
7	5 and 6

Matrix $\boldsymbol{W}$ and $\boldsymbol{w_{ij}}$ [5]

Situation 2: These types of spatial matrices are **Distance-Based**

- K-nearest neighbours based on distance: We explicitly limit the number of neighbours by specifying a threshold distance (or distance band weights) to also construct $\boldsymbol{W}$

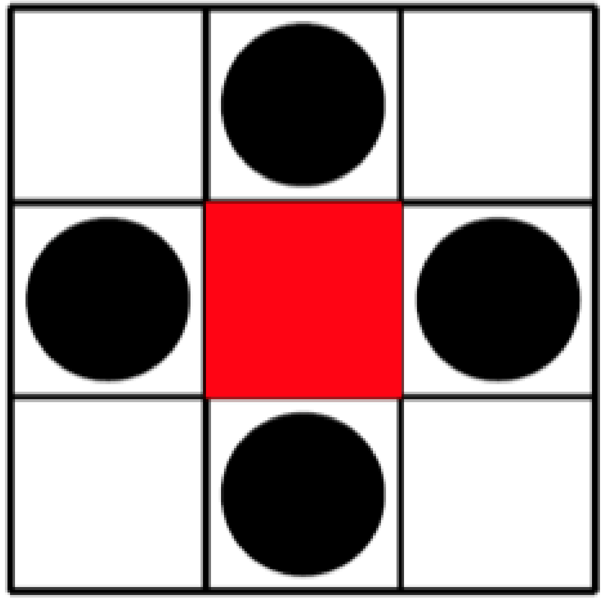


$$W = \begin{array}{cc} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} & \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \end{array}$$

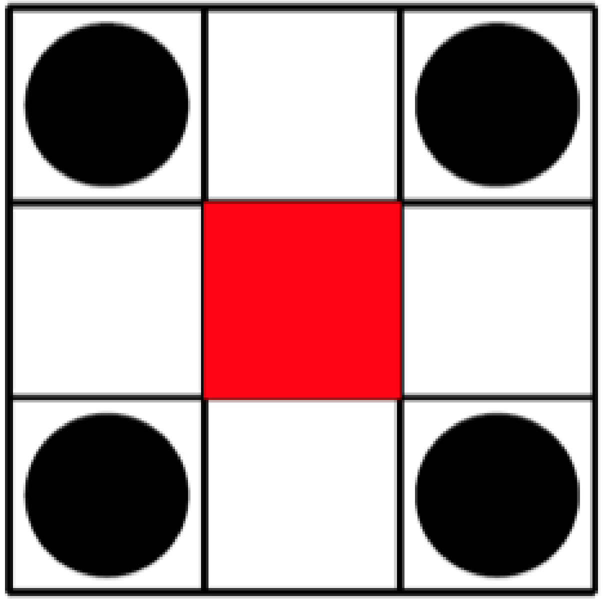
Contiguity-based weights [1]

