GEOG0114: PRINCIPLES OF SPATIAL ANALYSIS

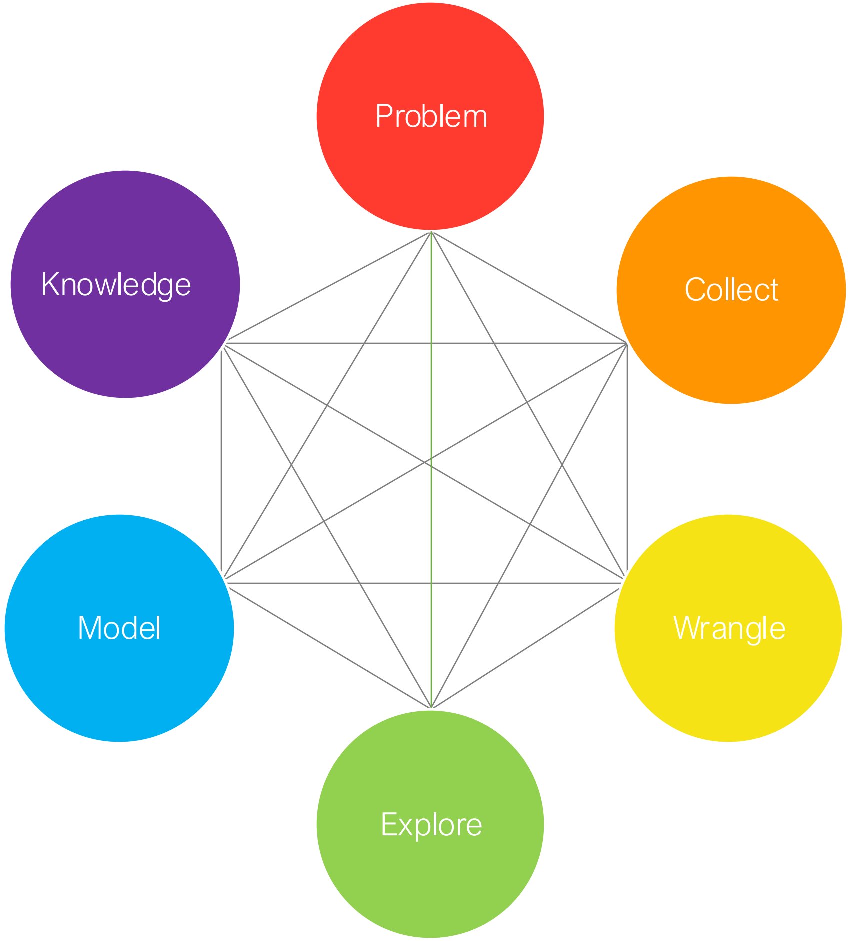
WEEK 2: GRAPHICAL REPRESENTATION OF SPATIAL DATA

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

Contents

1. **Types of spatial data formats used to describe reality**
   - Vector
   - Raster
2. **Spatial Operations (or Geoprocessing)**
3. **Thematic Mapping**
   - Choropleths
   - Proportional symbol map
   - Raster map
   - Isopleth / Contour maps

4. Issues
   - Ecological Fallacy bias
   - Absolute versus Relative
   - Resolution matters…



QUICK RECAP

1. Spatial statistics is grounded on **spatial dependence** and from **Tobler's 1st Law of Geography**

2. We mentioned some fundamental concepts, and two key words for explaining spatial dependence:
    - **Distance decay**
    - **Spatial autocorrelation**
    - **Spatial spillovers**

3. Event information presented spatially can either be:
    - **Areal/Regional (aka aggregated)**
    - **Point patterns**
    - **Geostatistical**

Let's rewind a bit to last week

Types of Spatial Data Formats

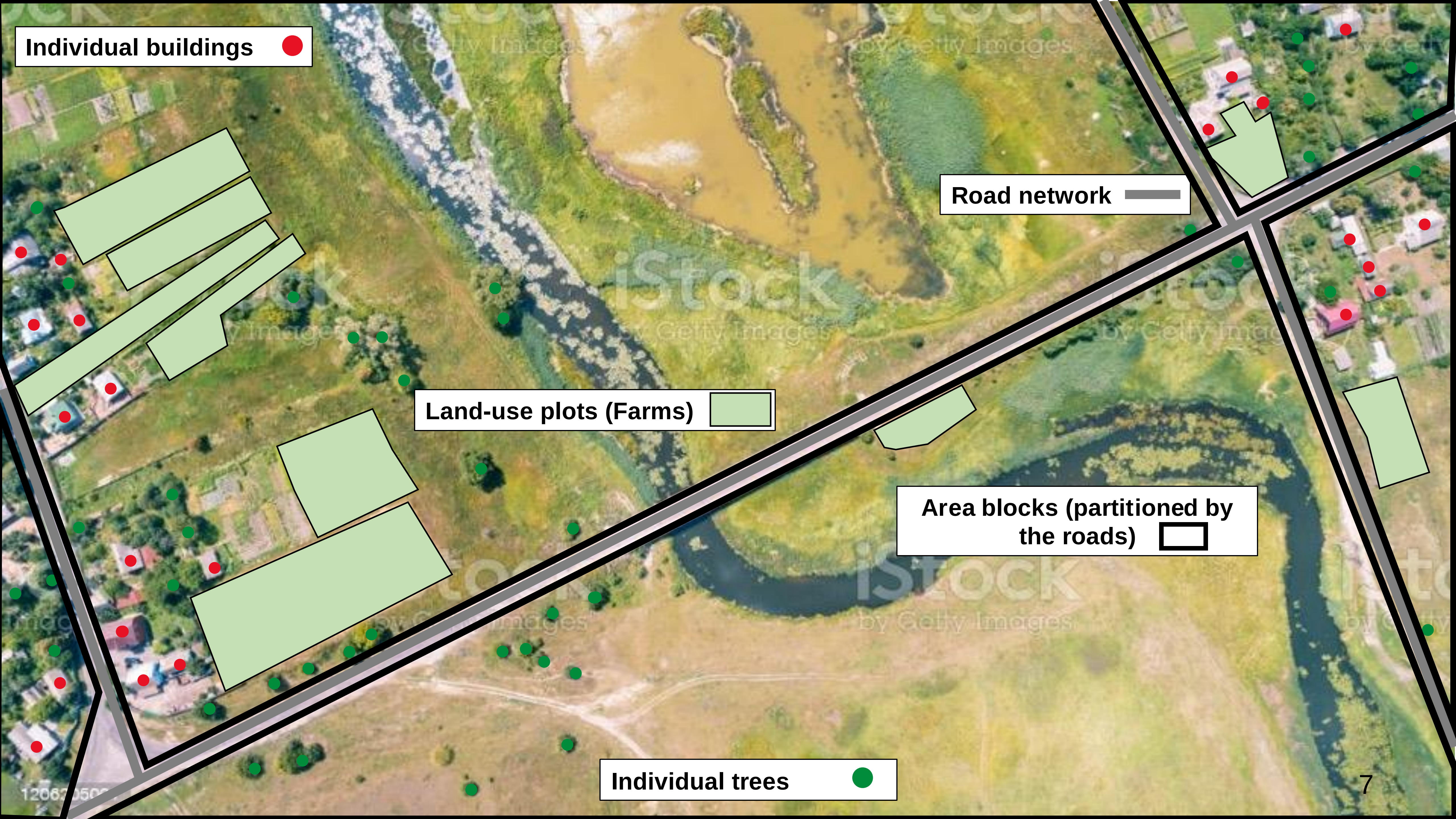
Suppose we want to map the following from this landscape:

1. Physical objects:

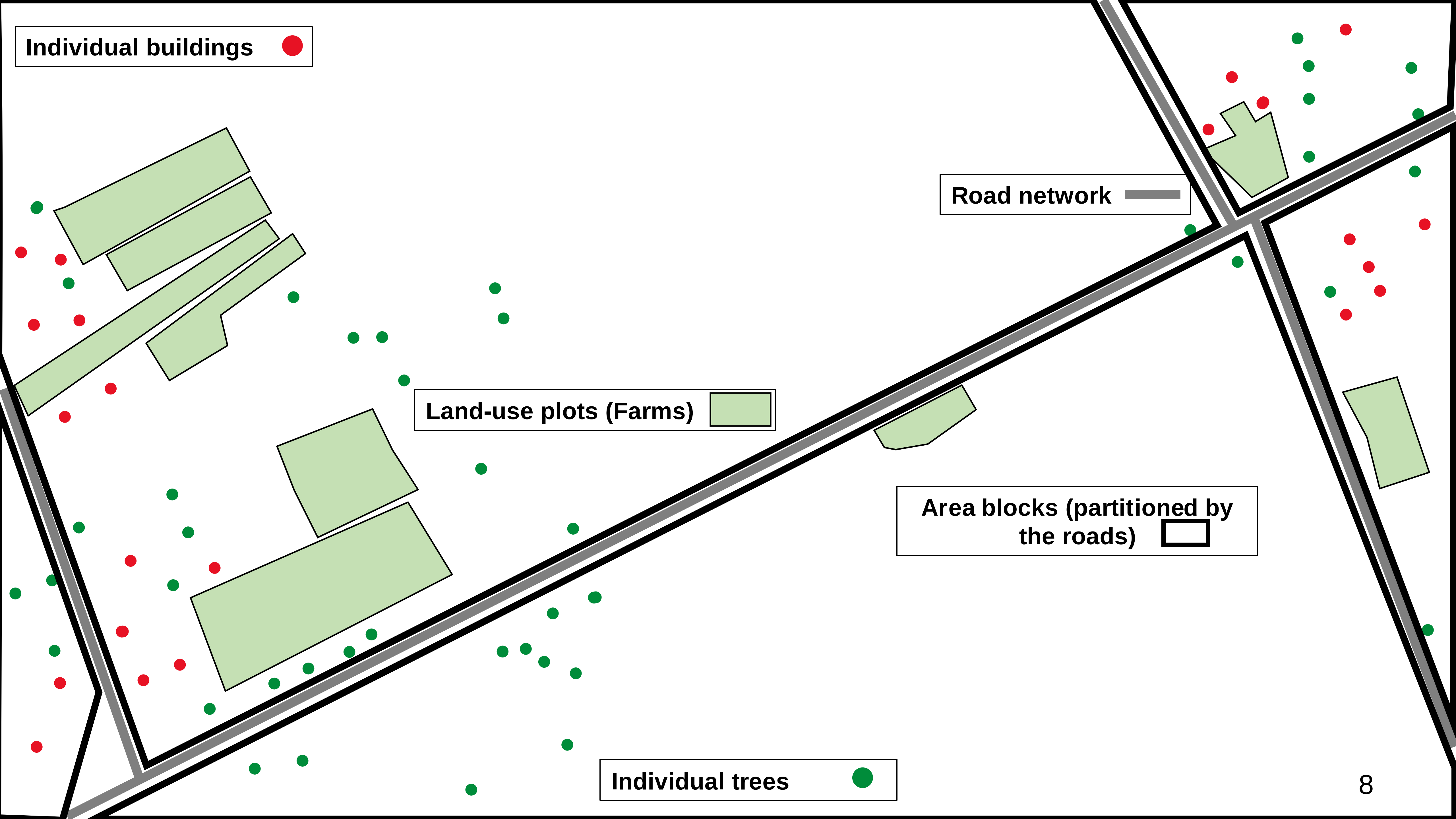
- Location of buildings
- Farm plots
- Locations of trees
- Road network
- Block areas (divided by the road)

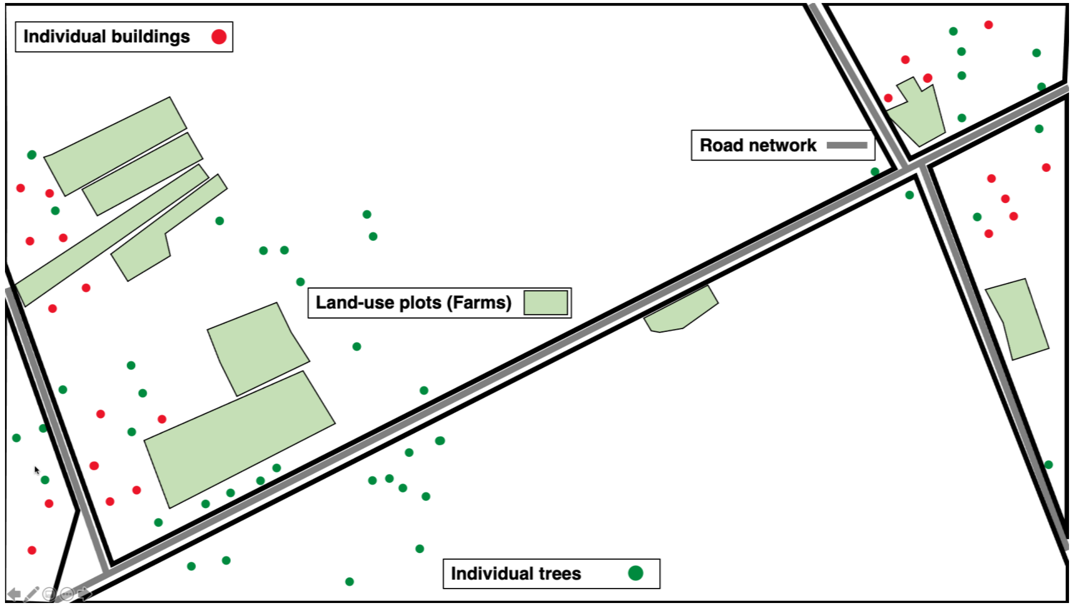
2. Levels of soil moisture across the landscape

