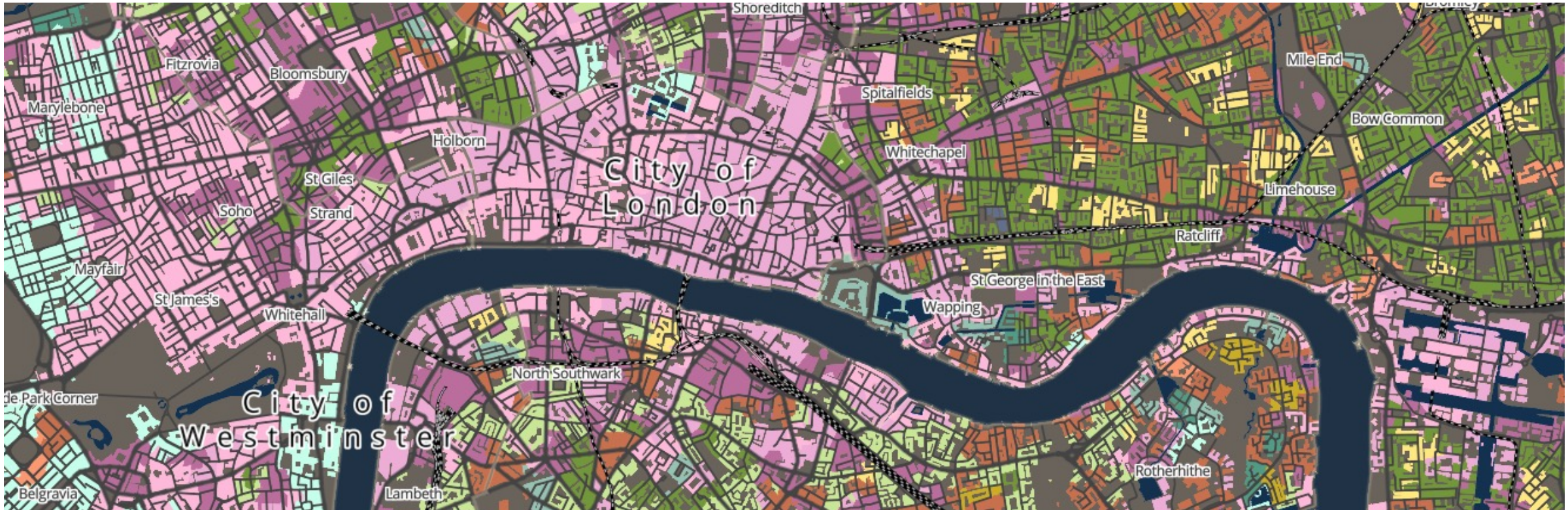
Geocomputation

GIScience and GIS software



UCL

Where are we at?

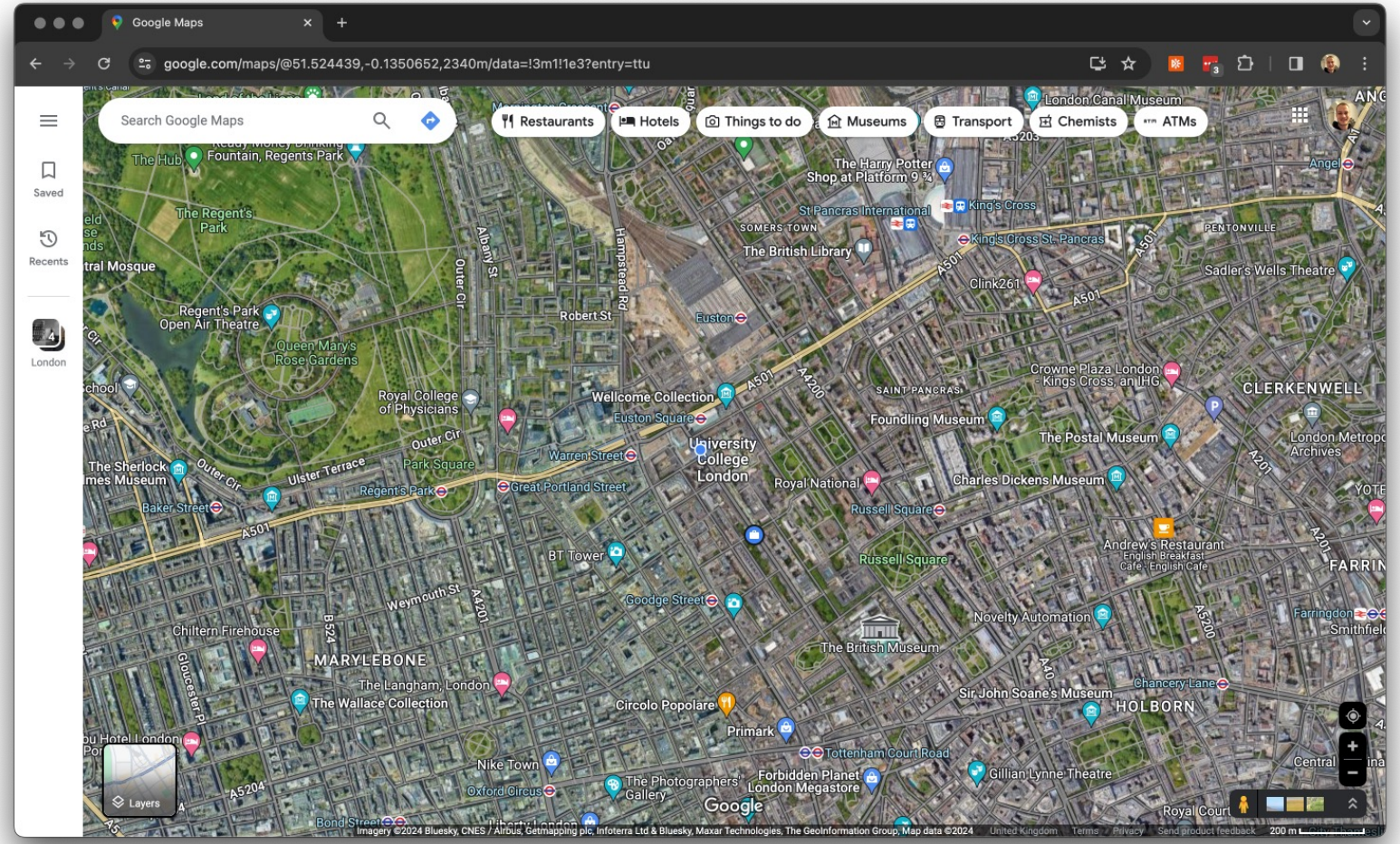
Part I: Foundational Concepts

W1	Geocomputation: An Introduction	
W2	**GIScience and GIS software**	QGIS
W3	Cartography and Visualisation	QGIS
W4	Programming for Data Analysis	R
W5	Programming for Spatial Analysis	R