Note: These are confined to area-based data

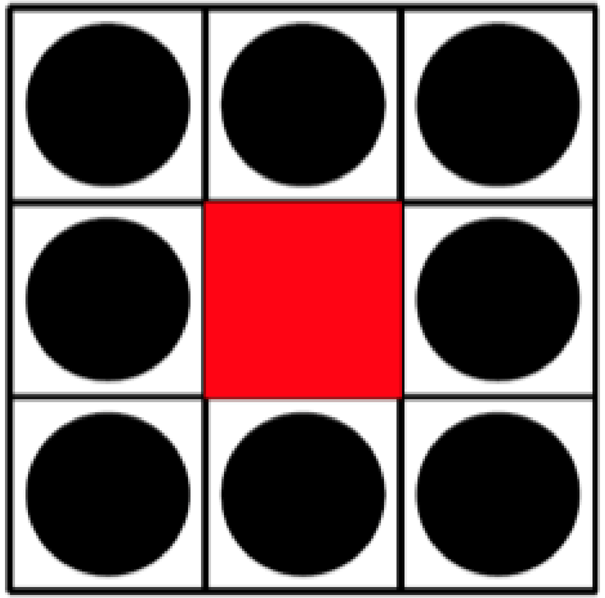
Sharing boundaries to any extent



Rook's Contiguity, these are neighbour locations that share a common border

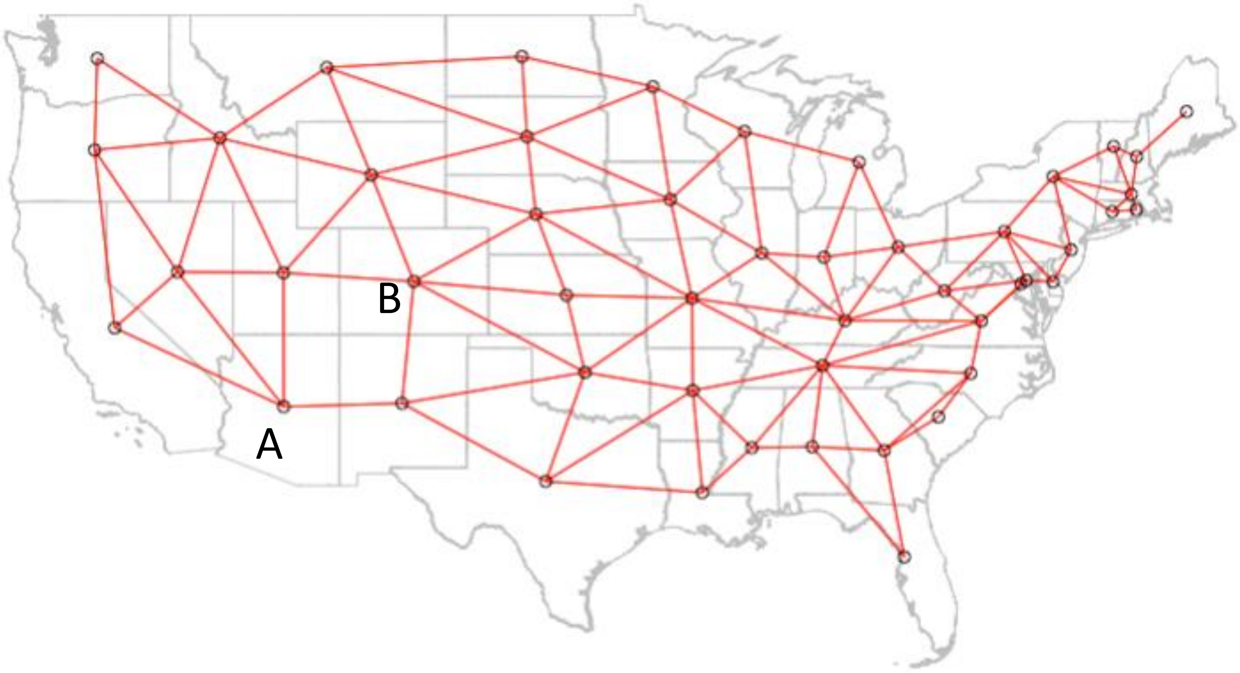


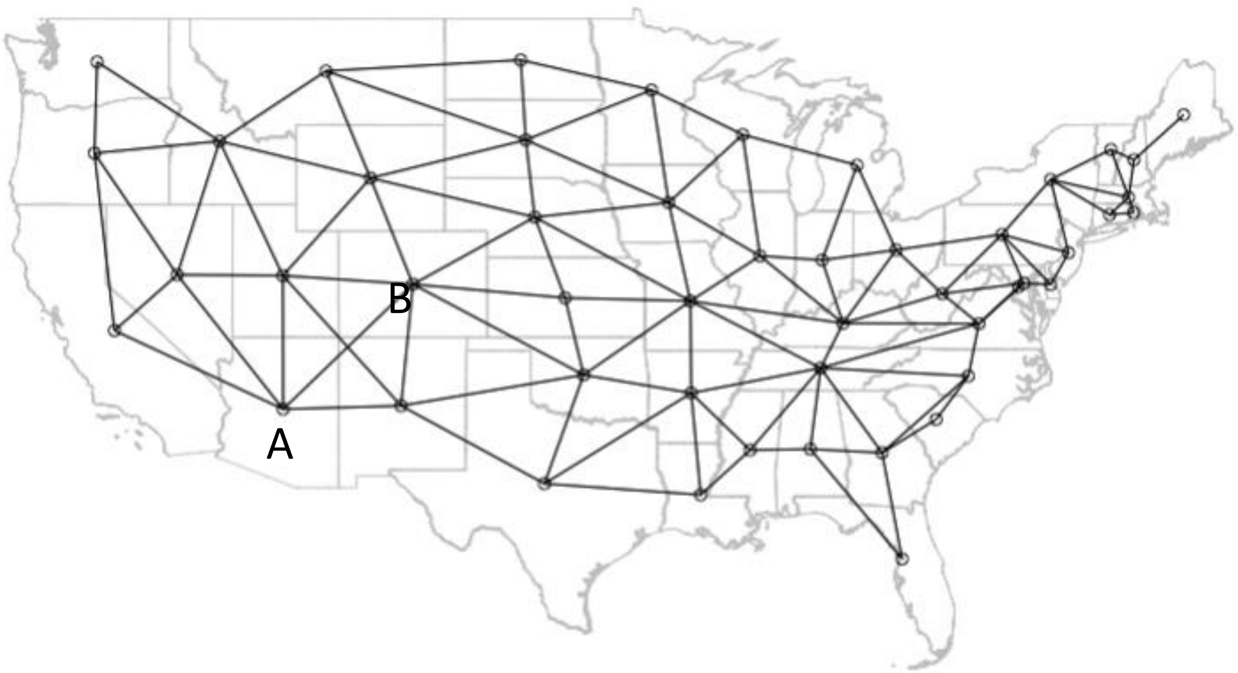
Bishops' Contiguity, refers to neighbour locations that ONLY share a single point along its border or boundary.