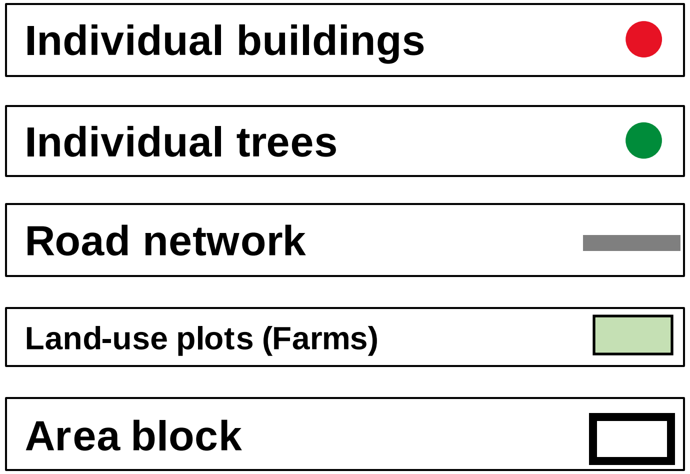

Individual buildings
Road network
Land-use plots (Farm)
Area blocks (partitioned by the roads)
Individual trees
1206205001



Individual buildings
Road network
Land-use plots (Farms)
Area blocks (partitioned by the roads)
Individual trees


Individual buildings
Road network
Land-use plots (Farms)
Area blocks (partitioned by the roads)
Individual trees

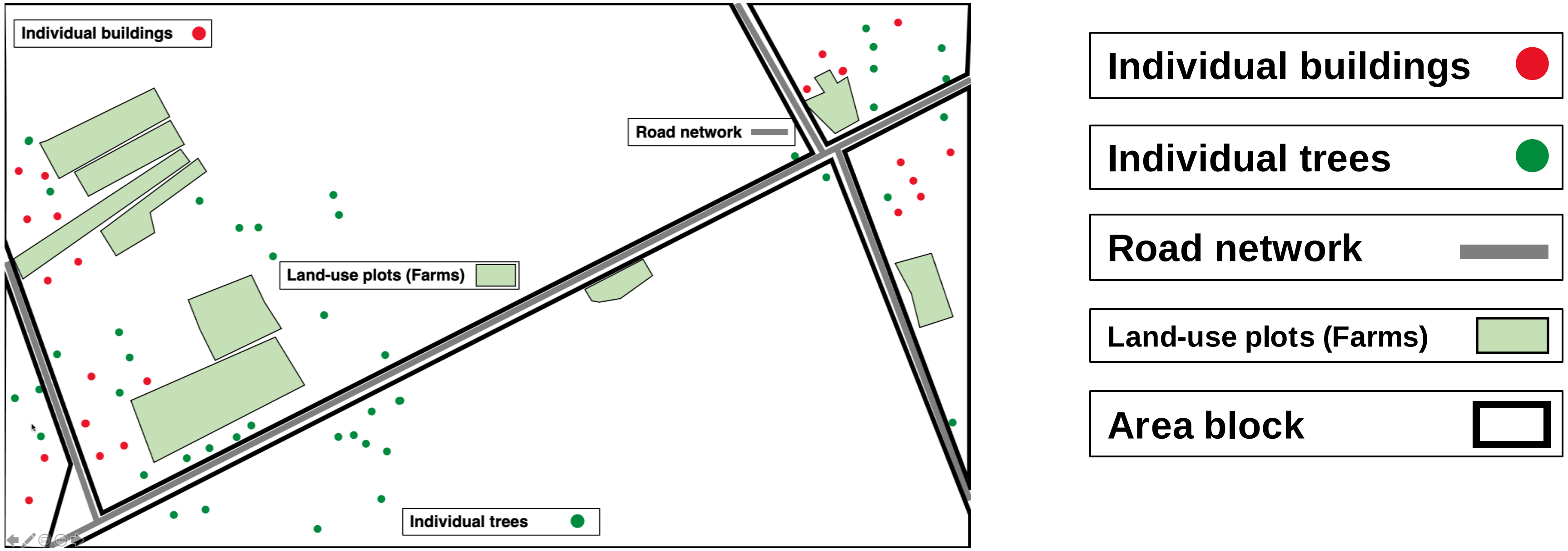




The above objects listed are called “**Features**”. A feature can be described according to its characteristics which is termed an “**Attribute**” in GIS. The attribute of a feature can be a **numeric** or **text** observation.

For example:

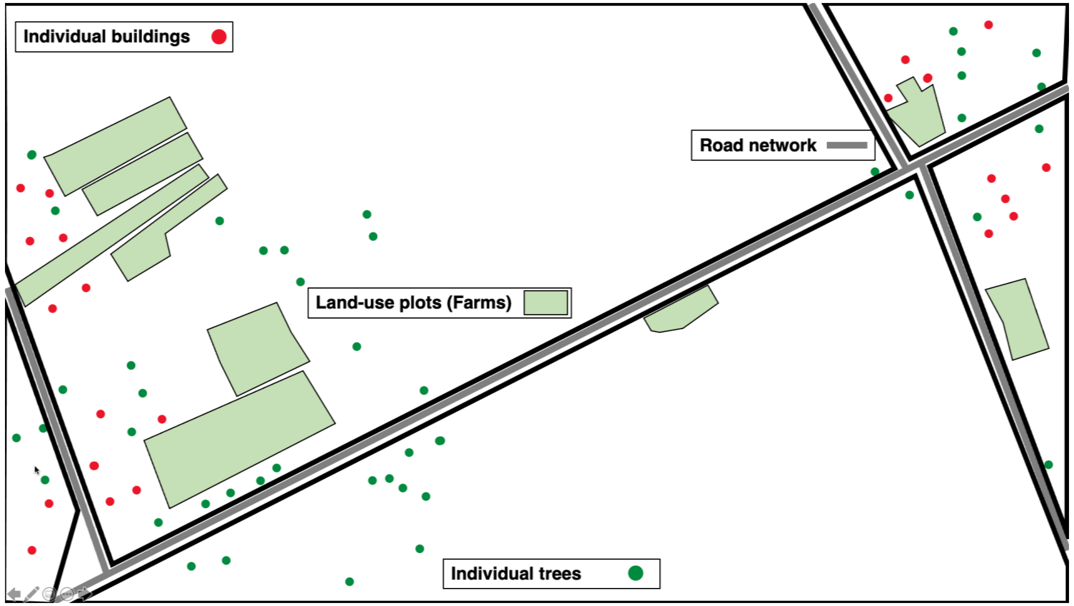
- A building is a **point feature** on this map, the number of people living a building is a **numeric attribute** describing this **feature**. Type of building (i.e., Victorian or modern) is a **text attribute**
- The road network is a **polyline feature**, the length (or distance (m)) of the road is a **numeric attribute** describing the road
- Land-use plot is a **polygon (or area) feature**, the type of land-use (i.e., farming) is the **text attribute** describing what that polygon is etc.

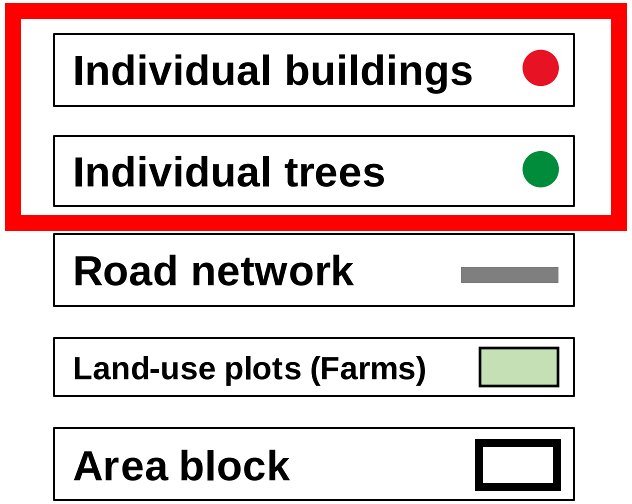


Now, the above objects are typically **Discrete** features, and therefore classed a **Vector Data**

There are three main types of **Vector Data**:

1. **Point vector**
2. **Polylines or Line vector**
3. **Polygon vector**





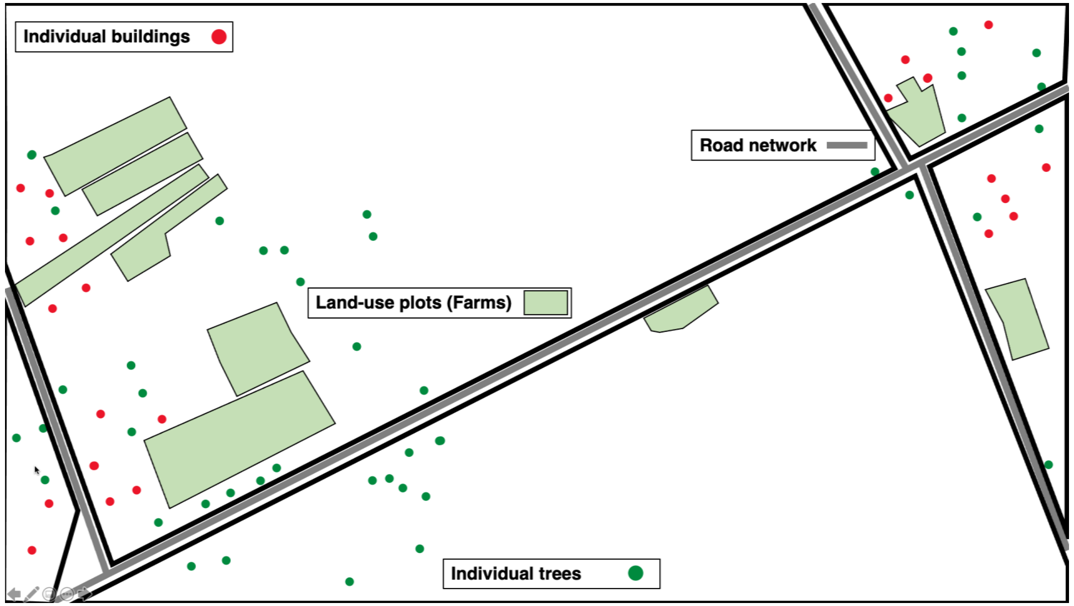
Now, the above objects are typically **Discrete** features, and therefore classed a **Vector Data**

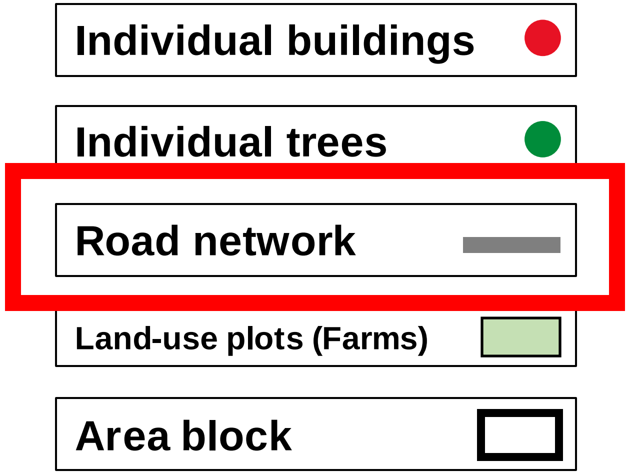
There are three main types of **Vector Data**:

1. **Point vector**
2. **Polylines or Line vector**
3. **Polygon vector**

Characteristics of a point vector

- X, Y location characterize as a coordinate
- Has no area
- Has no length
- This applies to discrete features of sample points
- Its geometry consists of a single **node** or **vertex**