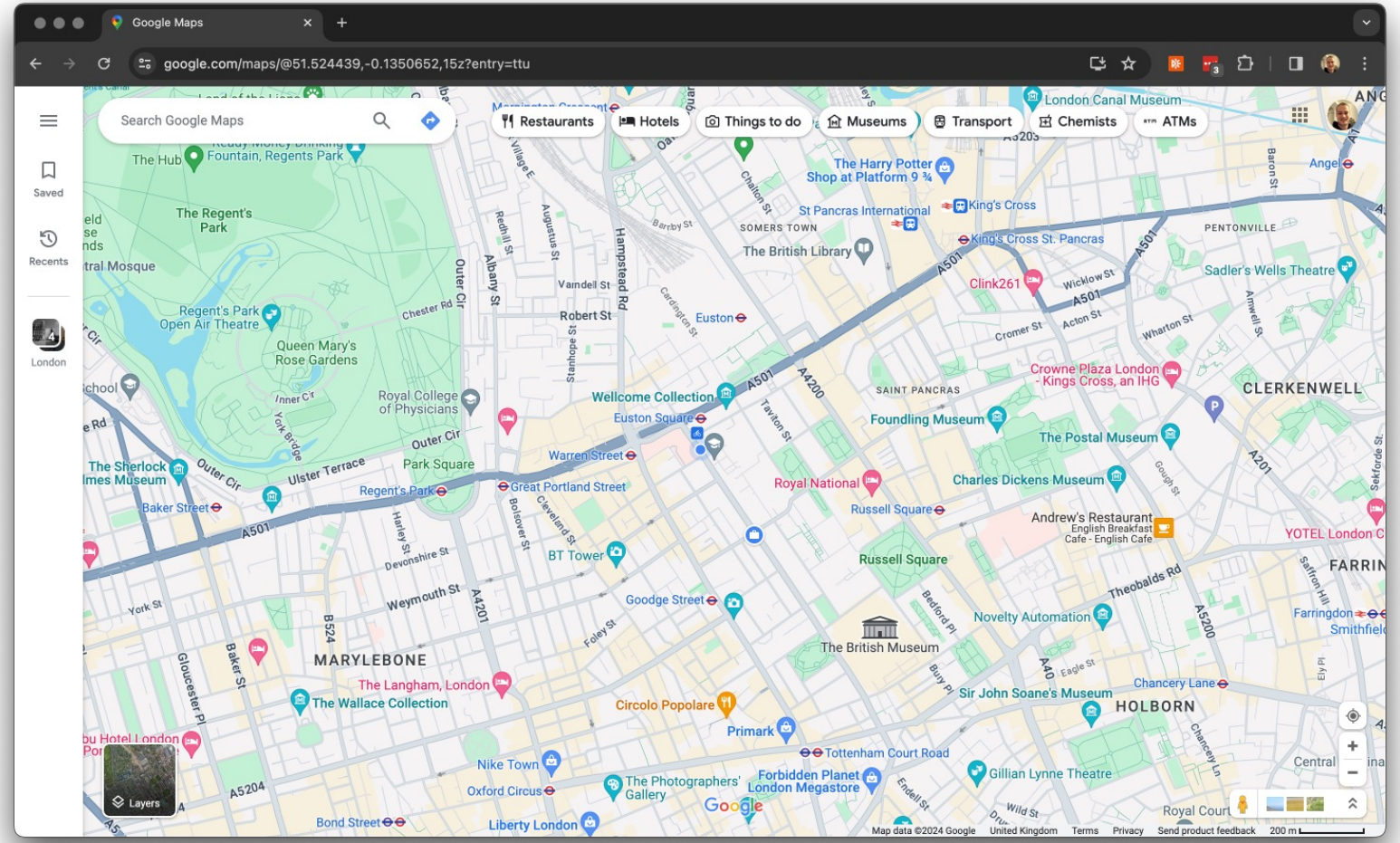
This week

- Digitally representing spatial data within GIS software
  - *Vector* data
  - *Raster* data
- GIS file types

Spatial modelling and digital representation



Spatial modelling and digital representation



Father of GIS

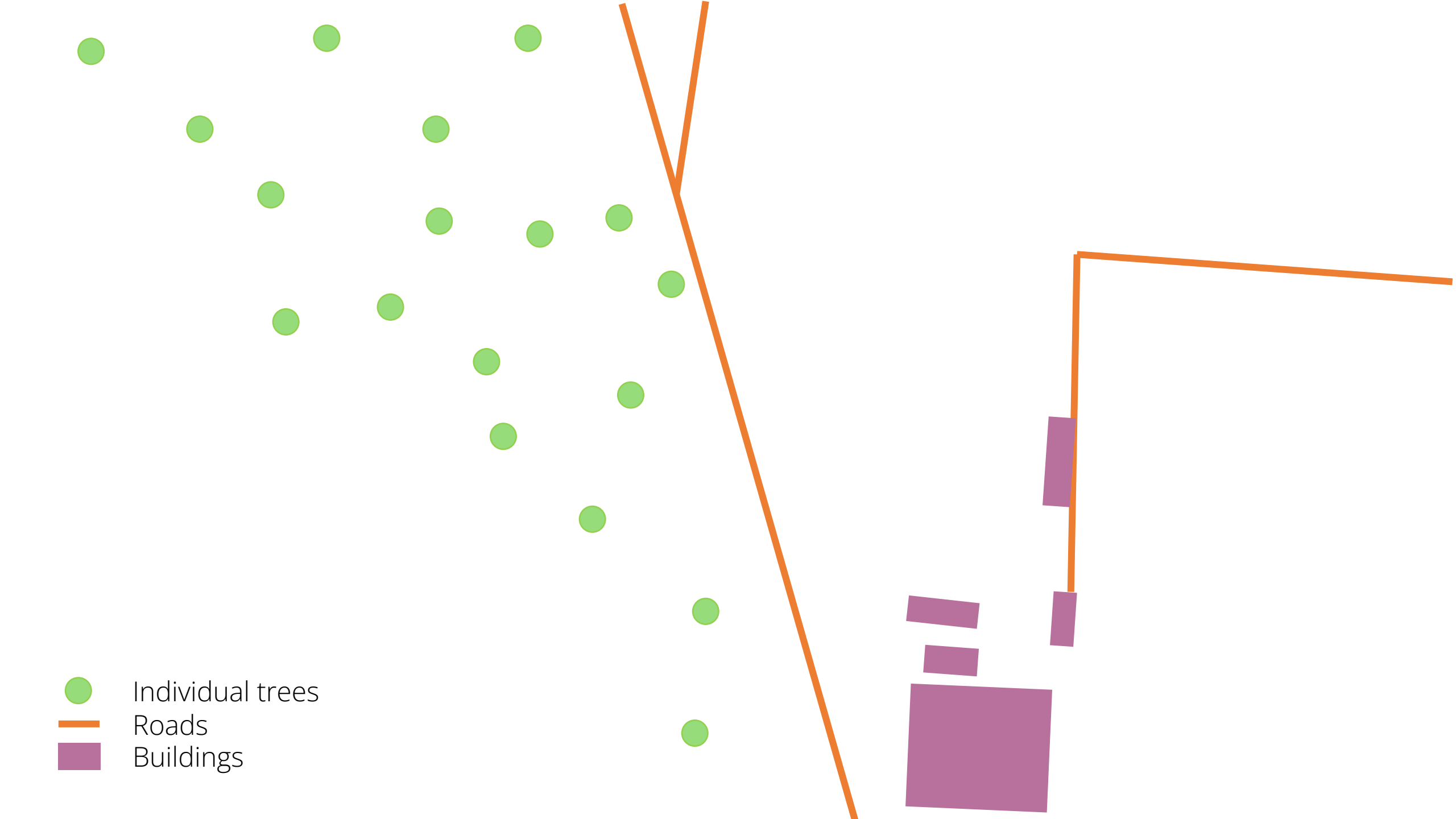
- Roger F. Tomlinson (1933-2014)
- Ph.D. dissertation: "*The application of electronic computing methods and techniques to the storage, compilation, and assessment of mapped data*" (1962, UCL).
- Conceived the idea of analysing multiple layers of spatial data within a single environment as well how to represent such spatial data in a digital format.

GIS data models

- GIScience requires spatial information to be represented in a digital format
- Traditionally, geographic information is represented in two ways:

  vector: a finite set of discrete geometric objects

  raster: images representing a surface (values, colours)


Individual trees
Roads
Buildings

Features

The objects created 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 textual observation.

Types of features

- Individual tree: type of tree, height, width. Point feature.
- Roads: type of road, length of road, speed limit. Polyline feature.
- Buildings: type of building (commercial, residential), number of people living in the building, number of stories. Polygon feature.

Vector data

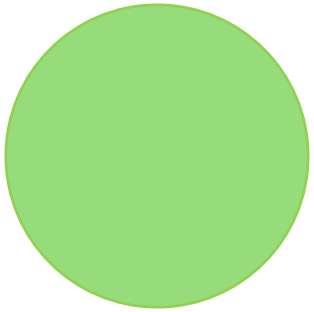
- The objects that we created of discrete entities are called vector data.
- Three types of vector data: point vectors, polylines or line vectors, and polygon vectors.