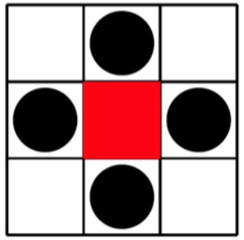
Queen's Contiguity is a combination of both the Rook and Bishop, where neighbour locations are allowed to share AT LEAST a single point along its border.

Contiguity-based weights [2]

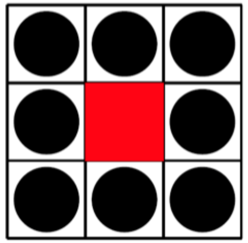




Rook’s contiguity



Queen’s contiguity



Spatial autocorrelation and statistical inference

Definition [1]

Spatial autocorrelation: This describes the degree of how spatial locations (i.e., points, areas, or raster cells) **close** to each other share similar values (i.e., locations that are akin to each other).

- We can test our data for spatial autocorrelation, with some form of statistical measure of the similarity of attributes of our data.
- We want to distinguish between areas of positively autocorrelated patterns (in which high values are surrounded by high values, and low values by low values, i.e., **clusters**);
- We want to test for random patterns (in which neighbouring values are independent of each other, i.e., CSR); and dispersed patterns (in which high values tend to be surrounded by low values and vice versa).

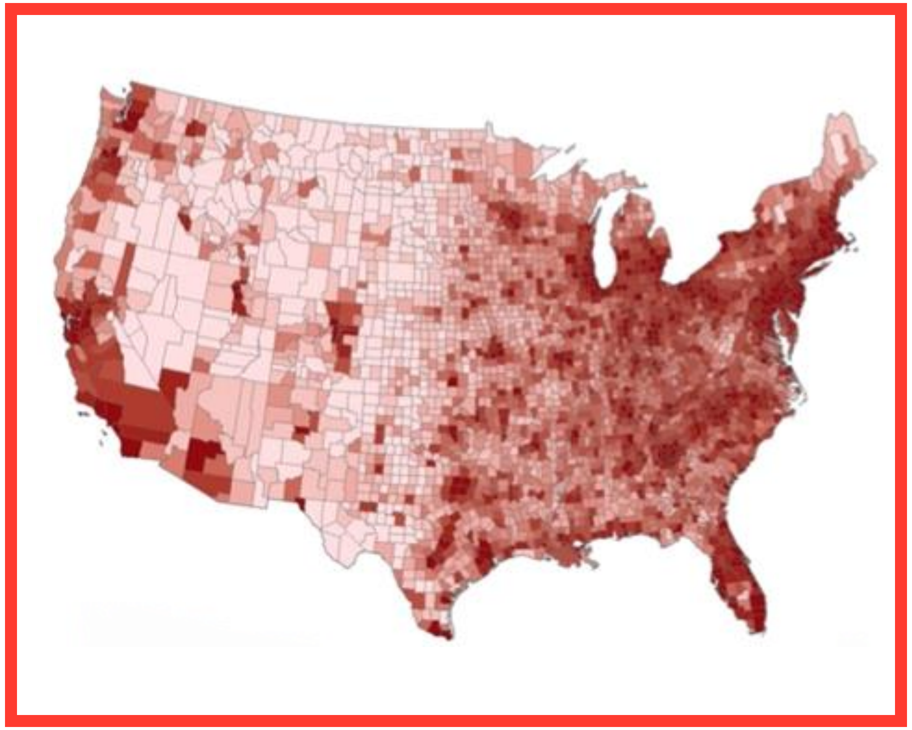
Definition [2]

Spatial autocorrelation: This describes the degree of how spatial locations (i.e., points, areas, or raster cells) **close** to each other share similar values (i.e., locations that are akin to each other).