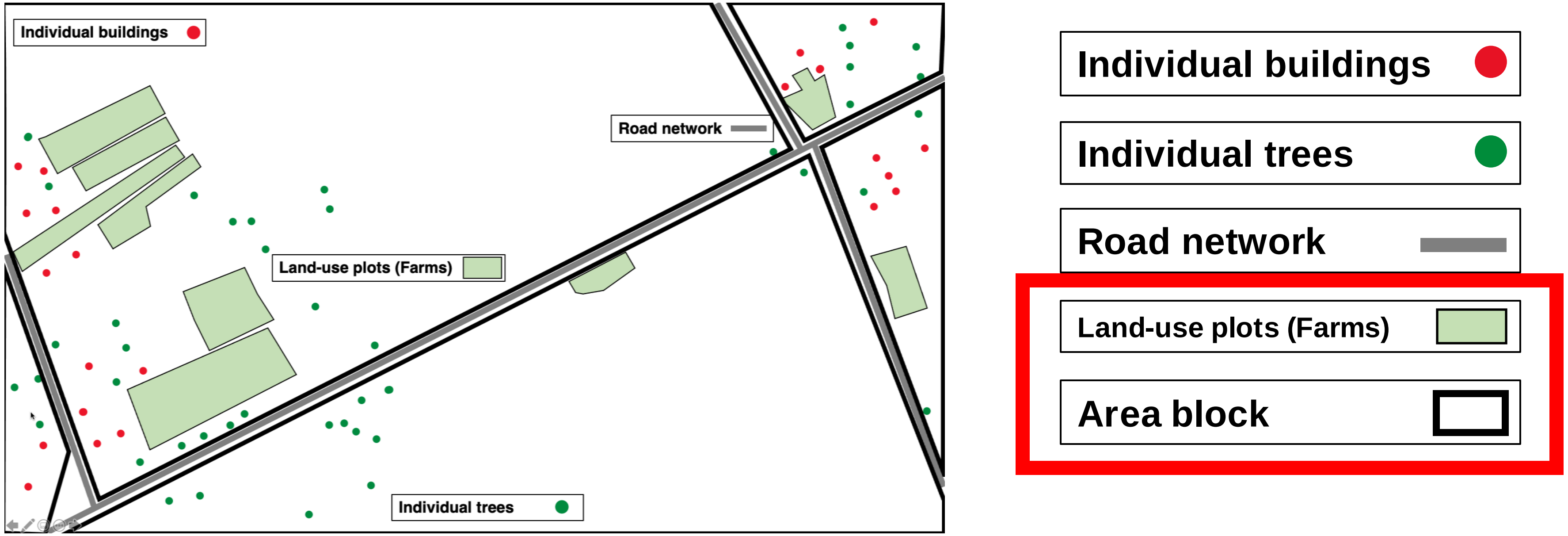
Now, the above objects are typically **Discrete** features, and therefore classed a **Vector Data**

There are three main types of **Vector Data**:

1. **Point vector**
2. **Polylines or Line vector**
3. **Polygon vector**

Characteristics of a polyline or line vector

- These are a series of X, Y points characterize by coordinates to form a line
- Has no area
- They have a length
- They have a direction (important for visualizing rivers steams & roads)
- Connectivity (it connecting to other lines segments in the network)
- This applies to features without an area but with a length – roads, rivers, railway tracks or migration flows between two or more locations.
- Its geometry consists of **2 nodes** (i.e., beginning and end of point of line) & can have more than one **vertex** (i.e., point(s) that connect different lines together)
- It is never enclosed



Now, the above objects are typically **Discrete** features, and therefore classed a **Vector Data**

There are three main types of **Vector Data**:

1. **Point vector**
2. **Polylines or Line vector**
3. **Polygon vector**

Characteristics of a polygon vector

- These are a series of X, Y points characterize by coordinates to form an enclosed region
- It has an area
- It has no length but rather a perimeter instead
- This applies to features with enclosed regions – e.g., postcode areas, area of residential premise, counties (other administrative boundaries) etc.
- Polygons have **three vertices** or more each connecting sequentially where the first vertex connects with the last vertex.

What about the soil moisture

Amount of moisture in soil at any of spot in these rectangle pixels to see change in gradient

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6	7	8	10	10	0	10	10	0	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	0	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	0	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

Soil moisture index	
	10
	7 to 9
	4 to 6
	1 to 3
	0

Unlike the vector data. The above feature describes how moisture levels across the surface of the landscape – the feature is is not measured discretely but on a **continuous** surface to show gradient in changes for soil moisture across the landscape

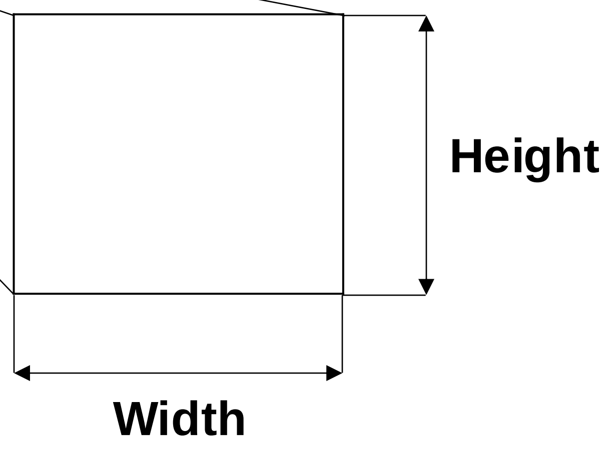
Now, this **Non-discrete** feature is classed a **Raster Data**

What is Raster Data?

- **It is a matrix of pixels or grid-cells** that contains a numeric or text value for a feature its representing
- It is composed of **rows and columns**
- Each pixel or grid-cell has a **resolution** (or **size for height and width**)

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

Soil moisture index	
	10
	7 to 9
	4 to 6
	1 to 3
	0



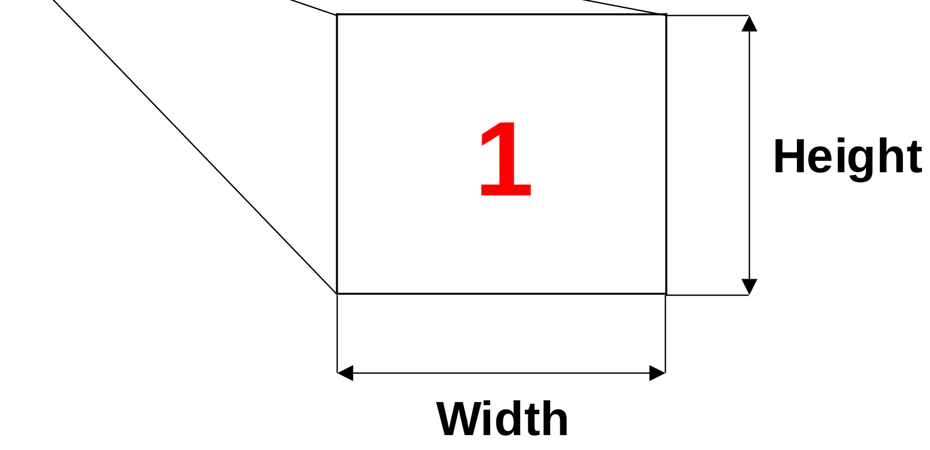
Characteristics of Raster Data and its pixels (or grid cells):

- It can take any shape (circle, square, rectangle, hexagon etc.). The usual shape you will encounter is a **square**
- Resolution (width and height: units - m, km, miles etc.)

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

Soil moisture index	
	10
	7 to 9
	4 to 6
	1 to 3
	0

Example with Boolean



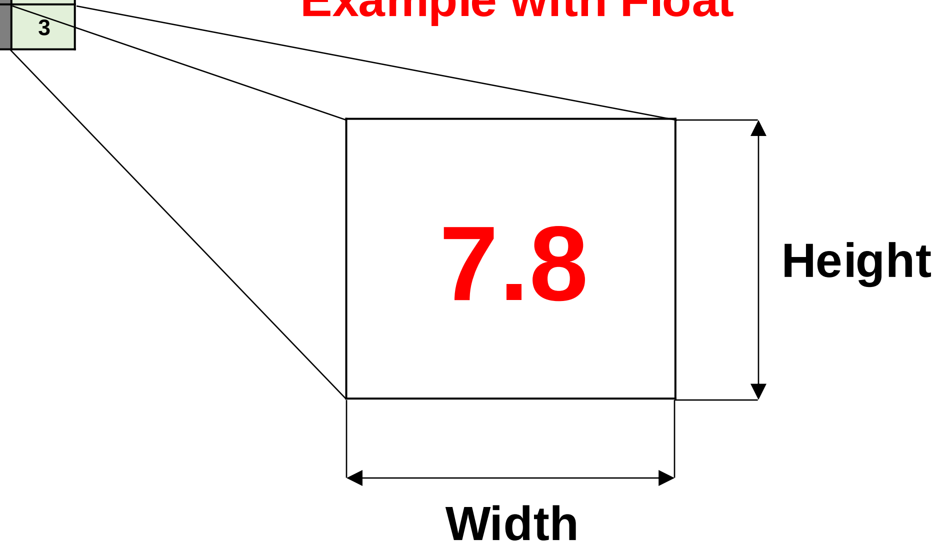
Characteristics of Raster Data and its pixels (or grid cells):

- It can take any shape (circle, square, rectangle, hexagon etc.). The usual shape you will encounter is a **square**
- Resolution (width and height: units - m, km, miles etc.)
- Data types:
  - Integer: for discrete values (e.g., above raster for moisture)
  - Boolean (or Binary) 1 = “Yes” and 0 = “No” (Presence/absence)

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

Soil moisture index	
	10
	7 to 9
	4 to 6
	1 to 3
	0

Example with Float



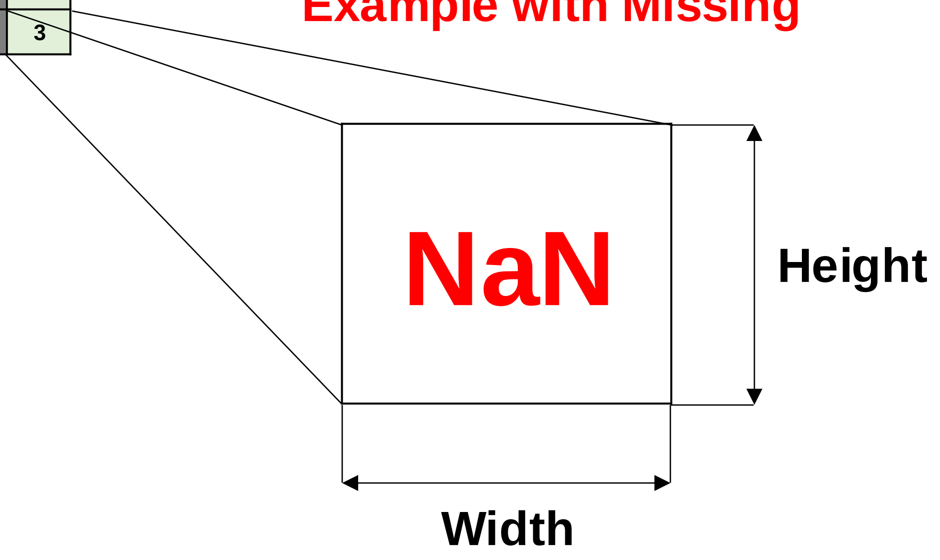
Characteristics of Raster Data and its pixels (or grid cells):

- It can take any shape (circle, square, rectangle, hexagon etc.). The usual shape you will encounter is a **square**
- Resolution (width and height: units - m, km, miles etc.)
- Data types:
  - Integer: for discrete values (e.g., above raster for moisture)
  - Boolean (or Binary) 1 = “Yes” and 0 = “No” (Presence/absence)
  - Float (or continuous data)