Point vector

Characteristics of a point vector in a GIS data model:

- Single XY location (coordinate)
- Has no area
- Has no length
- Geometry consists of a single node or vertex
- Used for: discrete features or ‘events’

Point vector



Point vector

Point vector

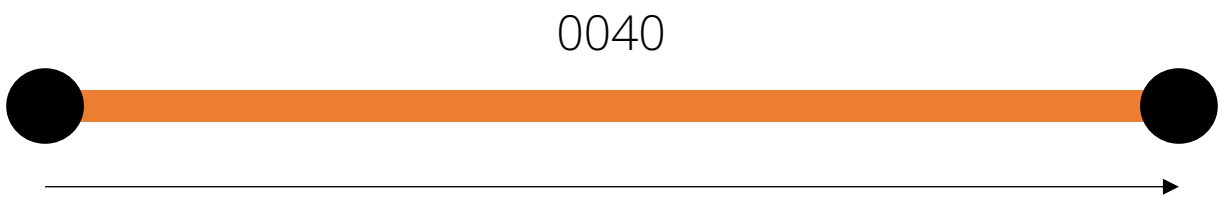
FeatureID	Type	Height
0030	Ent	500

Polyline vector

Characteristics of a polyline vector in a GIS data model:

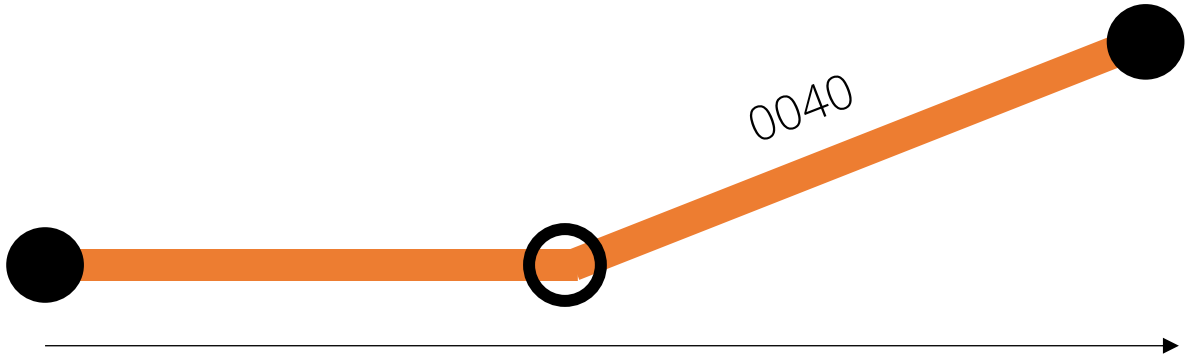
- Series of XY locations (coordinates) that form a line
- Has no area
- Has a length
- Has a direction (importance when it comes to roads, rivers, etc.)
- Can be connected to other polyline vectors to form a network
- Geometry consists of two nodes (start node and end node) and can have one or more vertices
- Used for: features without an area but with a length

Polyline vector



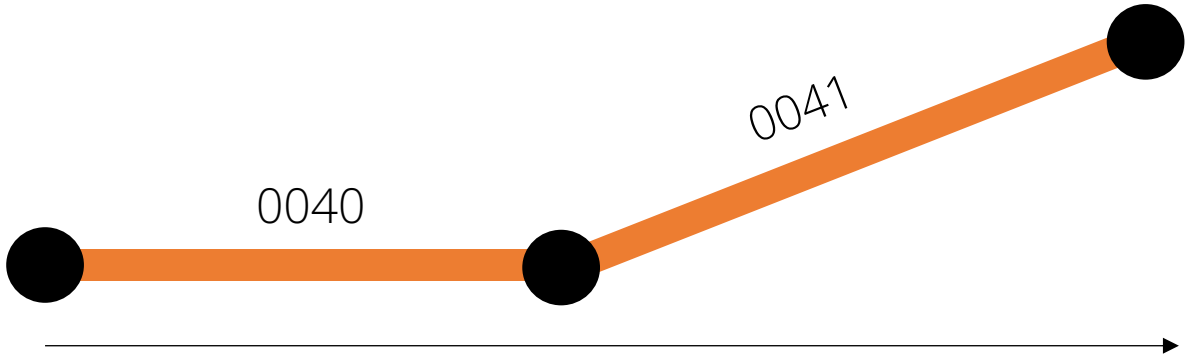
FeatureID	Type	Length
0040	Bicycle lane	1,500

Polyline vector



FeatureID	Type	Length
0040	Bicycle lane	1,650

Polyline vector



FeatureID	Type	Length
0040	Bicycle lane	600
0041	Bicycle lane	1,050

Polygon vector

Characteristics of a polyline vector in a GIS data model:

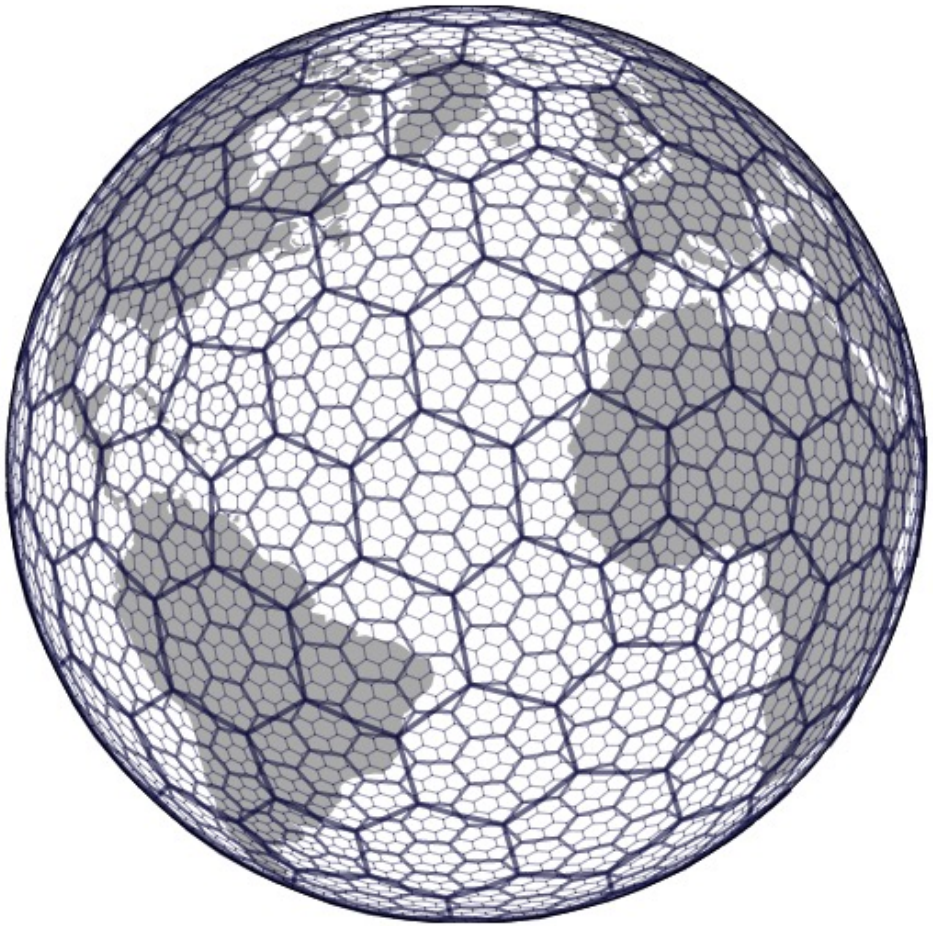
- Series of XY locations (coordinates) to form an enclosed region
- Has an area
- Has no length, but does have a perimeter
- Geometry consists of at least three nodes or vertices whereby the first node or vertex connects with the last one.
- Used for: features with enclosed regions such as buildings and administrative areas

Polygon vector

FeatureID	Type	Area
0050	University building	2000

Vector

Vector



Uber. 2018. *H3: Uber's Hexagonal Hierarchical Spatial Index*. [online] https://eng.uber.com/h3/