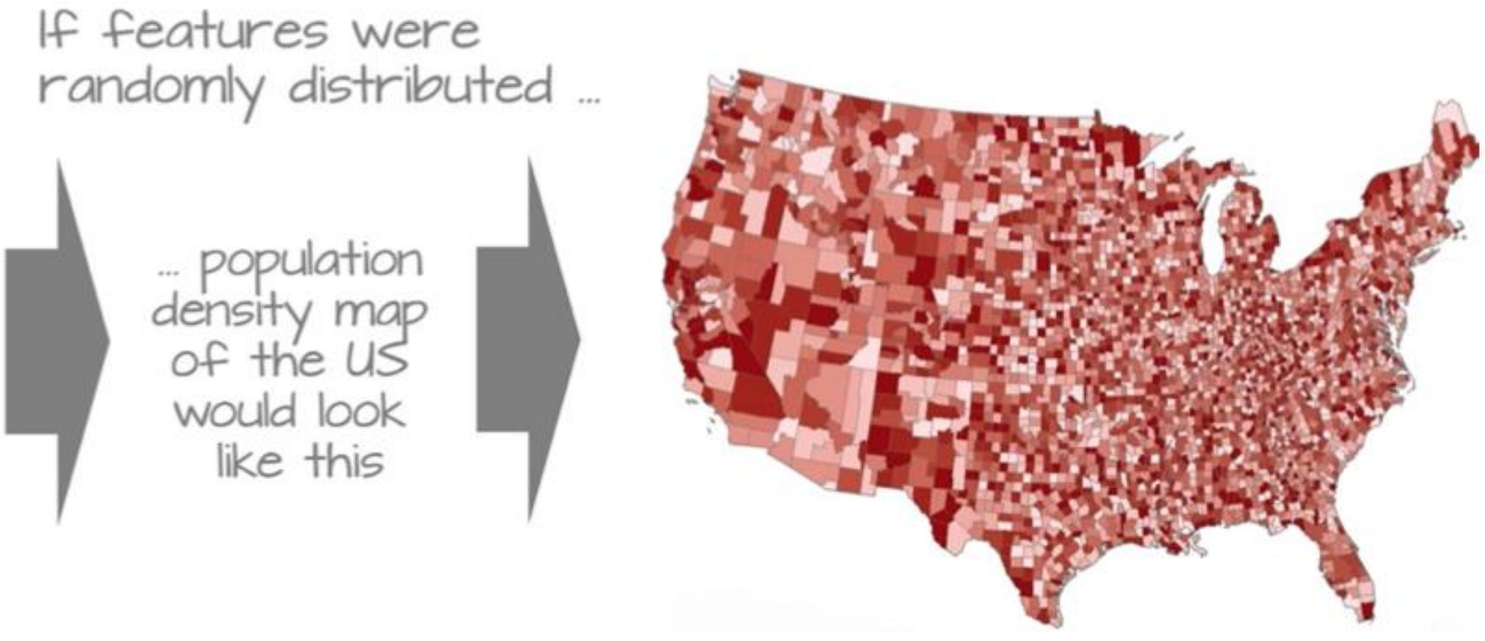
The hypothesis statement for testing evidence for spatial autocorrelation

- **Null hypothesis:** The outcome of interest are spatially independent (i.e., patterns are random)
- **Alternative hypothesis:** The outcome of interest are not spatially independent (i.e., hence, there is evidence of **clustering** or dispersion)

We can apply these hypotheses tests on these scenarios

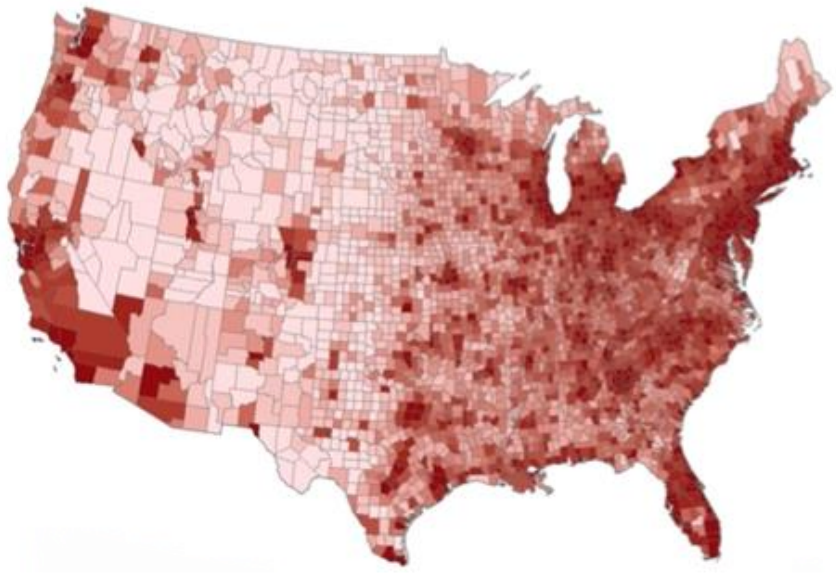


Scenario 1: US Population Density



- **Null hypothesis:** The spatial patterns for the US Population Density are independent. They are random. **[Here, we would reject the null hypothesis]**
- **Alternative hypothesis:** The patterns for the US Population Density are not random. They are indeed clustered. **[Here, we would accept the alternative hypothesis]**