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

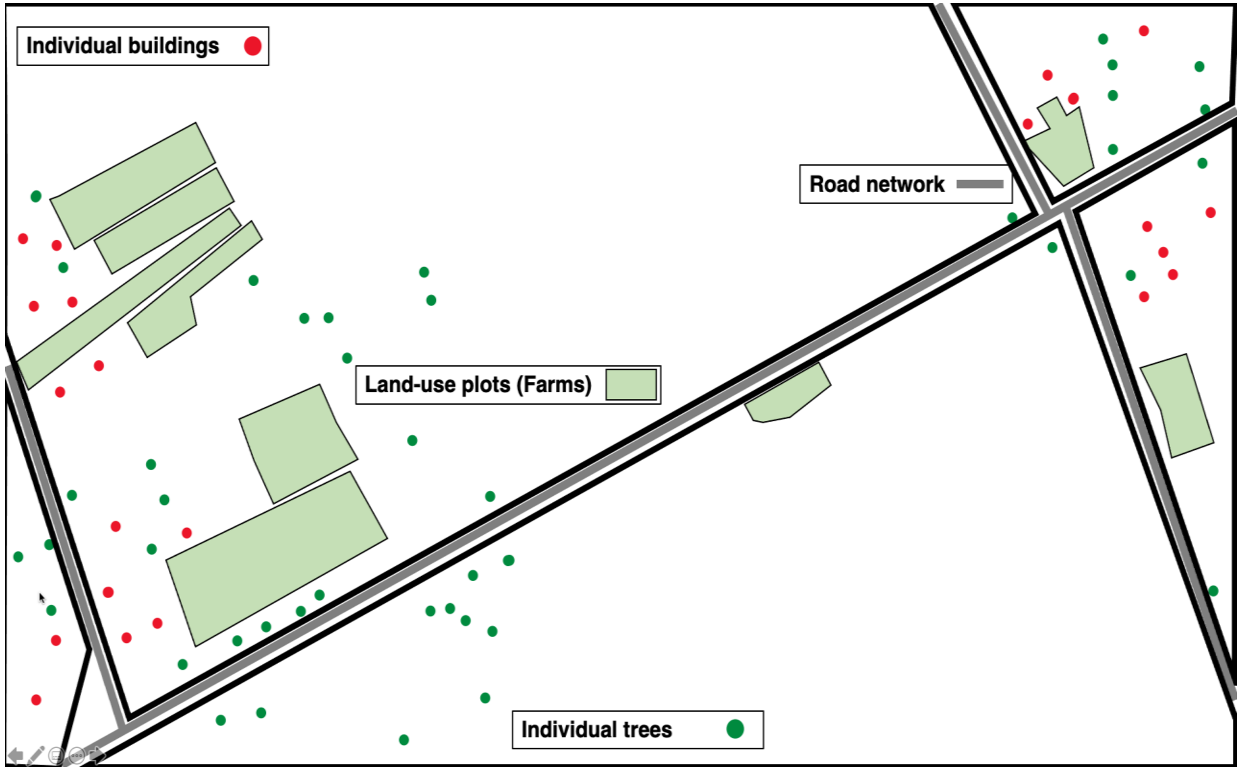
Soil moisture index	
	10
	7 to 9
	4 to 6
	1 to 3
	0

Example with Missing



Characteristics of Raster Data and its pixels (or grid cells):

- It can take any shape (circle, square, rectangle, hexagon etc.). The usual shape you will encounter is a **square**
- Resolution (width and height: units - m, km, miles etc.)
- Data types:
  - Integer: for discrete values (e.g., above raster for moisture)
  - Boolean (or Binary) 1 = “Yes” and 0 = “No” (Presence/absence)
  - Float (or continuous data)
  - No data, missing value in pixel: NA, NaN



Vector data

6	7	8	10	0	0	10	10	0	0	0	0	10	5	3	0	0	0	0	0
6	7	8	10	0	0	10	10	0	0	10	0	10	6	3	0	0	0	0	0
6	7	8	10	10	0	10	10	0	0	10	0	10	7	5	3	0	0	0	0
5	6	8	9	10	10	0	10	0	0	10	0	10	7	5	3	0	0	0	0
1	4	8	9	9	10	0	10	10	10	0	0	0	7	5	3	0	0	0	0
0	4	8	9	9	10	10	0	10	9	9	0	0	5	3	0	0	0	0	0
0	4	8	8	9	9	10	0	0	9	8	7	5	0	0	0	1	0	0	0
0	3	5	8	8	9	10	10	0	9	7	5	0	0	5	5	5	0	0	0
0	2	3	5	8	9	9	10	0	0	3	0	0	0	5	0	0	1	0	0
0	2	2	5	8	8	9	9	10	0	0	0	1	5	0	0	0	0	0	0
0	2	2	4	6	8	8	9	0	0	0	1	5	0	0	5	5	5	0	0
0	0	2	3	6	8	8	0	0	0	0	5	0	5	5	5	5	5	0	0
0	0	2	2	5	8	0	0	0	0	0	0	5	5	5	5	5	5	3	0
0	0	0	2	5	0	0	1	2	3	4	4	4	4	4	4	4	5	0	0
0	0	0	0	0	0	1	1	1	1	4	4	4	4	4	4	4	5	0	0
0	0	0	0	1	1	2	2	2	2	3	3	3	3	3	3	3	4	0	3
0	0	1	1	1	1	2	2	3	3	3	3	3	3	3	3	3	4	0	3

Raster data

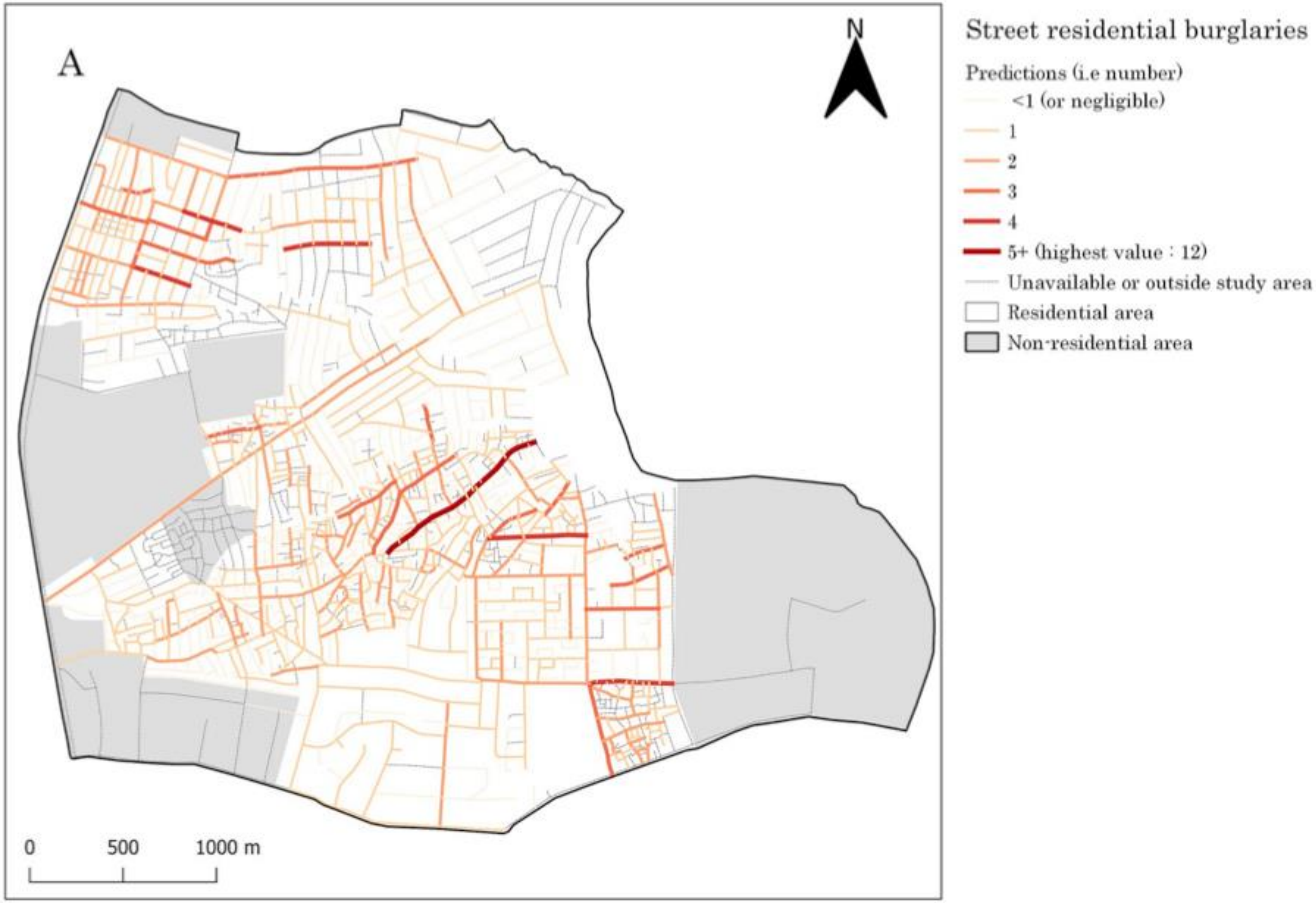


Image: Show predicted cases of residential burglaries in study area

Sources:
Musah, A., *et al.* (2020): https://doi.org/10.1016/j.apgeog.2019.102126

Let's look at this first example:

The map on the left shows the predicted numbers of burglaries across streets for a small city in Kaduna, Nigeria (*Musah et al. 2020*)

- What types of spatial data are shown in the map on the left?
- What kinds of **features** are included in this visualisation?
- For the street segments, what is the main **attribute** for their visualisation?

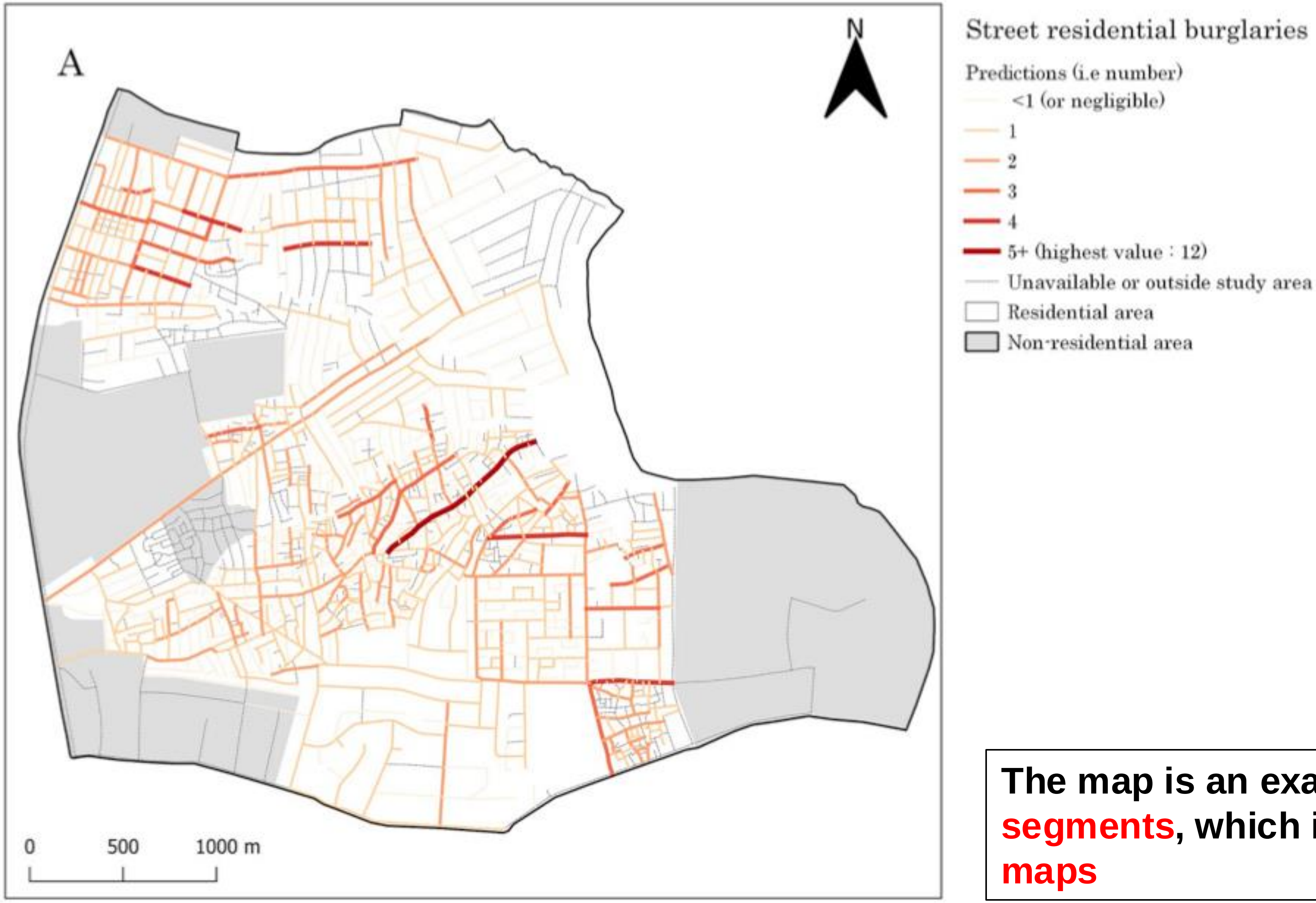


Image: Show predicted cases of residential burglaries in study area

- What types spatial data are shown in the map on the left?

  Vector data only

- What kinds of **features** are included in this visualisation?

  Polyline (Street segments)
  Polygon (Residential area (white))
  Polygon (Non-residential area (grey))
  Polygon (Black outline of study area)

- For the street segments, what is the main **attribute** for their visualisation?