Attribute joins

GeoID	Population
GEO0030	540
GEO0031	320

Attribute joins

FeatureID	GeoID
0050	GEO0030

Attribute joins

FeatureID	GeoID
0050	GEO0030

GeoID	Population
GEO0030	540
GEO0031	320

Attribute joins

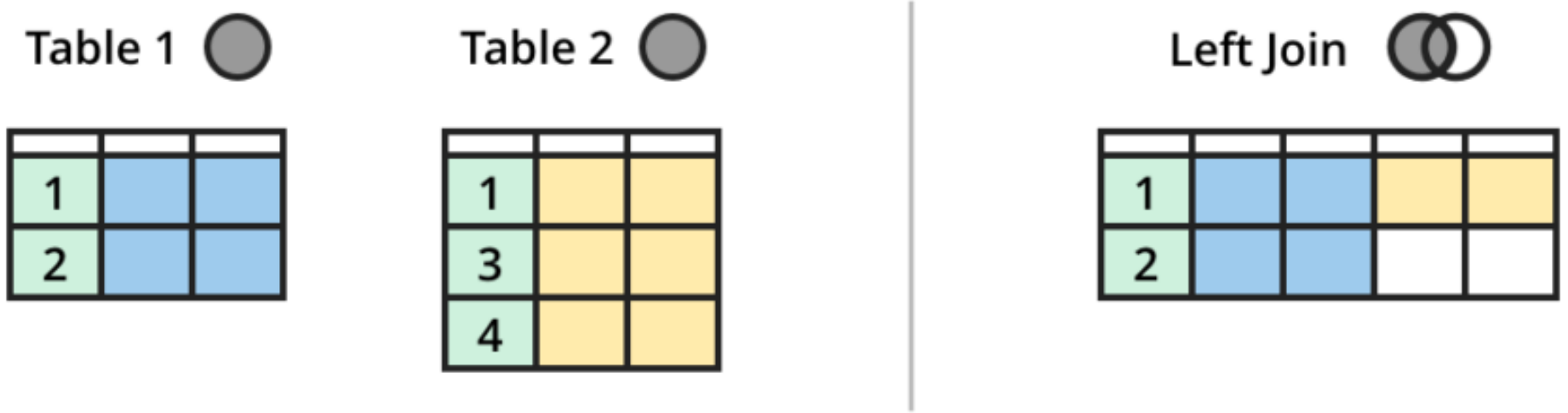
FeatureID	GeoID
0050	GEO0030

GeoID	Population
GEO0030	540
GEO0031	320

Attribute joins

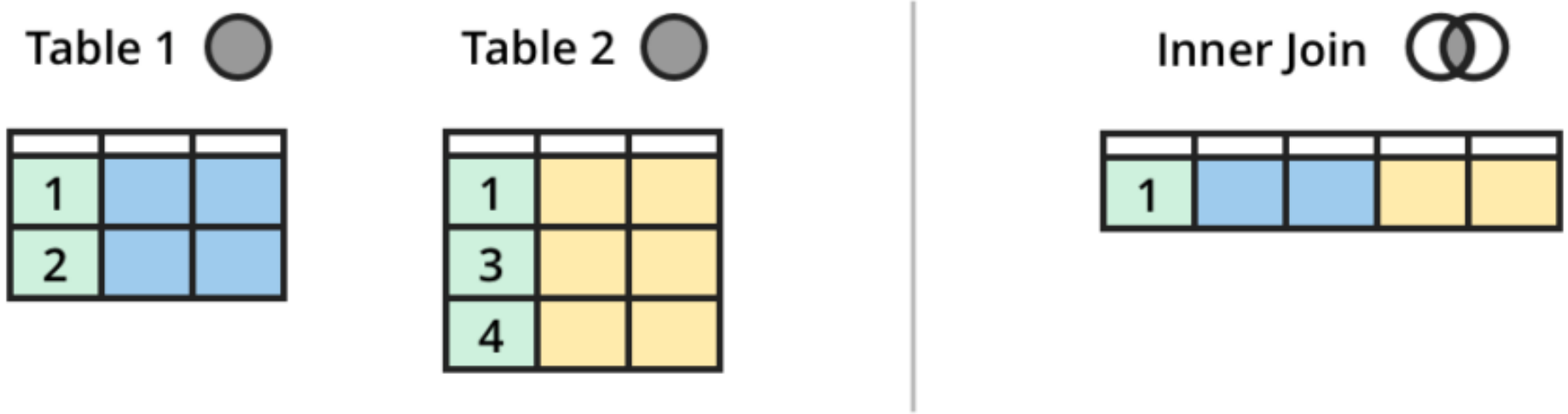
FeatureID	GeoID	Population
0050	GEO0030	540

Left joins



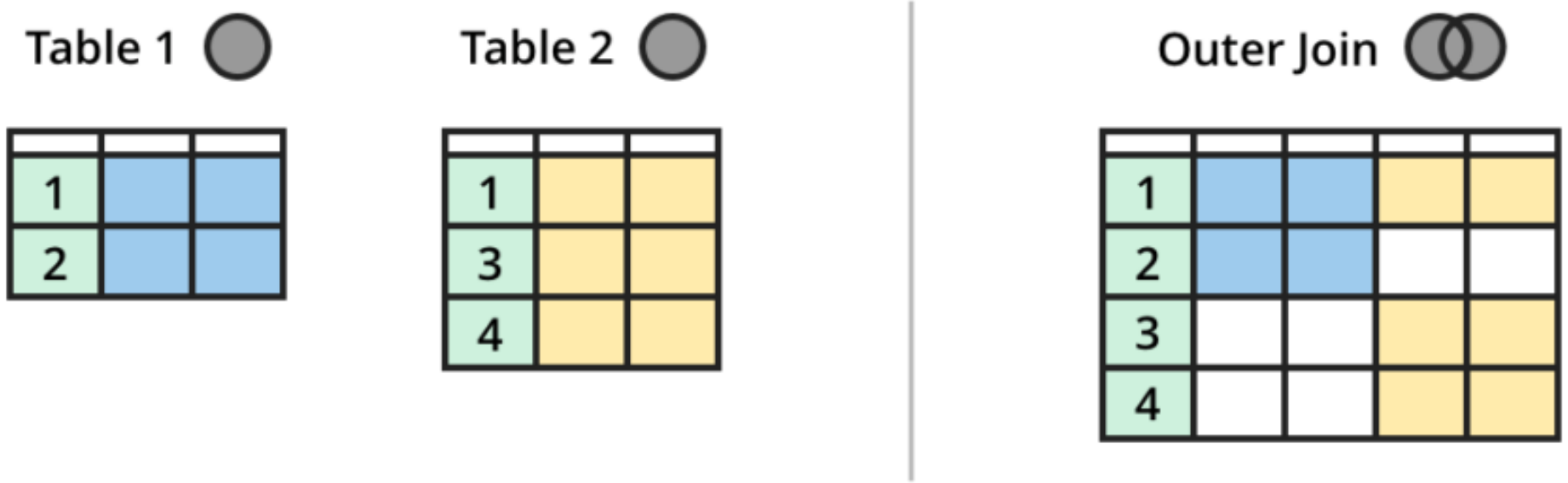
The Data School. 2019. *SQL Join Types Explained Visually*.

Inner joins



The Data School. 2019. *SQL Join Types Explained Visually*.