We can apply these hypotheses tests on these scenarios

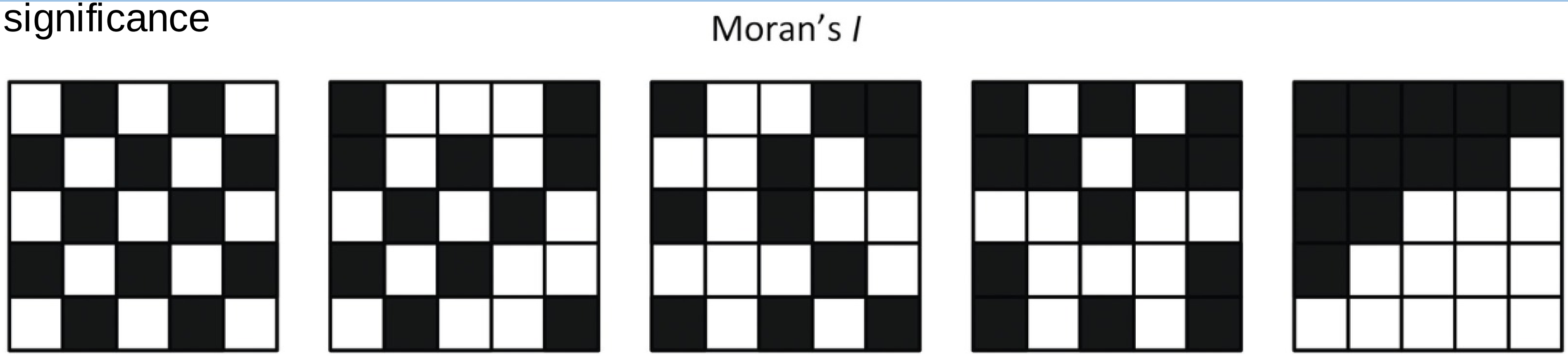


Scenario 2: US Population Density

- **Null hypothesis:** The spatial patterns for the US Population Density are independent. They are random. **[Here, we would accept the null hypothesis]**
- **Alternative hypothesis:** The patterns for the US Population Density are not random. They are indeed clustered. **[Here, we would reject the alternative hypothesis]**

Statistical analysis: Moran’s I statistic [1]

In testing for evidence of spatial autocorrelation, the **Moran’s I statistic,** which is a weighted correlation coefficient, is the statistical test used to detect departures from spatial randomness. Departures from randomness indicate spatial patterns such as clusters (+) or dispersed (-). This statistic is accompanied with a p-value to determine significance



Moran’s I statistic:	**Dispersed**	**Random**	**Clustered**
	-1	**0**	**+1**

- Positive Moran’s I value is when spatial autocorrelation generally indicates that nearby area have similar values, indicating spatial clustering. **This is a spatial pattern!**

- Negative Moran’s I spatial autocorrelation generally indicates that nearby area have dissimilar values, this is dispersion of values is indeed a **spatial pattern!**

- A Moran’s I value closer to 0 means no evidence of spatial autocorrelation. **No discernible spatial pattern!**

Statistical analysis: Moran’s I statistic [2]

In testing for evidence of spatial autocorrelation, the **Moran’s I statistic,** which is a weighted correlation coefficient, is the statistical test used to detect departures from spatial randomness. Departures from randomness indicate spatial patterns such as clusters (+) or dispersed (-). This statistic is accompanied with a p-value to determine significance

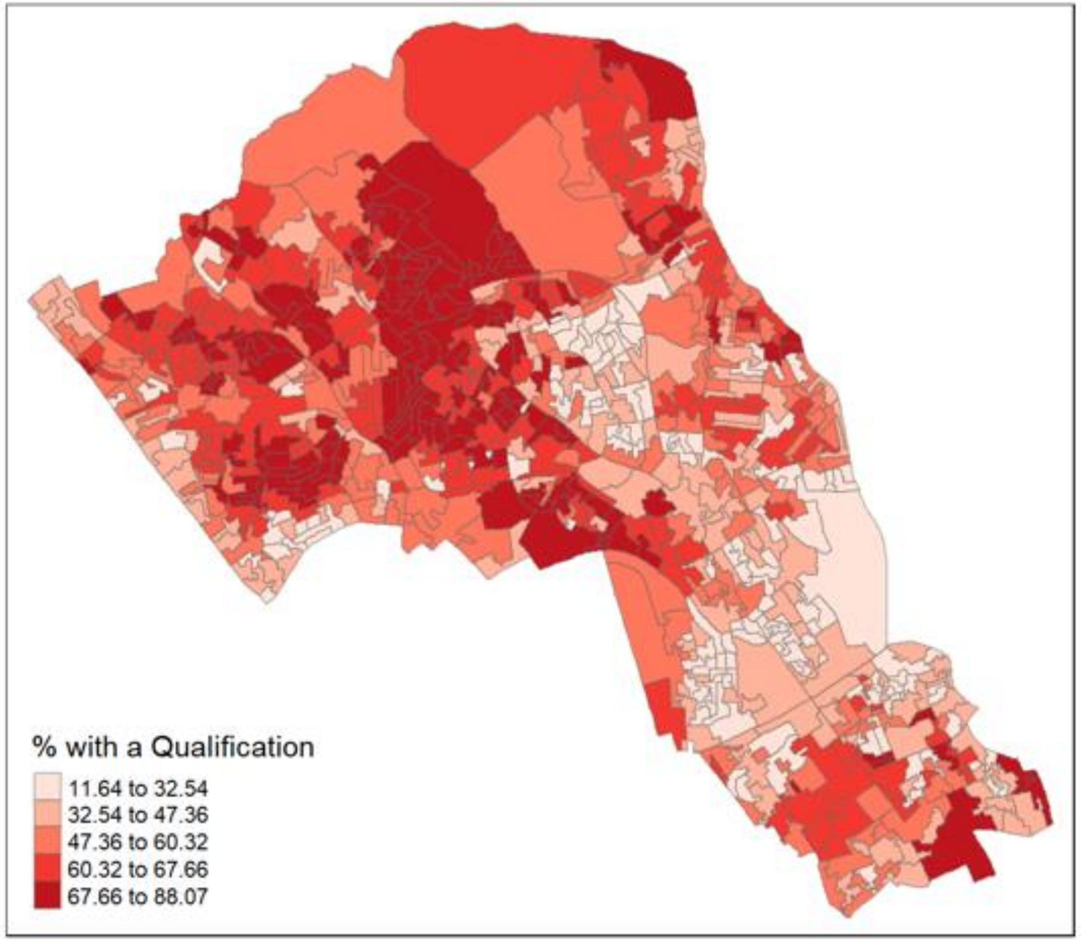
There are two ways for statistically measuring spatial autocorrelation:

- **Global Moran’s I statistic**: What is the overall spatial dependence across the entire data set area? Studying at a global level will tell you how clustered, dispersed or random the data is distributed over the entire area studied.

- **LISA (Local Indicators of Spatial Association)**: What is the difference between each unit of analysis (e.g., areal unit) and its neighbours? We use it for studying at the local level, you can find areas of greater contrast by seeing if places are quantifiably more like or dislike with their neighbours than expected on average.

If the p-value is less than 0.05, we can reject the null hypothesis in favour of the alternative hypothesis and arrive at the conclusion that the patterns are significantly clustered. If the p-value is above 0.05, it means that the clustering/dispersed patterns are not significant, thus we can conclude patterns are random.

Example: What is the spatial dependence in qualification in Camden, London?



Map shows the prevalence (%) of qualification or education attainment levels in postcodes across Camden

<u>Step one</u>: Ask, what is the overall spatial dependence in prevalence of qualification across the entire study area i.e., is it clustered, random or dispersed?

- **Generate the hypothesis**
  - Null hypothesis: The patterns are random
  - Alternative hypothesis: The patterns are not random (i.e., clustered/dispersed).

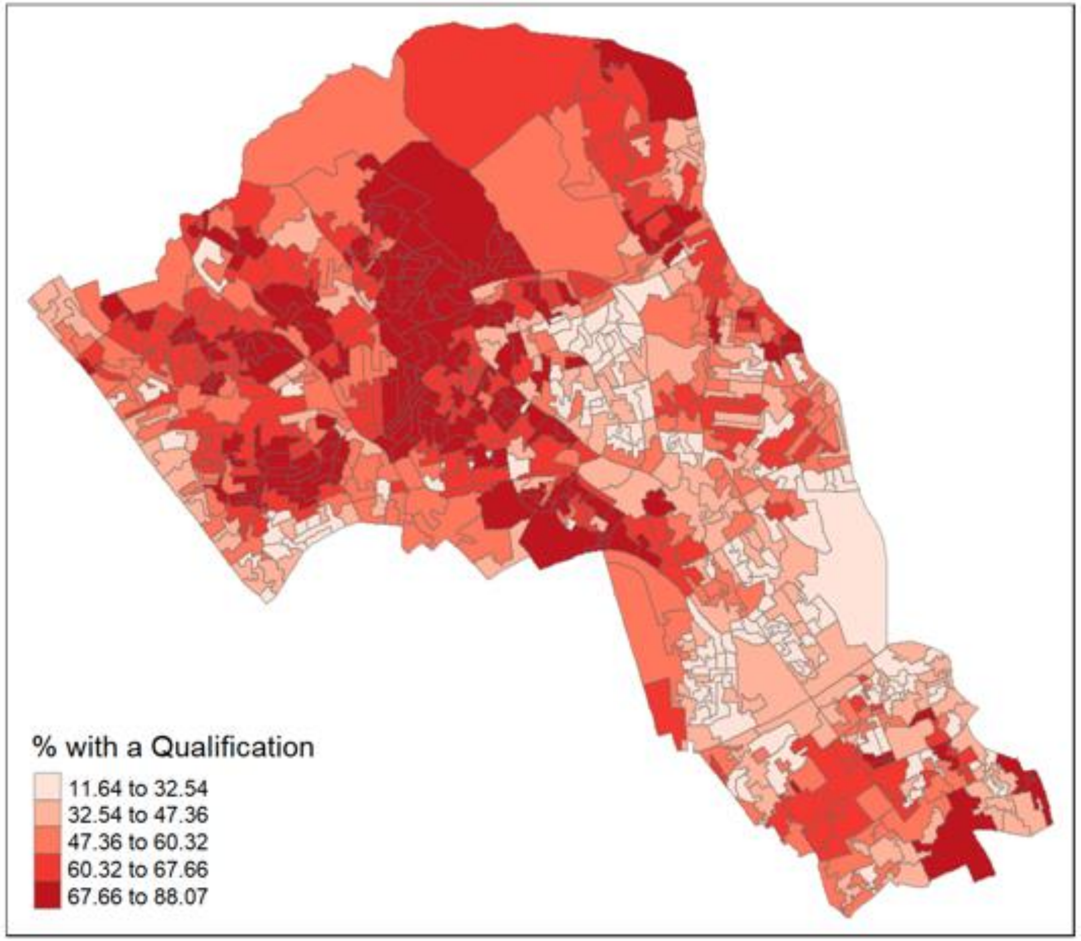
<u>Step two</u>: Perform Global Moran's I test and get a p-value to accept, or reject the null hypothesis