  Feature: Polyline (Streets)
  Attribute: Predicted burglaries

The map is an example of vector map with line segments, which is one of many types of thematic maps

Sources:

Musah, A., *et al.* (2020): https://doi.org/10.1016/j.apgeog.2019.102126

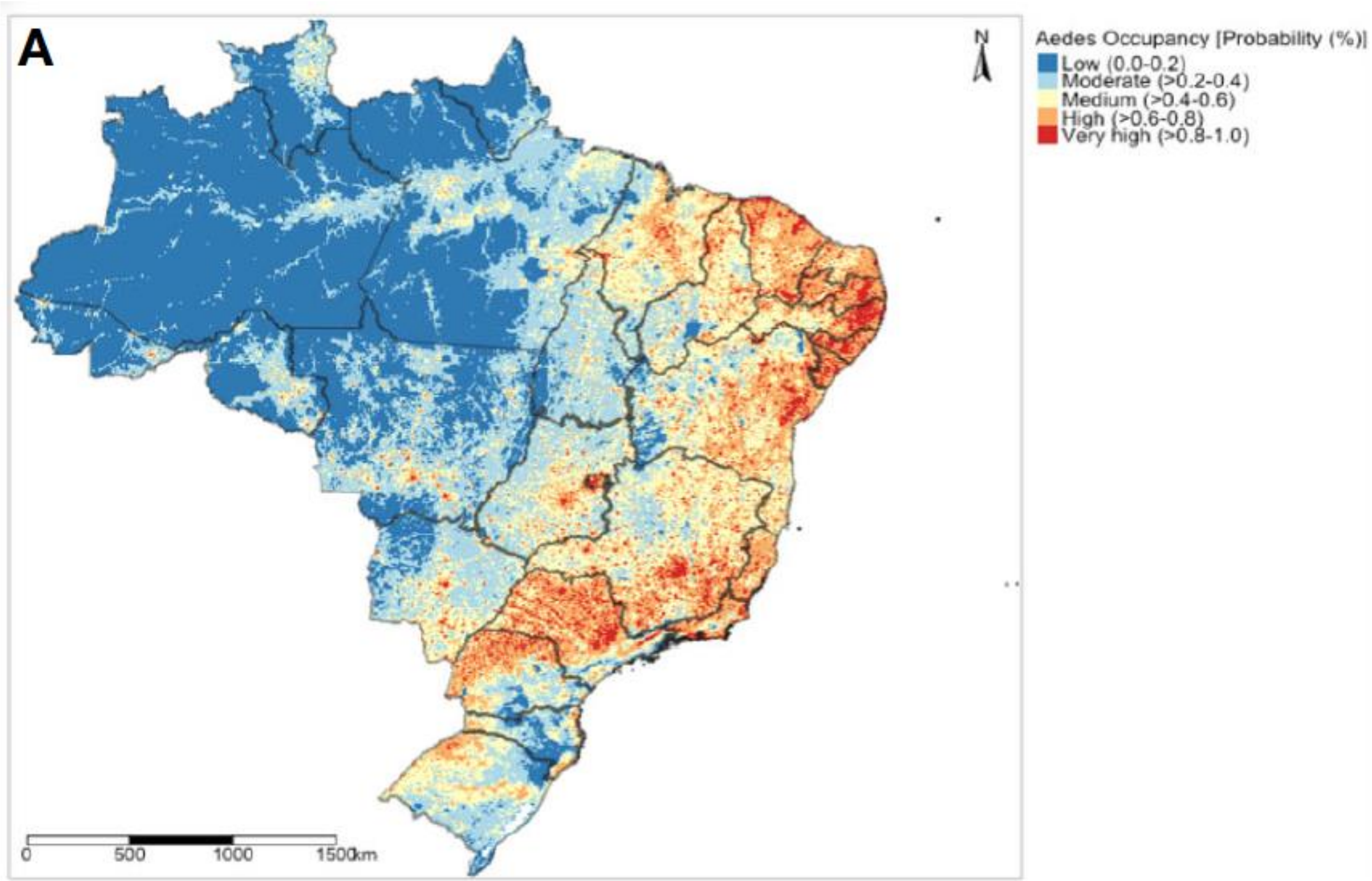


Image: Show predicted probability mosquito occupancy to indicate where Aedes aegypti thrive in Brazil

Let’s look at this second example:

The map on the left shows the predicted risk of mosquito (Aedes aegypti) occupancy in Brazil (*Musah et al. 2023*)

- What types of spatial data are shown in the map on the left?
- What kinds of **features** are included in this visualisation?
- For the raster, what is the main **attribute** for their visualisation?

Sources:

Musah, A., *et al*. (2023): https://doi.org/10.3389/fitd.2023.1039735

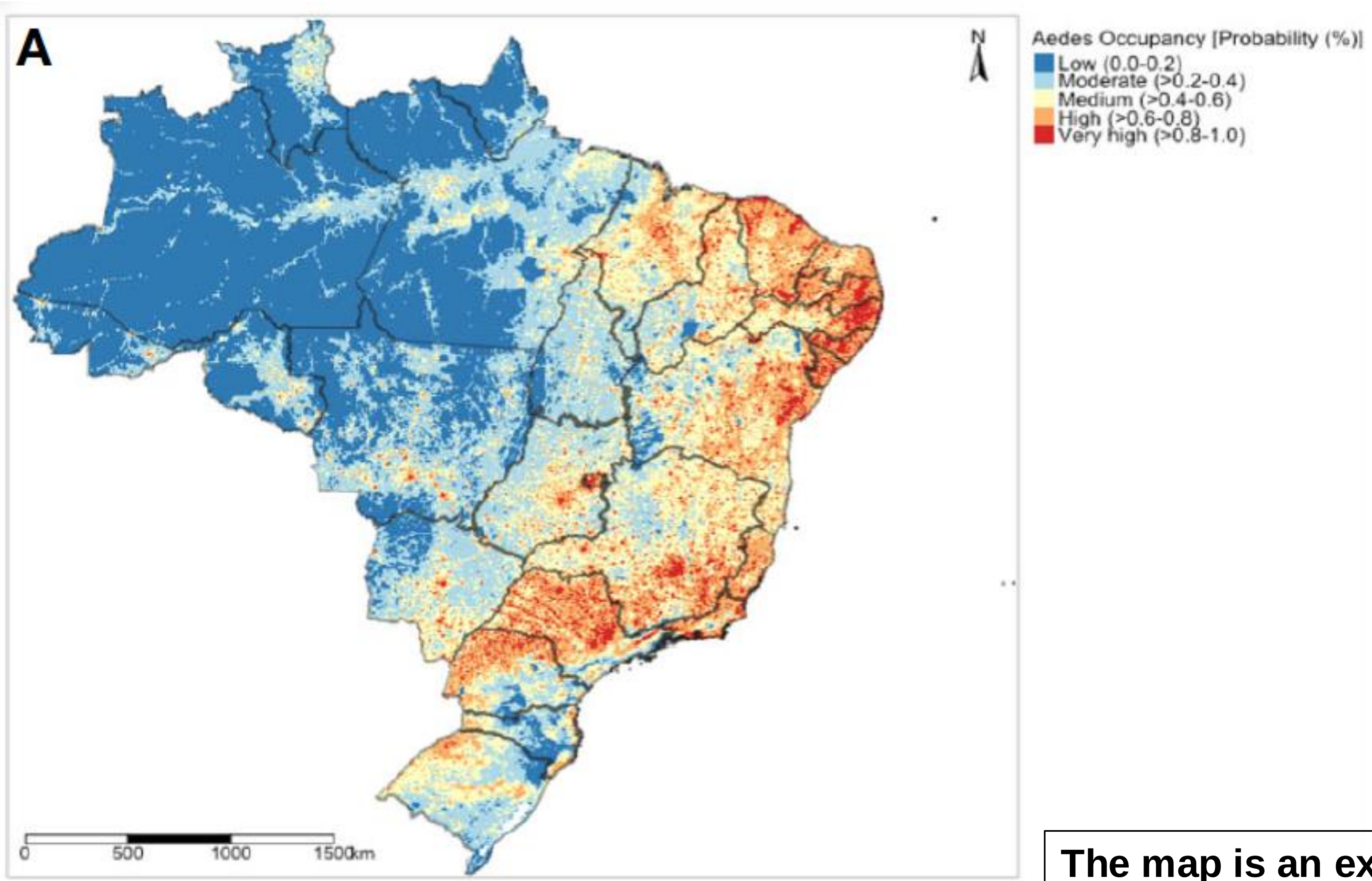


Image: Show predicted probability mosquito occupancy to indicate where Aedes aegypti thrive in Brazil

- What types spatial data are shown in the map on the left?

  Mainly raster data, and vector data

- What kinds of **features** are included in this visualisation?

  Polygon (Brazilian States)
  Raster (Prediction of risk)

- For the street segments (polylines), what is the main **attribute** for their visualisation?

  Feature: Raster (4.5km by 4.5km)
  Attribute: Predicted probability

The map is an example of raster map with predictions store in pixels/grid points, which is one of many types of thematic maps

Sources:
Musah, A., *et al*. (2023): https://doi.org/10.3389/fitd.2023.1039735

Spatial Operations, or Geoprocessing

Definition: Geoprocessing is a collection of tools that offers a framework for dealing with geographic data

- Its purpose is to help users in the data management and automation of GIS tasks for spatial data.

- For performing geospatial modelling and analysis of vector and raster data.

In terms of spatial operations, the most problem you will encounter are:

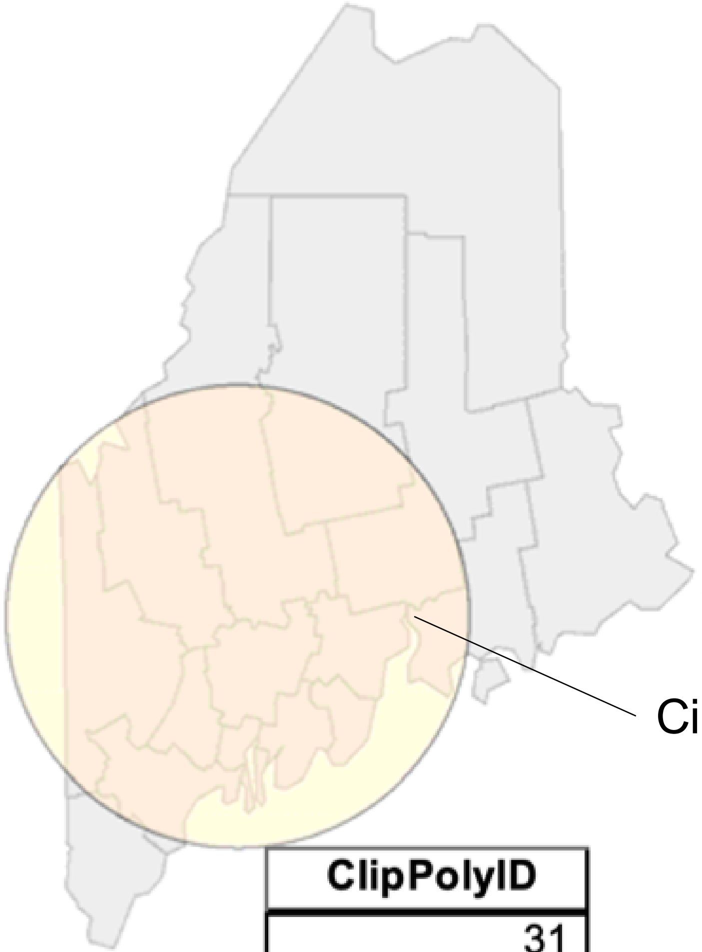
1. **Clipping**
2. **Intersection**
3. **Union**
4. **Dissolving**
5. **Aggregation**
6. **Buffer**

Clipping

Capture spatial data within a region of interest

NAME
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York

State of Maine



ClipPolyID
31

Circular polygon is our 'cookie cutter' for the clipping

Suppose we want to capture only the counties within that circular shape.

We are going to use the circular shapefile as our "**cookie cutter**" to capture the counties within it and clip out the areas that fall outside it.