Outer joins

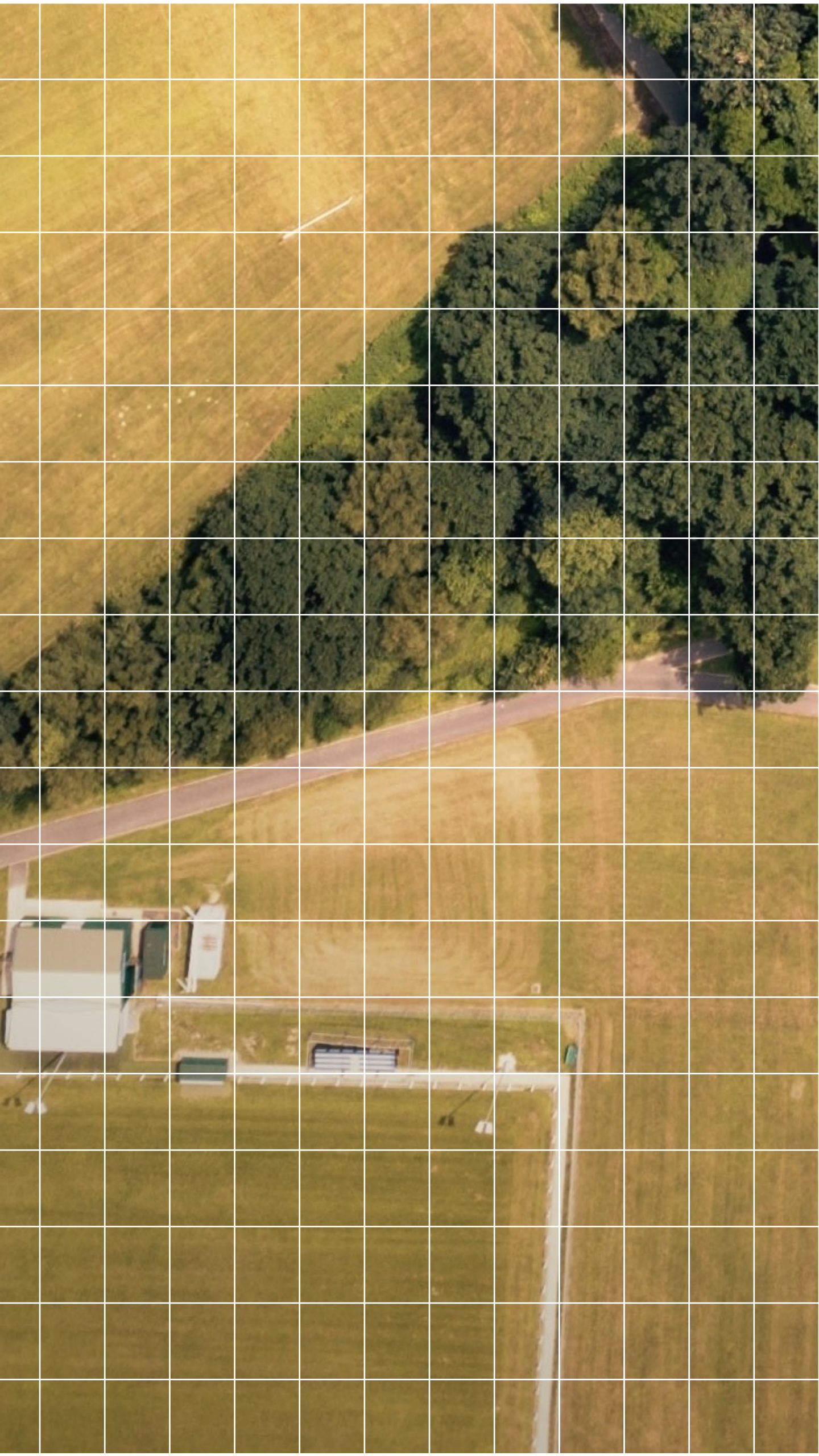


The Data School. 2019. *SQL Join Types Explained Visually*.

GIS data models

Traditionally, geographic information is represented in two ways:

- vector: a finite set of geometric objects
- raster: images representing a surface (values, colours)



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

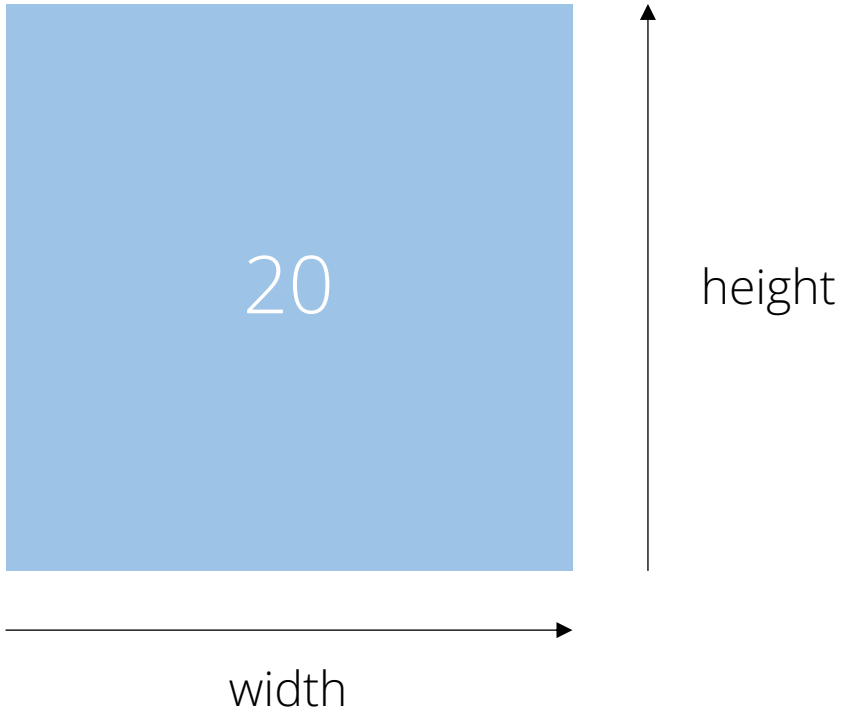
Raster data

- Unlike the vector data. The above feature describes rainfall levels across the surface of the landscape – the feature is is not measured discretely but on a continuous surface to show gradient in changes.
- This non-discrete feature is classed a raster data.

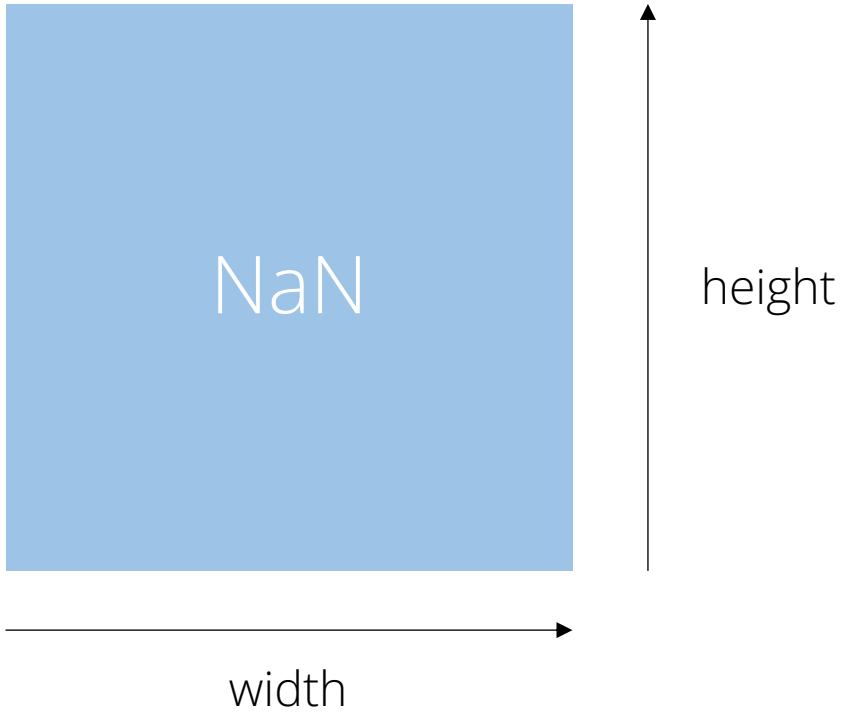
Raster data

- Raster data is represented by 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).