- **Result (Global Moran's I = 0.5448 [p=0.0001 <**

<u>Step three</u>: Interpretation of Global Moran's I test

- **We can reject the null hypothesis in favour of the alternative since the Moran's I statistic's p-value is less than 0.05. Therefore, The Moran I statistic is 0.54, we can, therefore, determine that there our qualification variable is significantly positively autocorrelated in Camden. In other words, the data does spatially cluster.**

Example: What is the spatial dependence in qualification in Camden, London?

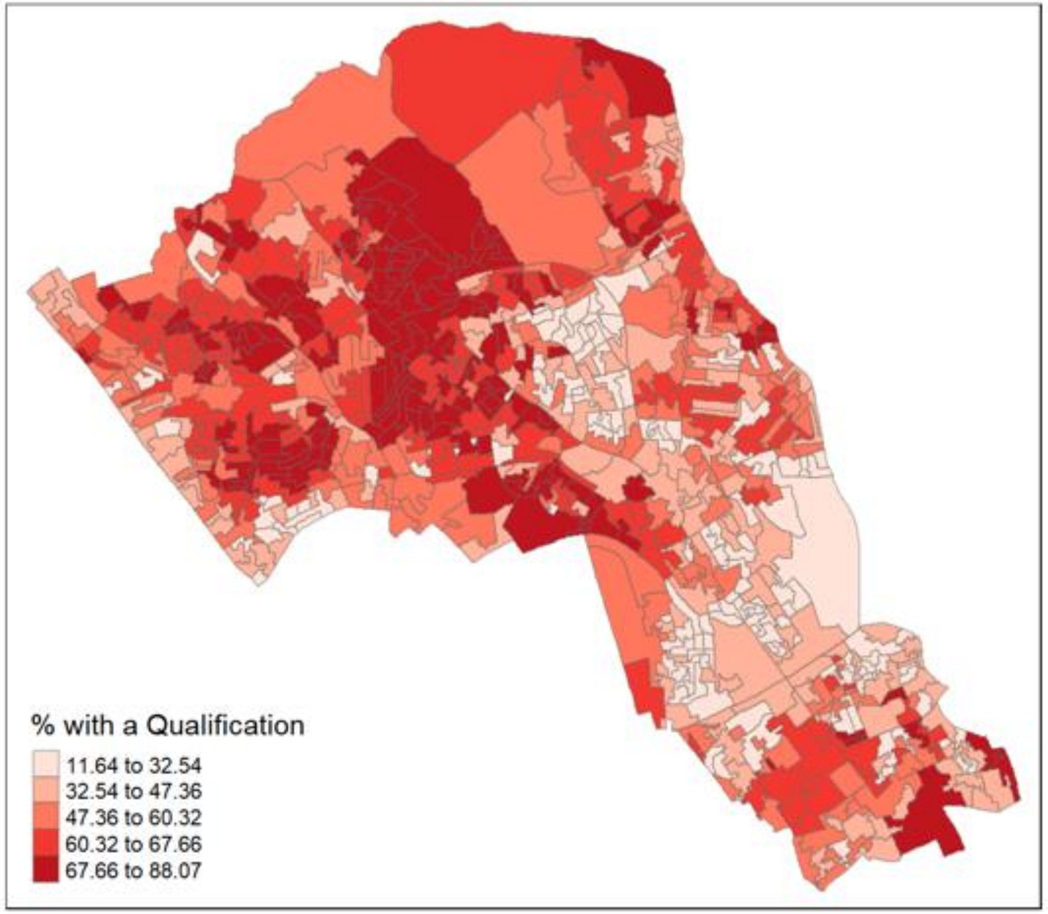


Map shows the prevalence (%) of qualification or education attainment levels in postcodes across Camden

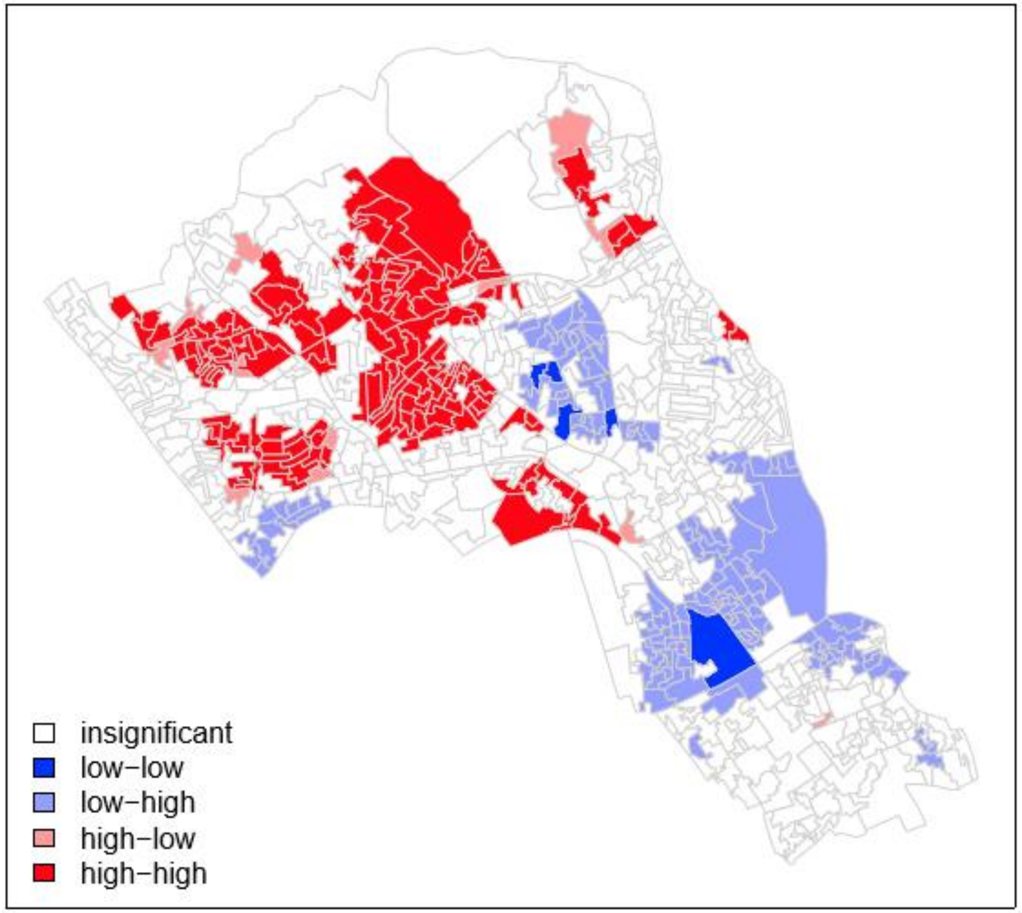
<u>Step four</u>: Ask, are there any place-to-place differences in the distribution of qualification holders across the entire study area i.e., hot or cold spots for clusters, and where are they exactly?

- Perform the Local Moran's I test and estimate the following:
    - Obtain the I values for each get each postcode, and use a divergent colour scheme to categorise them accordingly (negative values are cold spots, positive values are hot spots)
    - Obtain the I-values' corresponding p-values for each postcode, and use to determine if the hot/cold spots are significant (i.e., $p < 0.05$)
    - P-values are categorised accordingly in this order as 'Low-low', 'Low-high', 'Not significant', 'High-low' and 'High-high'
    - Plot a map of I-values and p-values.

Example: What is the spatial dependence in qualification in Camden, London?



Map shows the prevalence (%) of qualification or education attainment levels in postcodes across Camden



This is what we want. The p-values to know whether the LISA estimates are significant or not.

Summary

The take home message is:

1. Know the key characteristics of spatial dependence i.e., spatial heterogeneity and spatial autocorrelation

2. The key for most spatial analysis, especially for spatial autocorrelation, is the derivation of the Spatial Weight Matrix (W)

3. Know how to quantify, measure and interpret the degree spatial dependence "globally" and "locally" using the Global Moran's I statistic and Local Indicators of Spatial Association (LISA)

4. Know how to generate the hypothesis, as well as interpret the results churned by methods highlighted in Point 3

Any questions?