The operations is called "**Clipping**"

Source: Gimond, M. 2021. Intro to GIS and Spatial Analysis. [online]. https://mgimond.github.io/Spatial/introGIS.html

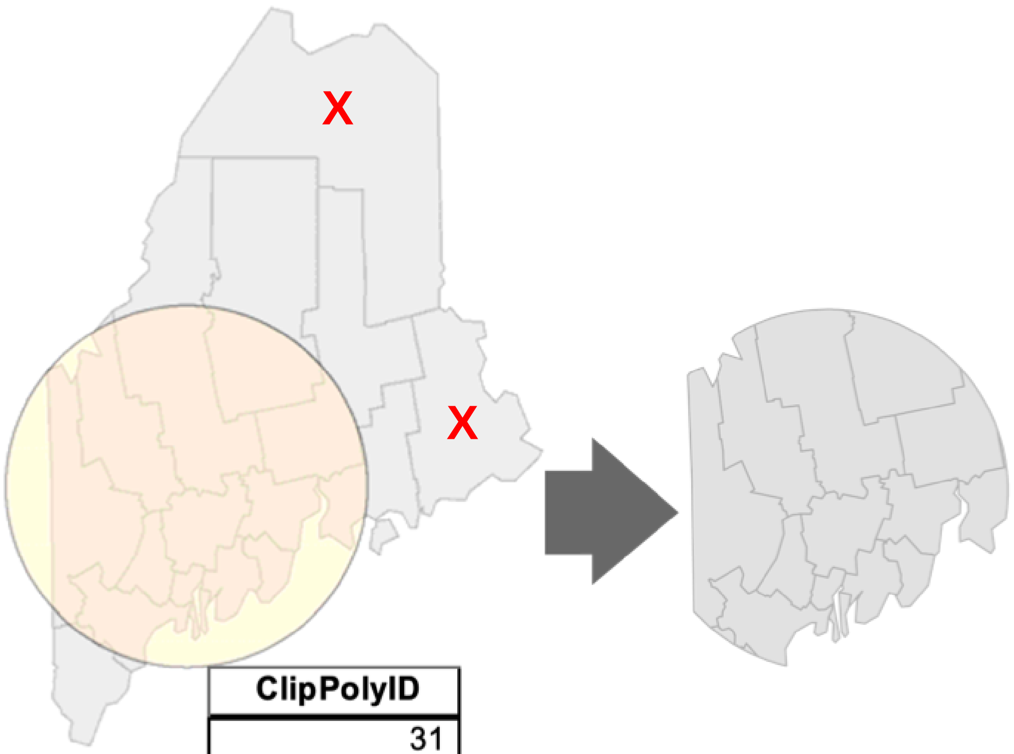
Clipping

Capture spatial data within a region of interest

Before

NAME
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York

State of Maine



ClipPolyID
31

Clipped area

NAME
Androscoggin
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
York

X = Aroostook and Somerset are excluded

Source: Gimond, M. 2021. Intro to GIS and Spatial Analysis. [online]. https://mgimond.github.io/Spatial/introGIS.html

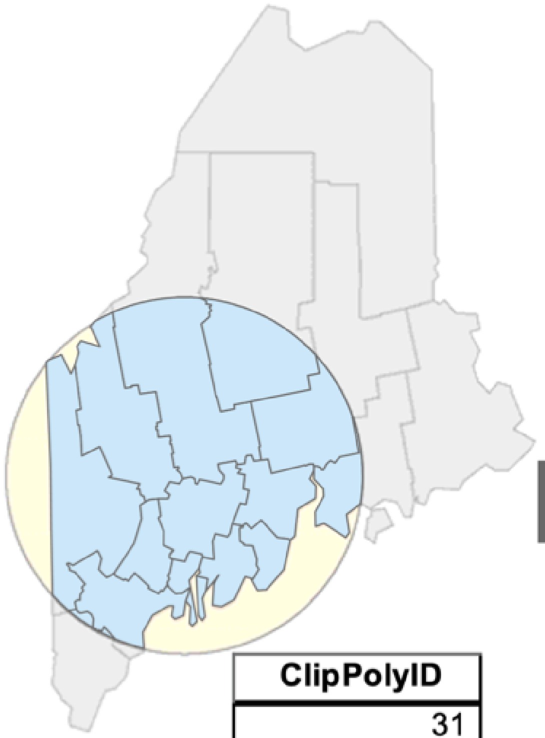
This spatial operation can be applied to all types of spatial data:

- Clipping points to fall within an area (i.e., vector to vector)
- Clipping smaller areal units to fall within a larger area (i.e., vector to vector)
- Clipping raster cells to fall within a spatial area (i.e., vector to raster)

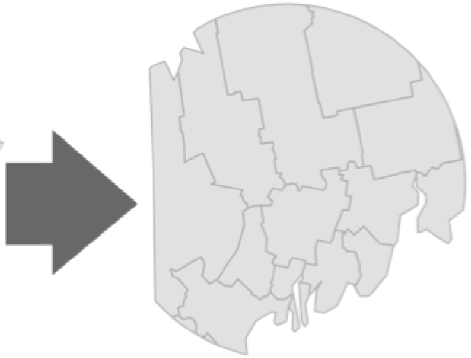
Intersection

Capture spatial data between two or more overlapping spatial data

NAME
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York



ClipPolyID
31



NAME	ClipPolyID
Androscoggin	31
Cumberland	31
Franklin	31
Hancock	31
Kennebec	31
Knox	31
Lincoln	31
Oxford	31
Penobscot	31
Piscataquis	31
Sagadahoc	31
Somerset	31
Waldo	31
York	31

Source: Gimond, M. 2021. Intro to GIS and Spatial Analysis. [online]. https://mgimond.github.io/Spatial/introGIS.html

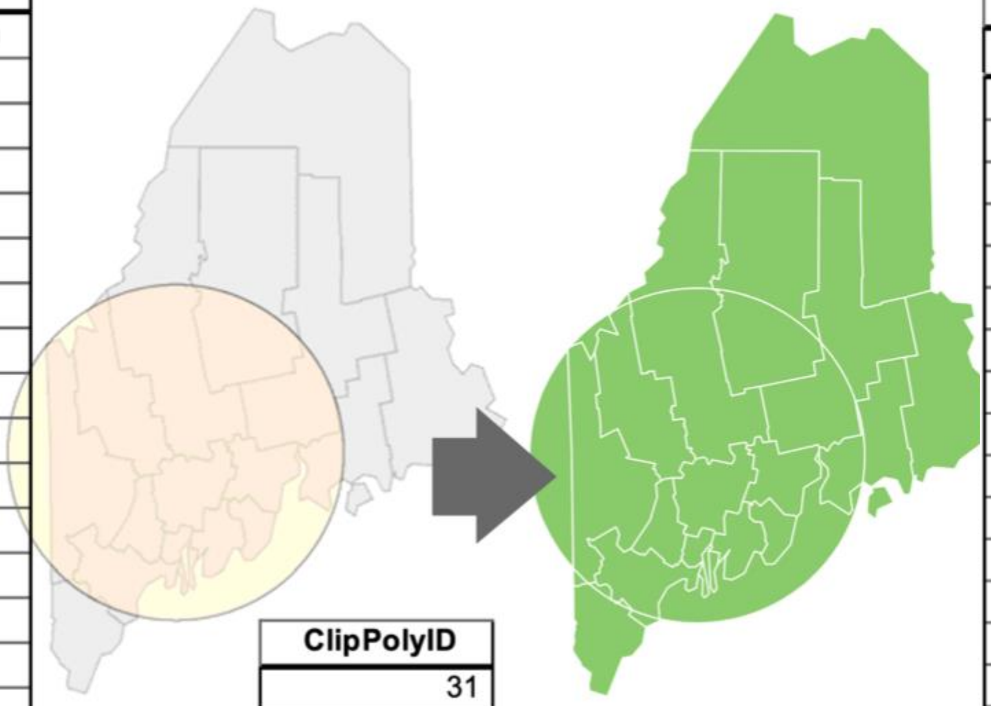
This type of spatial operation can be applied to the following:

- Intersecting points to areas (i.e., vector to vector)
- Intersecting smaller areal units that fall within a larger area (i.e., vector to vector)
- Intersecting points with raster (i.e., vector to raster)

Union

Bringing together two or more different vectors objects to one

NAME
Androscoggin
Aroostook
Cumberland
Franklin
Hancock
Kennebec
Knox
Lincoln
Oxford
Penobscot
Piscataquis
Sagadahoc
Somerset
Waldo
Washington
York



ClipPolyID
31

NAME	ClipPolyID
	31
Androscoggin	N/A
Aroostook	N/A
Cumberland	N/A
Franklin	N/A
Hancock	N/A
Kennebec	N/A
Knox	N/A
Lincoln	N/A
Oxford	N/A
Penobscot	N/A
Piscataquis	N/A
Sagadahoc	N/A
Somerset	N/A
Waldo	N/A
Washington	N/A
York	N/A

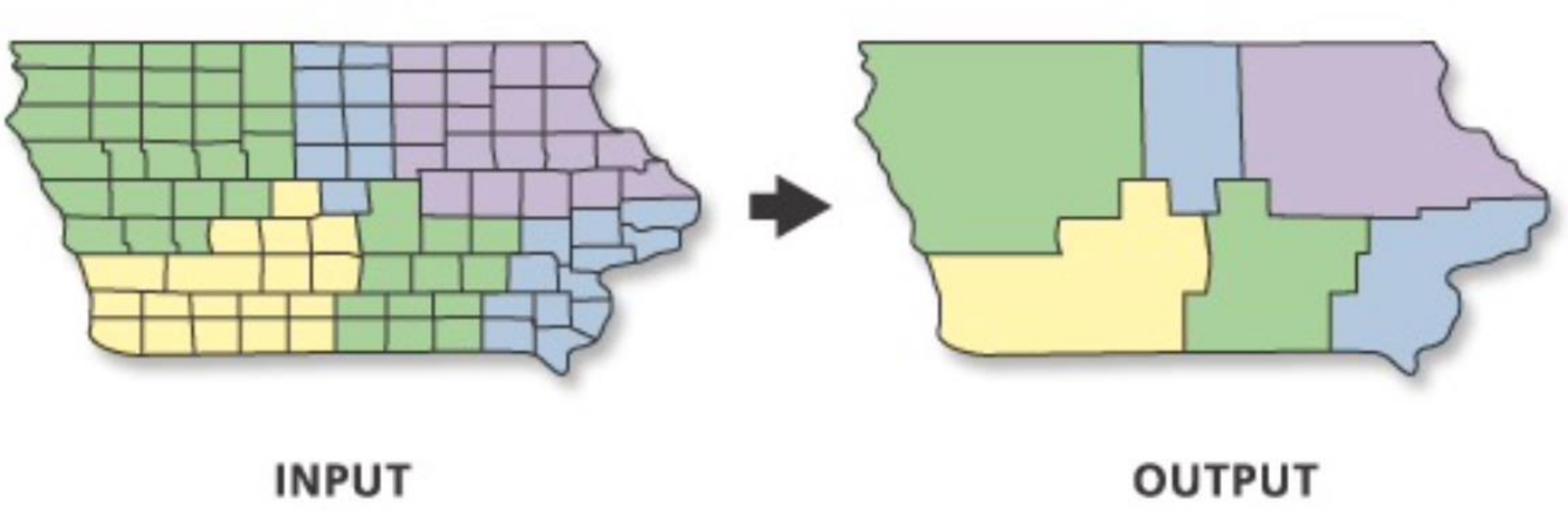
Source: Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]. https://mgimond.github.io/Spatial/introGIS.html

This type of spatial operation can be applied to the following:

- You union all types of vector data (i.e., points, lines and polygons)
- It’s a vector-to-vector spatial operation only. You cannot union a vector with a raster. The raster must be vectorised before it can be in union with another vector

Dissolving

Amalgamate or merge smaller adjacent vector units to form another vector unit.



Source: ESRI (2021) Dissolve https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/dissolve.htm

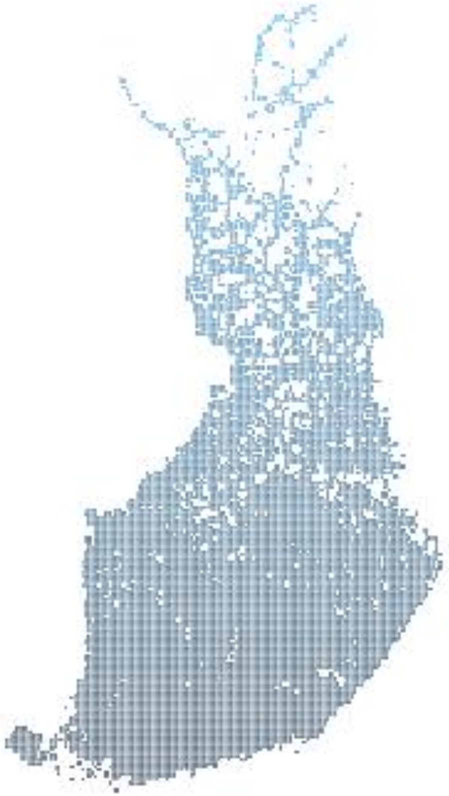
This type of spatial operation can be applied to the following:

- You dissolve only the following types of vector data (i.e., lines and polygons)
- It cannot be applied to vector points

Aggregation

Aggregating values (i.e., summing, or averaging) of smaller spatial units that fall within a larger spatial unit.