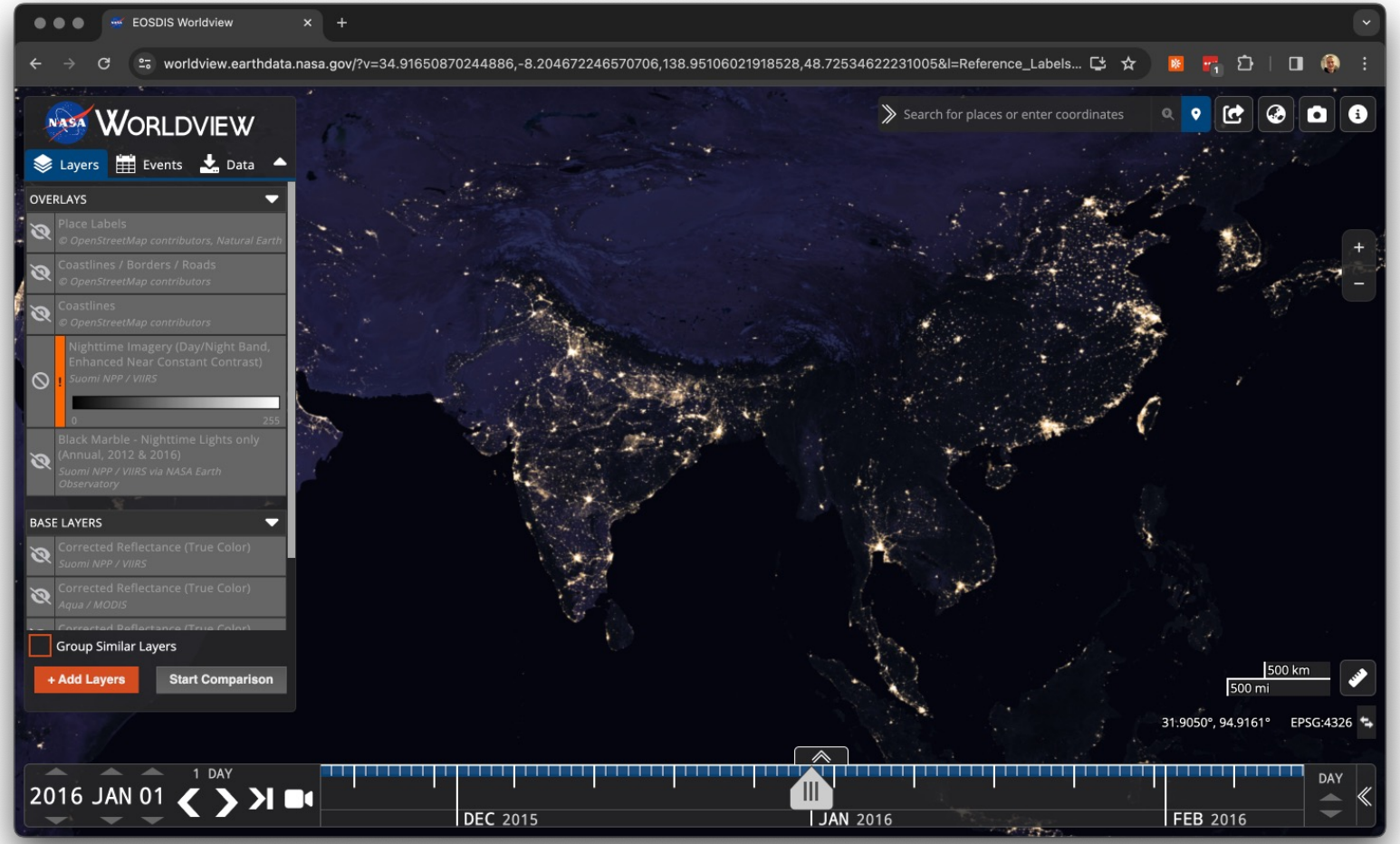
Raster data



Raster data



Raster data

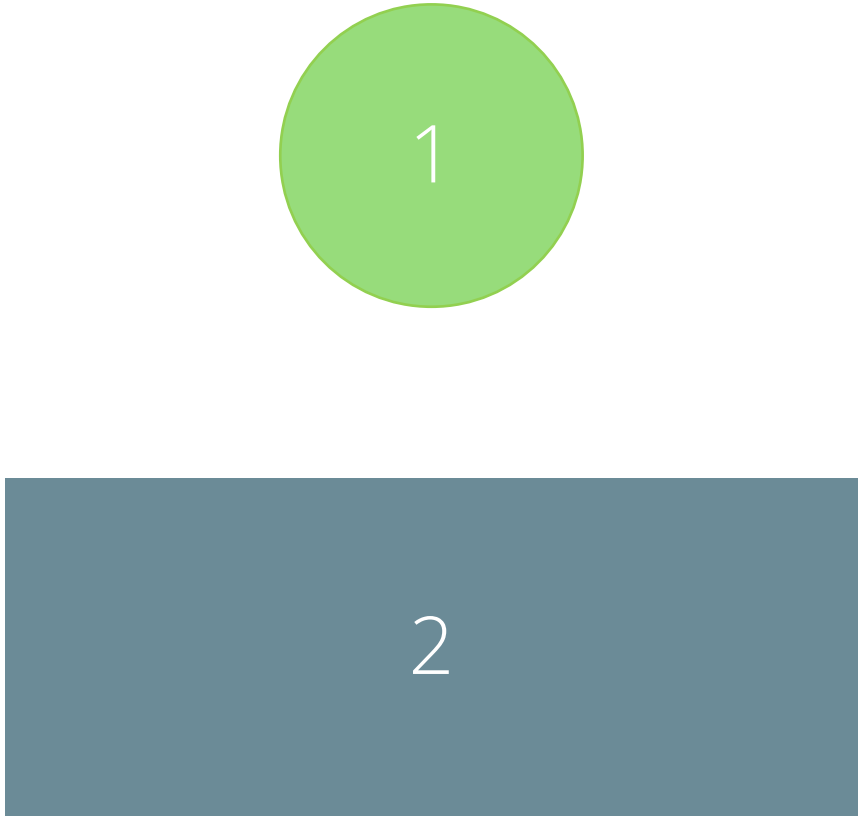

EOSDIS Worldview
worldview.earthdata.nasa.gov/?v=34.91650870244886,-8.204672246570706,138.95106021918528,48.72534622231005&l=Reference_Labels...
WORLDVIEW
Search for places or enter coordinates
Layers
Events
Data
OVERLAYS
Place Labels
© OpenStreetMap contributors, Natural Earth
Coastlines / Borders / Roads
© OpenStreetMap contributors
Coastlines
© OpenStreetMap contributors
Nighttime Imagery (Day/Night Band, Enhanced Near Constant Contrast)
Suomi NPP / VIIRS
0
255
Black Marble - Nighttime Lights only (Annual, 2012 & 2016)
Suomi NPP / VIIRS via NASA Earth Observatory
BASE LAYERS
Corrected Reflectance (True Color)
Suomi NPP / VIIRS
Corrected Reflectance (True Color)
Aqua / MODIS
Group Similar Layers
+ Add Layers
Start Comparison
500 km
500 mi
31.9050°, 94.9161°
EPSG:4326
1 DAY
2016 JAN 01
DEC 2015
JAN 2016
FEB 2016
DAY

Sample scheme

- Both for the vector and raster data model real life features are ‘sampled’ or represented in a certain way. How to represent the spatial information? With which level of detail?
- It must be fine enough to provide general consistency in our feature or field as well as accurately represent its distribution.
- It must be fine enough also to capture the important changes in our feature or field, e.g. a turn in a road, or a certain measurement change in a variable, such as temperature or rainfall.
- But we must also not over sample – we need to consider efficiency and efficacy of our sampling as we collect the data and store it digitally.

Sample scheme

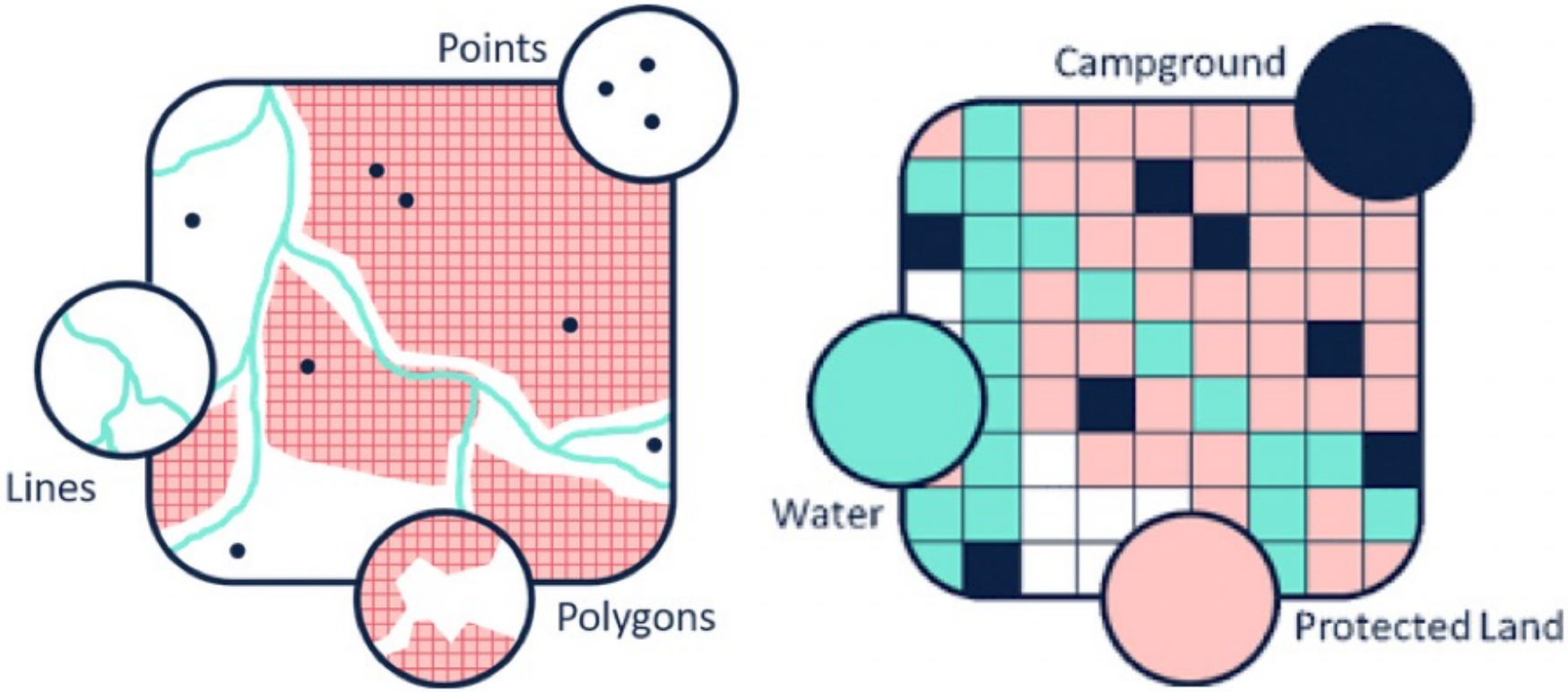
Representing the Medical Sciences and Anatomy Building