Population density grid (raster)

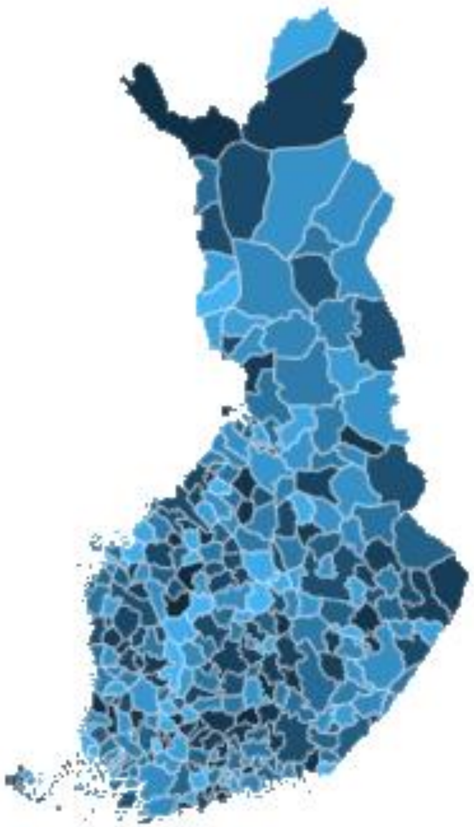


Highest resolution

Postcode areas (vector)

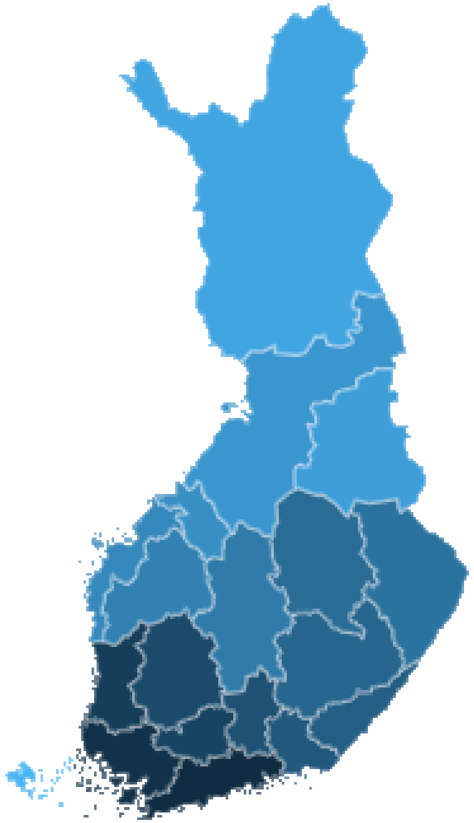
High resolution

Municipalities (vector)



Low resolution

Regions (vector)



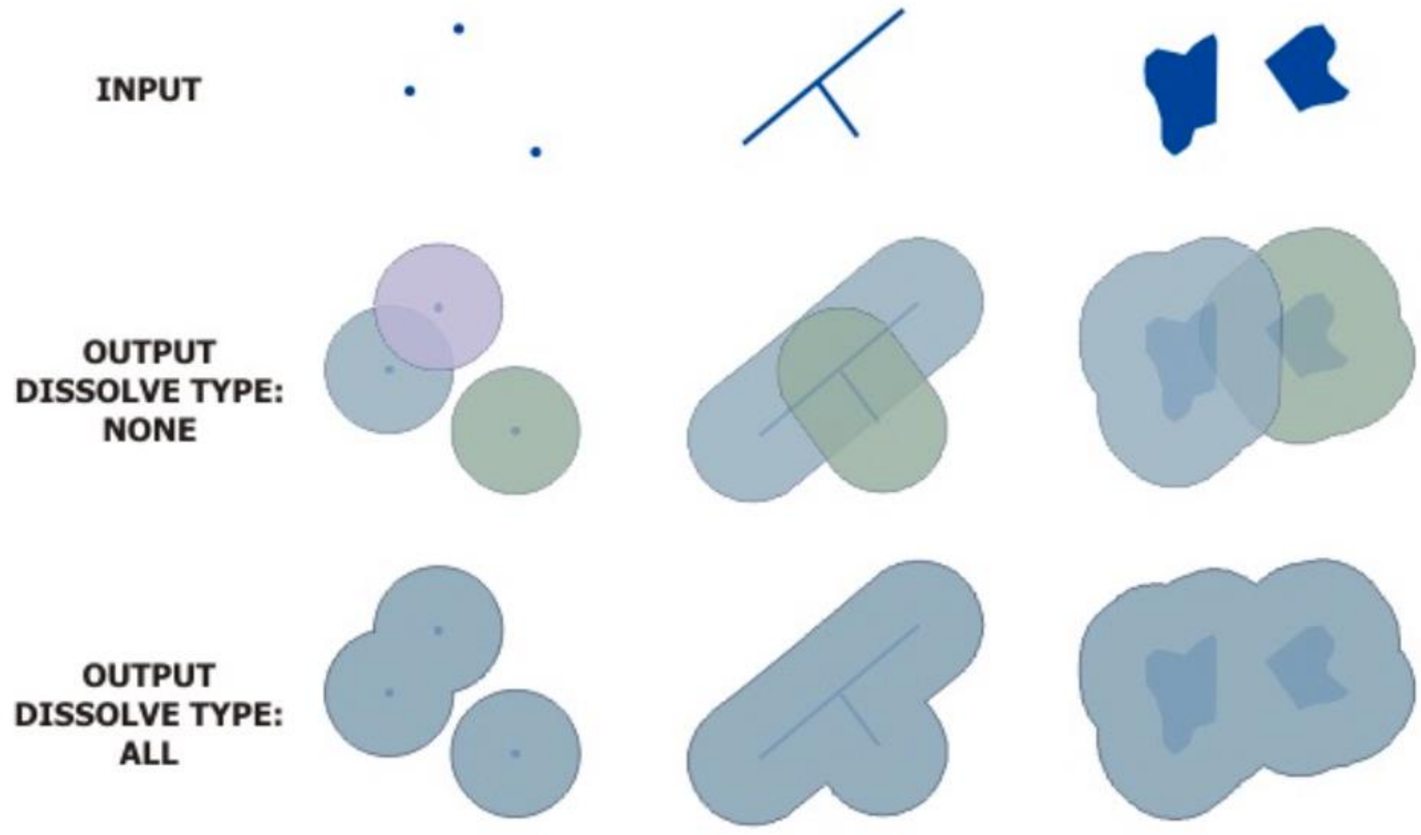
Coarser resolution

By aggregating (and summing) the gridded values within any of these vector shapefile, it will produce another shapefile containing the estimated total population within such boundaries

Source: geofi (access Finnish [Maakunta region] Geospatial data) https://cran.r-project.org/web/packages/geofi/readme/README.html

Buffering

Is an area, created by a specified distance, used to surround one or more features



Source: ESRI (2021). Buffer. https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/buffer.htm

This type of spatial operation can be applied to the following:

- You can create buffers using any type of vector data (i.e., points, lines and polygons)
- Buffers is a spatial operation reserved for vector only. This cannot be applied to raster data unless it's vectorised to a form that incredibly manageable for applying a buffer.

Definition: Geoprocessing is a collection of tools that offers a framework for dealing with geographic data

- Its purpose is to help users in the data management and automation of GIS tasks for spatial data.
- For performing geospatial modelling and analysis of vector and raster data.

In terms of spatial operations, the most problem you will encounter are:

1. **Clipping**
2. **Intersection**
3. **Union**
4. **Dissolving**
5. **Aggregation**
6. **Buffer**

Other types of spatial operations not covered here (but will see frequently in the practicals!)

- Spatial transforms of coordinates (reprojecting etc.)
- Coordinate systems and projections
- Distances

Thematic Mapping & Issues

Definition:

Thematic maps are graphical outputs that typically show geographic patterns of a particular theme in a geographic area

There are several types of thematic maps – the most common maps you will encounter are:

1. **Choropleths (e.g., Dot, Lines and Dasymetric maps)**
2. **Proportional point symbol maps**
3. **Raster maps**

Choropleth Map:

This type of map uses various colour schemes to represent aggregated values (i.e., sum, mean, and some other statistic) of an attribute within a predefined feature (i.e., points, lines or polygons)

- The key characteristics about Choropleths are the shading to show data on a map:
  - **Intensity of colours** – darker the colour (e.g., deep reds) may represent a higher value and the lighter colour (e.g., brighter reds) is vice versa
  - **Divergent colours** – it is best to represent two extremes e.g., deprivation map of London.
  - **Random colours** – it is best to use them for categorical variable whose categories have no order (e.g., urban rural classification, land use type etc.)

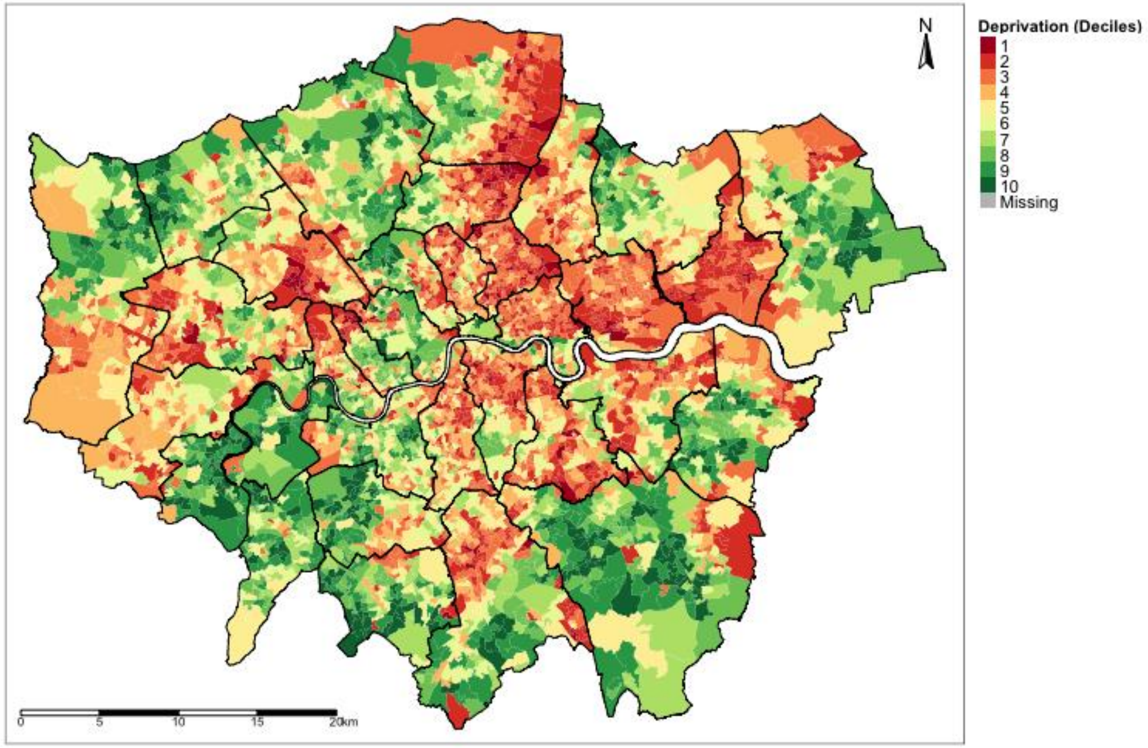
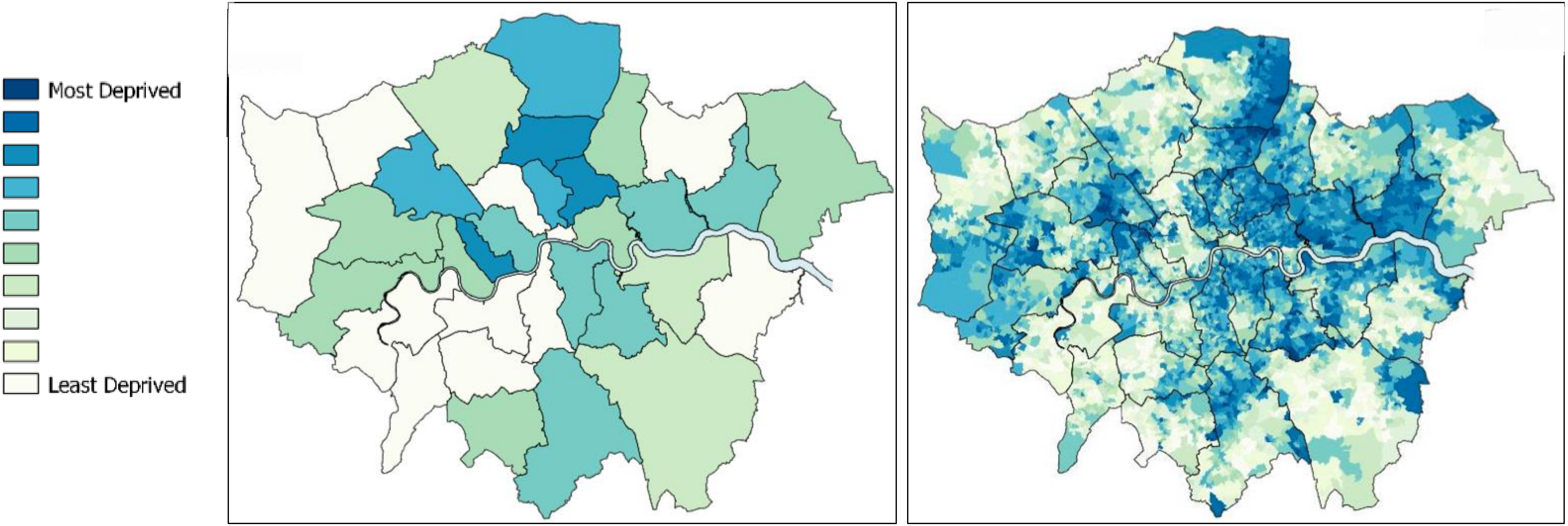


Image: We have generated a choropleth before in Week 1's practical. A map showing distribution of deprivation in London

Choropleths are great to use as they are easy to read and visually Impressive… but they come with their own issues thought!

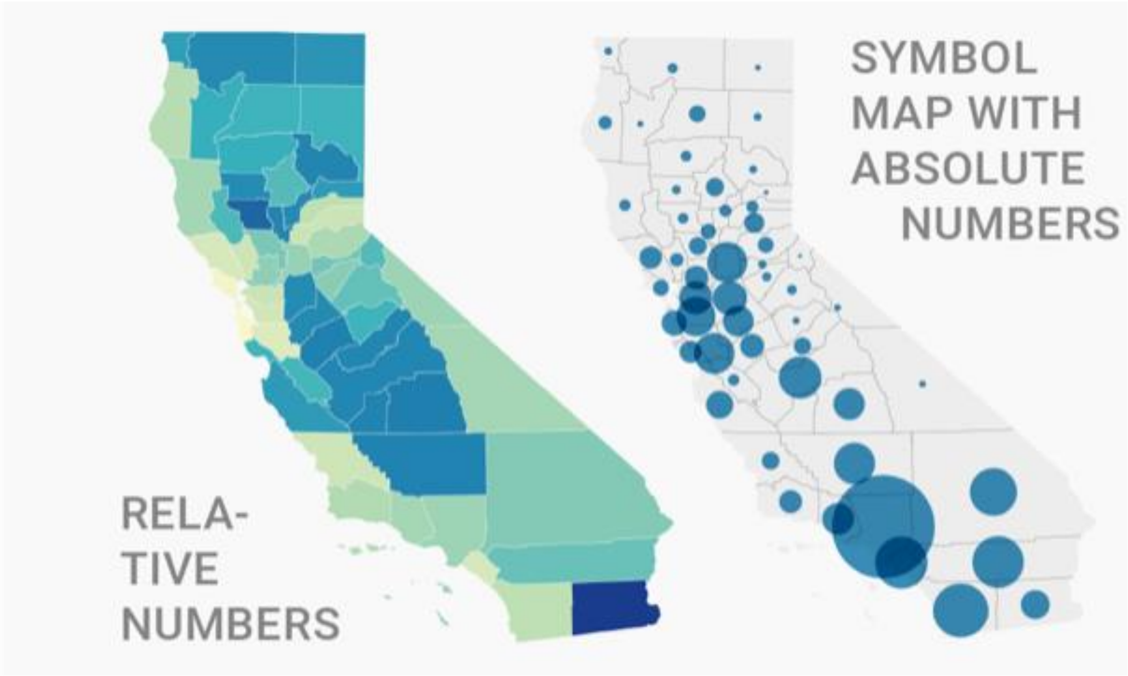


The two maps show the same thing – the one on the left shows levels of Deprivation based on the boroughs of London. While the one on the right shows it at a much higher resolution at a Lower Super Output Area level.

Problems:

1. The left map assumes the **whole area has the same value** – so it does not allow for variation within an area. We can see that this is not true because the right shows significant variation within an area.
2. The left map shows changes from one borough to another are **very abrupt** – which should not be the case. The variation from least to most deprived should be gradual (i.e., can be clearly seen on right map)
3. **Ecological Fallacy bias** is so pronounced in the left map.

The map on the right is a far better Choropleth – this is an example of **Dasymetric Map (more revealing)**



More Problems with Choropleth:

1. The far-left map uses **raw (or absolute) numbers** (i.e., number of people unemployed in counties in California). This is not an ideal way for visualizing such data because there is **no way** for us to compare difference across counties without knowing their population sizes. They are best visualized with **relative numbers** (i.e., unemployment rate or proportion) (middle map).

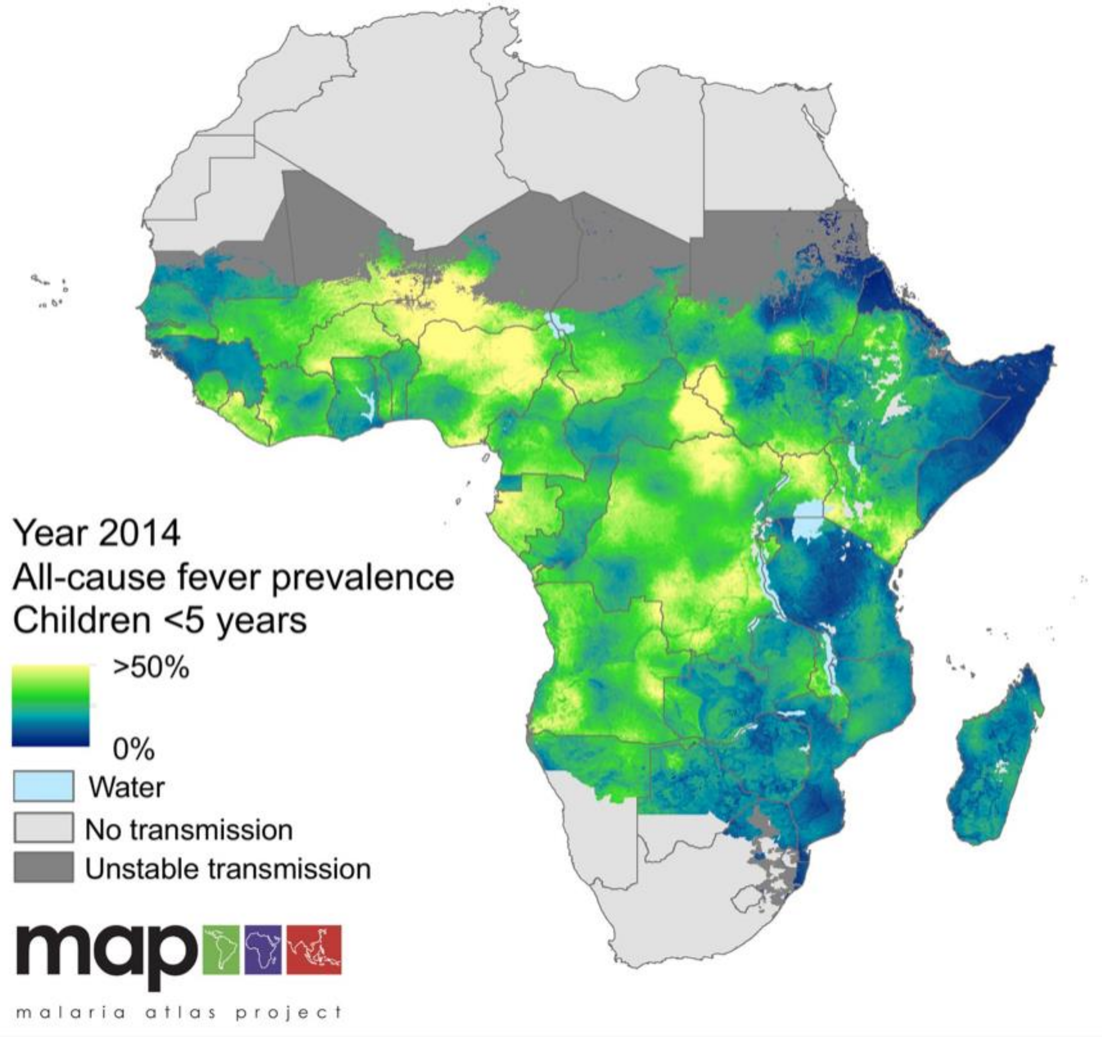
2. If you got to map data with absolute numbers – because the denominators are unavailable – then use symbols whose sizes are proportional to the absolute value (i.e., far-right map)

The map on the far-right is a far better Choropleth for absolute values – that example is called **Proportional Point Symbol Map**

Raster Map [1]:

Spatial resolution matters with raster data!

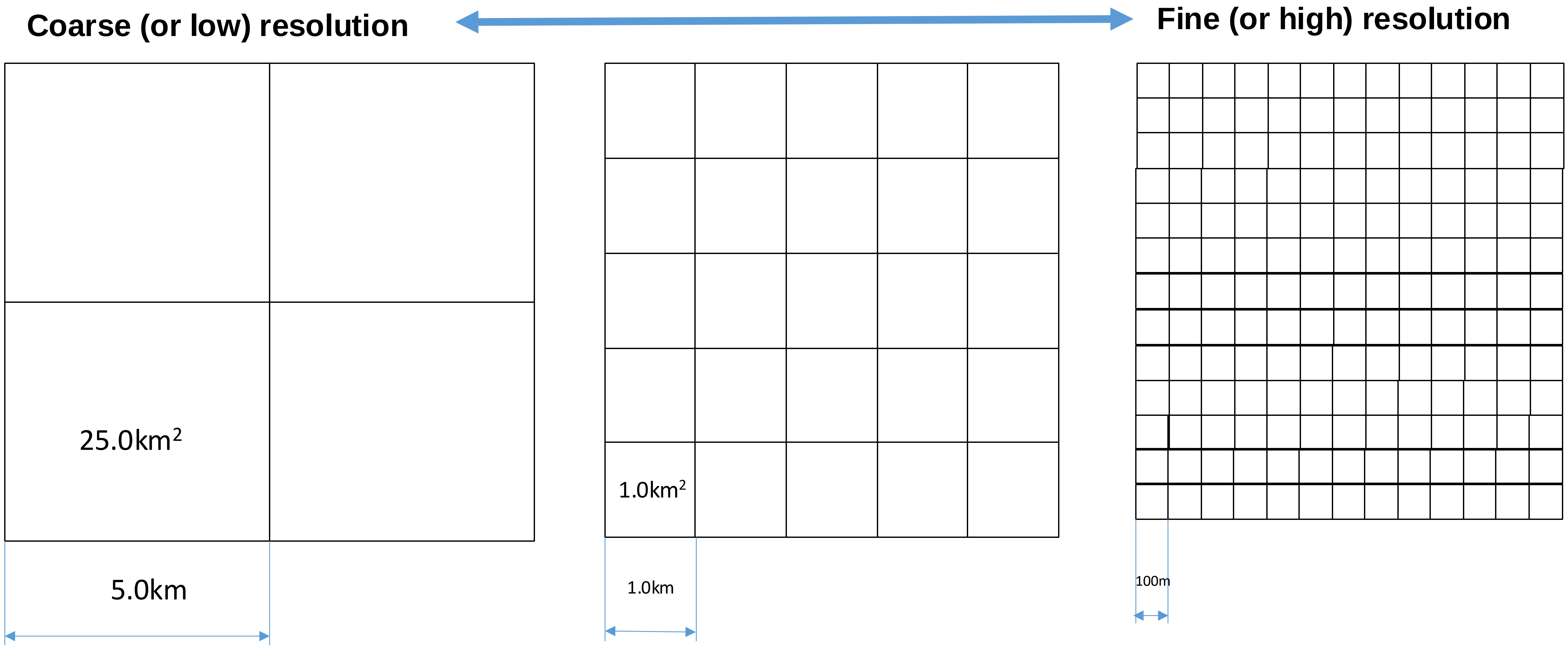
- Problems:
  - Re-analysis of the same data but at different pixel size can yield wildly different results which is annoying
  - Unlike vector data, which are series of vector layers can be easily overlaid spatial joining etc., Raster are difficult to achieve this (i.e., align one raster with another) if again, pixel sizes are different.
  - Incredibly difficult to work with – especially if they are in a high-dimensional format e.g., NetCDF
  - Depending on the number of pixels and amount of data stored in them – a single raster file can consume at a lot storage space on a PC.
  - Performing geostatistical modelling and generating raster at a high-resolution can be time consuming and cost-prohibitive for your PC.



Malaria Atlas Project: Modelling prevalence of Malaria Under < 5 year in Sub-Saharan Africa

Raster Map [2]:

Spatial resolution matters with raster data!



Summary

The take home message is:

- Vector and Raster data, are the two types of spatial data – the former maps out discrete entities and the latter visualizes entities over a continuous surface.
- Features are entities visualized on a map output, the attribute is a quantitative or qualitative description of that feature.
- Vectors are points, lines and polygon. A raster is a collection of pixels.
- Types of spatial operation
- Both types of outputs provide valuable information, but we should be aware of certain limitations for example – vector-based choropleths can result in a severe form of bias known as Ecological Fallacy
- Raster outputs on the other hand are dependent pixel size. Two can use the same data and yield different results due to this artefact.

Any questions?