Vector versus raster

- Why not vectorise everything? Can add as many attributes as we like?
- ”Raster is faster but vector is corrector”
- Depends on your application and intended analysis.
- Vector and raster data both have their advantages and disadvantages.

Vector versus raster



Vector versus raster

	Advantages	Disadvantages
Raster	Map Algebra with raster data is usually quick and easy to perform	Linear features and paths are difficult to display
	Some specific use cases can only be achieved with raster data (e.g. modeling water flow over the land surface)	Subject to a pixelated look and feel
		Datasets can become very large because they record values for each cell
Vector	Graphical output is generally more aesthetically-pleasing	Continuous data is poorly stored and displayed
	Higher geographic accuracy because data isn't dependent on grid size	Needs a lot of work and maintenance to ensure that it is accurate and reliable

Spatial data formats

- There are a number of commonly used file formats that store vector and raster data, some of which you will come across during this course and it is important to understand what they are, how they represent data and how you can use them.
- Different file formats for vector data and raster data.
- Common vector formats: shapefile, GeoJSON, GeoPackage
- Common raster formats: GeoTIFF, GeoPackage

Shapefiles

- Perhaps the most (in)famous file format.
- Widely used, despite being outdated, especially limitations of `.dbf` format.
- A shapefile is not a single file, but a collection of files of which at least three are needed for the data to be displayed in GIS software.

Shapefiles

- `.shp` contains the feature geometry. *Mandatory.*
- `.shx` index file which stores the position of the feature's ID in the `.shp` file. *Mandatory.*
- `.dbf` stores alle attribute information associated with the records. *Mandatory.*
- `.prj` contains the coordinate system information and projection. *Optional but not really.*
- `.xml` general metadata. *Optional.*
- `.cpg` encoding information. *Optional.*
- `.sbn` optimisation file for spatial queries. *Optional.*

Shapefiles

GeoJSON

- GeoJSON (Geospatial Data Interchange format for JavaScript Object Notation) is becoming an increasingly popular spatial data file, particularly for web-based mapping as it is based on JavaScript Object Notation.
- Unlike a shapefile in a GeoJSON, the attributes, boundaries and projection information are all contained in the same file.
- How to spot in the wild: `.json` or .geojson

GeoJSON

- Point
```
{"type": "Point", "coordinates": [30.0, 10.0]}
```
- LineString
```
{"type": "LineString", "coordinates": [[30.0, 10.0], [10.0, 30.0], [40.0, 40.0] ]}
```
- Polygon
```
{"type": "Polygon", "coordinates": [[[30.0, 10.0], [40.0, 40.0], [20.0, 40.0], [10.0, 20.0], [30.0, 10.0]]]}
```

GeoPackage

- A GeoPackage is an open, standards-based, platform-independent, portable, self-describing, compact format for transferring geospatial data.
- It stores spatial data layer as a single file, based upon an SQLite database.
- How to spot in the wild: .gpkg

GeoTIFF

- Geostationary Earth Orbit Tagged Image File Format.
- Created by NASA and is a standard public domain format.
- All necessary information to establish the location of the data on Earth’s surface is embedded into the image. This includes all details on map projection.
- How to spot in the wild: .tiff

GI file formats and GI systems

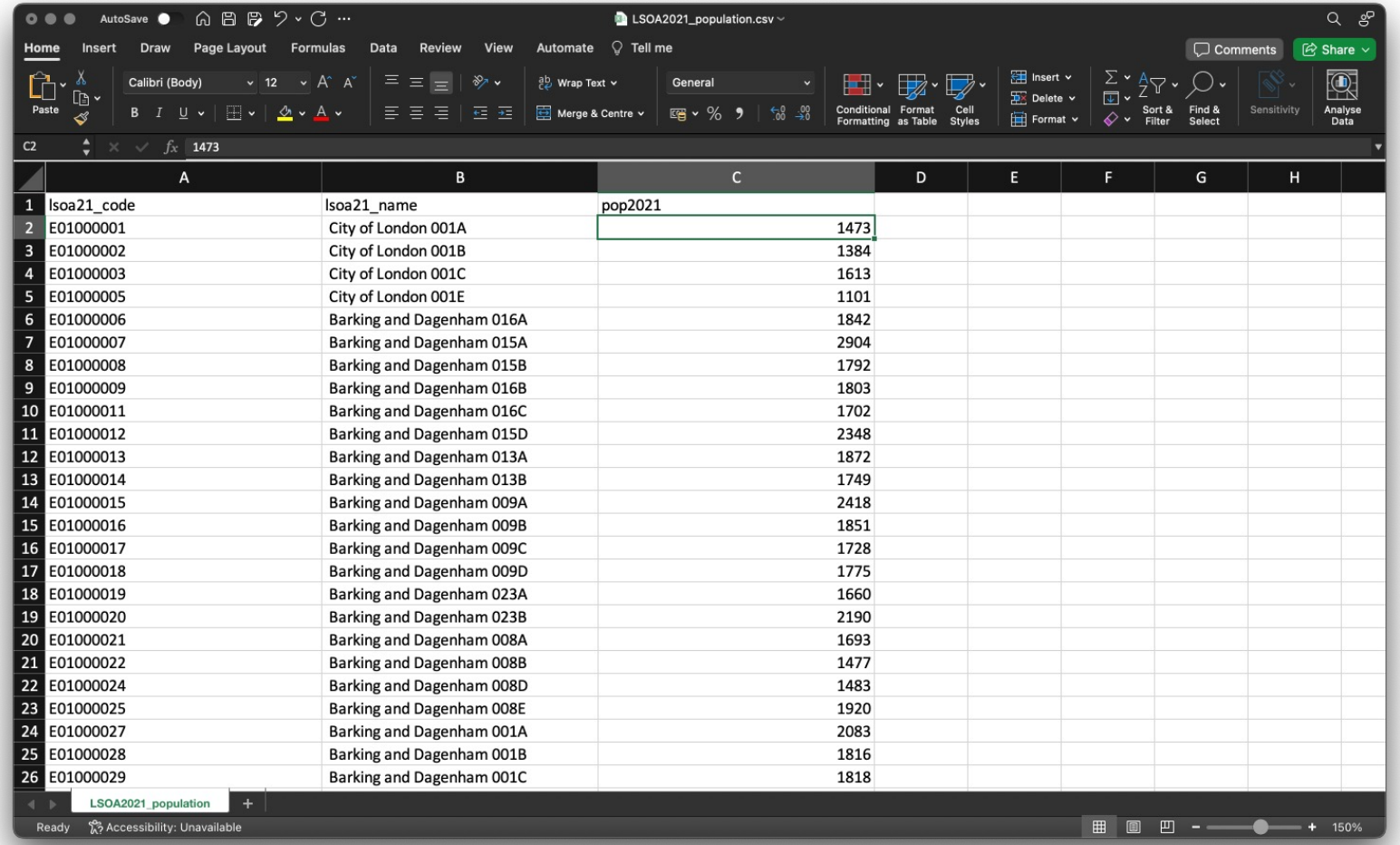
- Standard GI systems can read most file types without any problems or need for conversion.
- When using a programming language it is sometimes necessary to use a dedicated function to read in the data – so important to know which format you are dealing with (but libraries / packages exist to do just that).

On the topic of file formats: to join attribute data to your spatial files: .csv

CSV files

- “Comma Separated Values” or “Character Separated Values”.
- Format to store tabular data in rows and columns.
- Plain text rather than a binary file (e.g. Microsoft Excel).
- No limits on number of rows, columns, cells.

CSV files



AutoSave LSOA2021_population.csv

Home Insert Draw Page Layout Formulas Data Review View Automate Tell me Comments Share

Paste Calibri (Body) 12 Wrap Text General Merge & Centre Conditional Formatting Format as Table Cell Styles Insert Delete Format Sort & Filter Find & Select Sensitivity Analyse Data

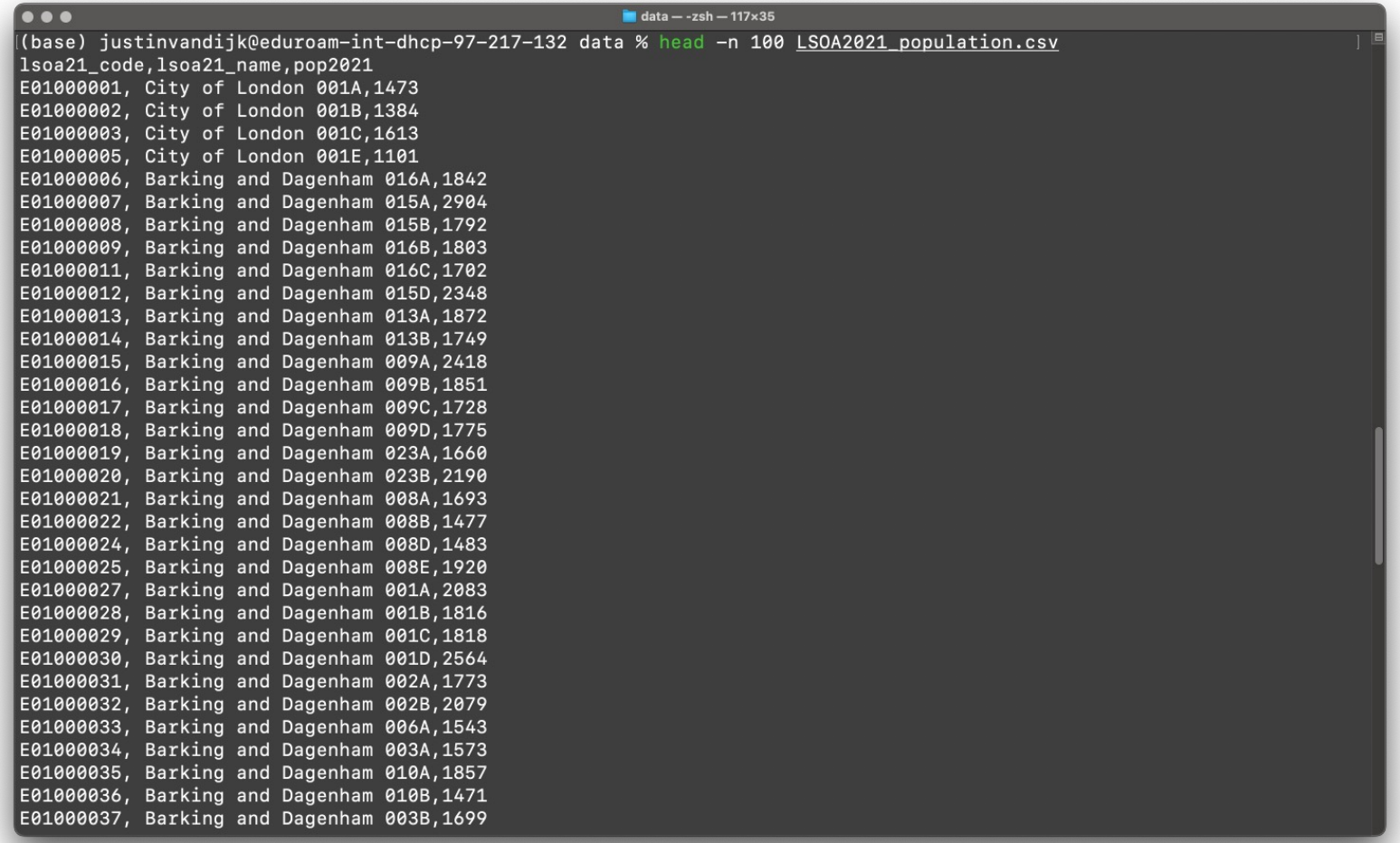
C2 1473

	A	B	C	D	E	F	G	H
1	lsoa21_code	lsoa21_name	pop2021					
2	E01000001	City of London 001A	1473					
3	E01000002	City of London 001B	1384					
4	E01000003	City of London 001C	1613					
5	E01000005	City of London 001E	1101					
6	E01000006	Barking and Dagenham 016A	1842					
7	E01000007	Barking and Dagenham 015A	2904					
8	E01000008	Barking and Dagenham 015B	1792					
9	E01000009	Barking and Dagenham 016B	1803					
10	E01000011	Barking and Dagenham 016C	1702					
11	E01000012	Barking and Dagenham 015D	2348					
12	E01000013	Barking and Dagenham 013A	1872					
13	E01000014	Barking and Dagenham 013B	1749					
14	E01000015	Barking and Dagenham 009A	2418					
15	E01000016	Barking and Dagenham 009B	1851					
16	E01000017	Barking and Dagenham 009C	1728					
17	E01000018	Barking and Dagenham 009D	1775					
18	E01000019	Barking and Dagenham 023A	1660					
19	E01000020	Barking and Dagenham 023B	2190					
20	E01000021	Barking and Dagenham 008A	1693					
21	E01000022	Barking and Dagenham 008B	1477					
22	E01000024	Barking and Dagenham 008D	1483					
23	E01000025	Barking and Dagenham 008E	1920					
24	E01000027	Barking and Dagenham 001A	2083					
25	E01000028	Barking and Dagenham 001B	1816					
26	E01000029	Barking and Dagenham 001C	1818					

LSOA2021_population

Ready Accessibility: Unavailable 150%

CSV files


data — -zsh — 117×35
(base) justinvandijk@eduroam-int-dhcp-97-217-132 data % head -n 100 LSOA2021_population.csv
lsoa21_code,lsoa21_name,pop2021
E01000001, City of London 001A,1473
E01000002, City of London 001B,1384
E01000003, City of London 001C,1613
E01000005, City of London 001E,1101
E01000006, Barking and Dagenham 016A,1842
E01000007, Barking and Dagenham 015A,2904
E01000008, Barking and Dagenham 015B,1792
E01000009, Barking and Dagenham 016B,1803
E01000011, Barking and Dagenham 016C,1702
E01000012, Barking and Dagenham 015D,2348
E01000013, Barking and Dagenham 013A,1872
E01000014, Barking and Dagenham 013B,1749
E01000015, Barking and Dagenham 009A,2418
E01000016, Barking and Dagenham 009B,1851
E01000017, Barking and Dagenham 009C,1728
E01000018, Barking and Dagenham 009D,1775
E01000019, Barking and Dagenham 023A,1660
E01000020, Barking and Dagenham 023B,2190
E01000021, Barking and Dagenham 008A,1693
E01000022, Barking and Dagenham 008B,1477
E01000024, Barking and Dagenham 008D,1483
E01000025, Barking and Dagenham 008E,1920
E01000027, Barking and Dagenham 001A,2083
E01000028, Barking and Dagenham 001B,1816
E01000029, Barking and Dagenham 001C,1818
E01000030, Barking and Dagenham 001D,2564
E01000031, Barking and Dagenham 002A,1773
E01000032, Barking and Dagenham 002B,2079
E01000033, Barking and Dagenham 006A,1543
E01000034, Barking and Dagenham 003A,1573
E01000035, Barking and Dagenham 010A,1857
E01000036, Barking and Dagenham 010B,1471
E01000037, Barking and Dagenham 003B,1699

Conclusion

- Two GIS data models: the vector data model and the raster data model.
- The vector model uses points, line, and polygon segments to identify locations on the earth while the raster model uses a series of cells to represent locations on the earth.
- Both GIS data models accommodate attributes: the qualitative or quantitative descriptions of the feature.
- Per definition any data model is an incomplete representation of reality.
- GI systems have been designed to work with a variety of different file types.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk