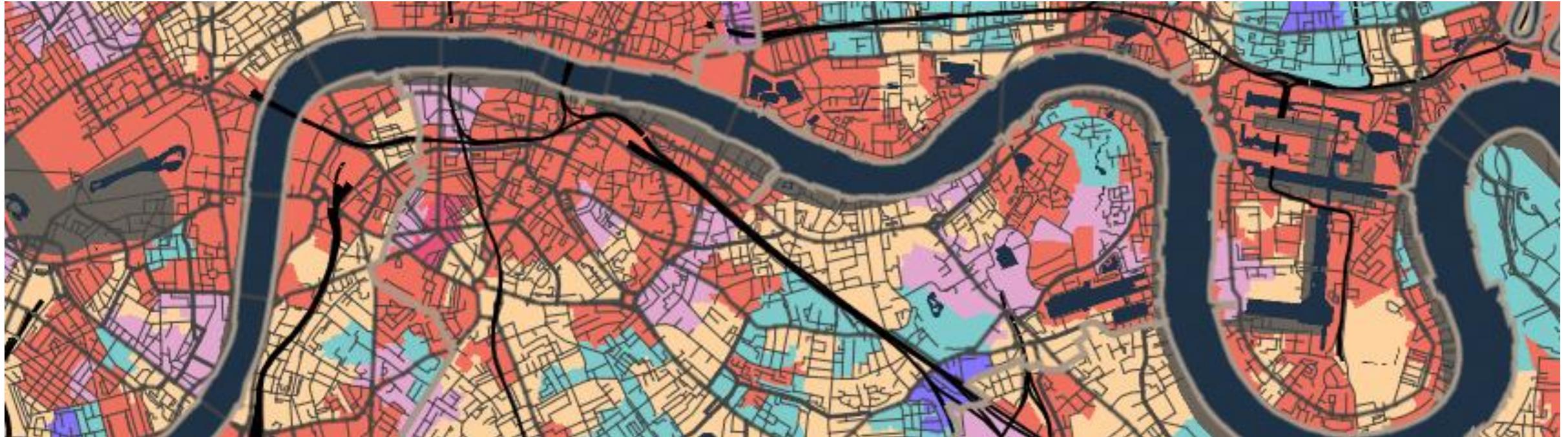


# SA-TIED Geospatial Analysis Workshop

## Overview



 Dr Justin van Dijk  
 j.t.vandijk@ucl.ac.uk



# Objectives

We will explore the following topics:

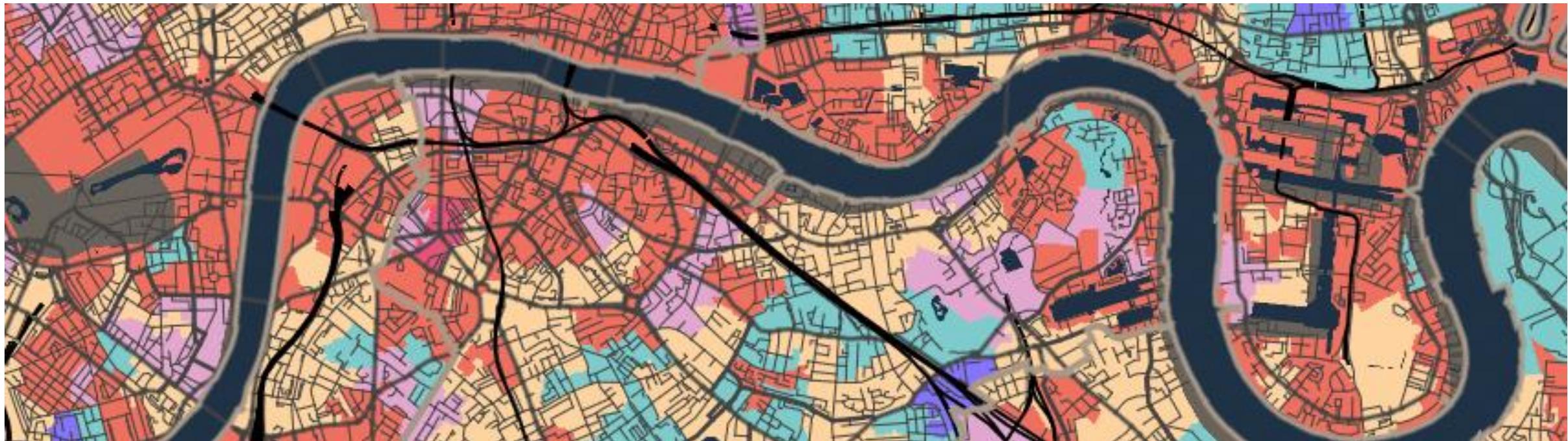
- Fundamentals of using R for data analysis.
- Creating thematic maps using R.
- Quantifying the degree of spatial dependence in a dataset.
- Incorporating space into statistical models.

# Schedule

Day 1 - Morning	R for Data Analysis
Day 1 - Afternoon	R for Spatial Analysis
Day 2 - Morning	Spatial Autocorrelation
Day 2 - Afternoon	Spatial Models

# SA-TIED Geospatial Analysis Workshop

## S02 – R for Spatial Analysis



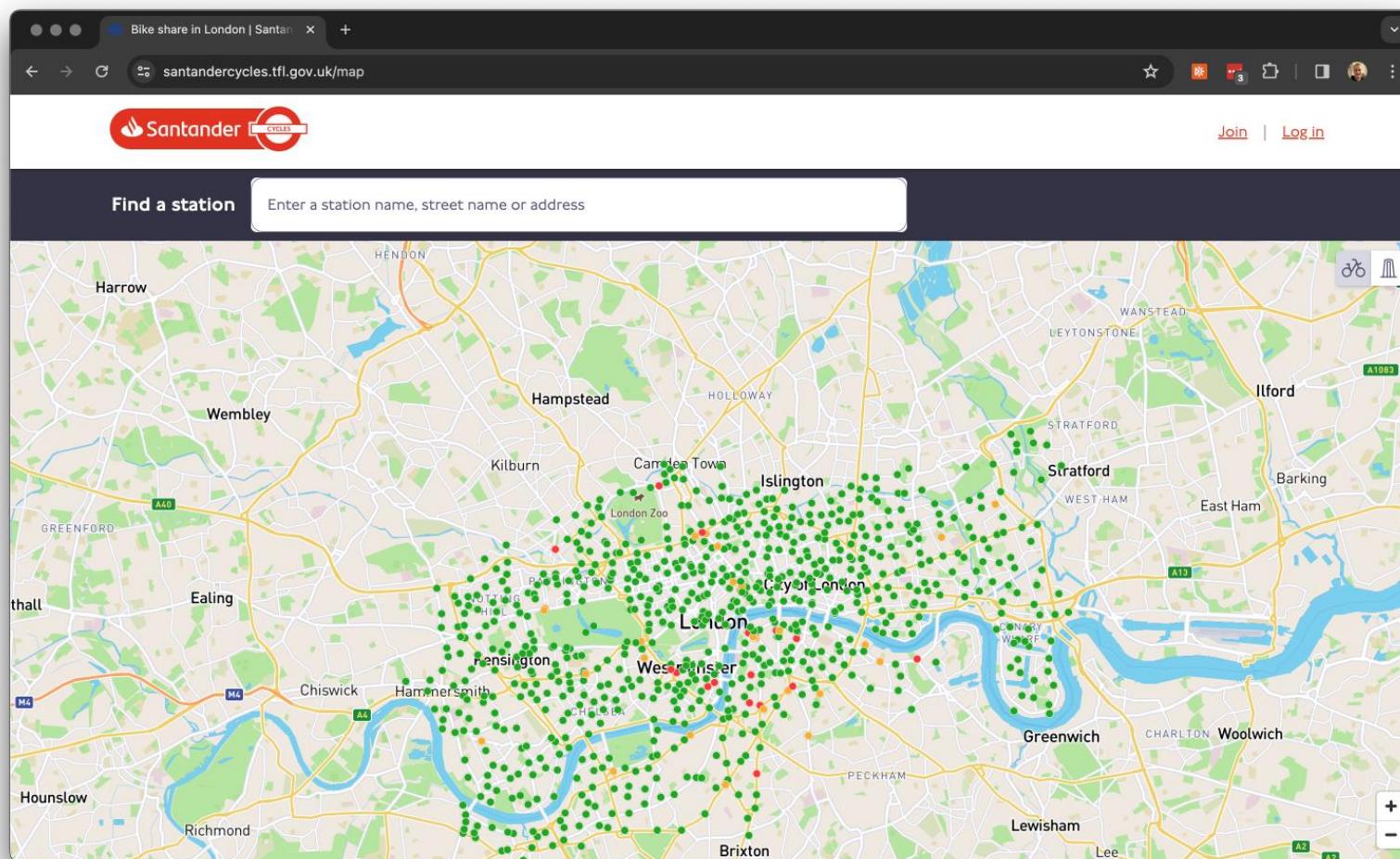
# This session

- Why is mapping important?
- How do GIScience and GISystems fit in?
- Mapping 101.

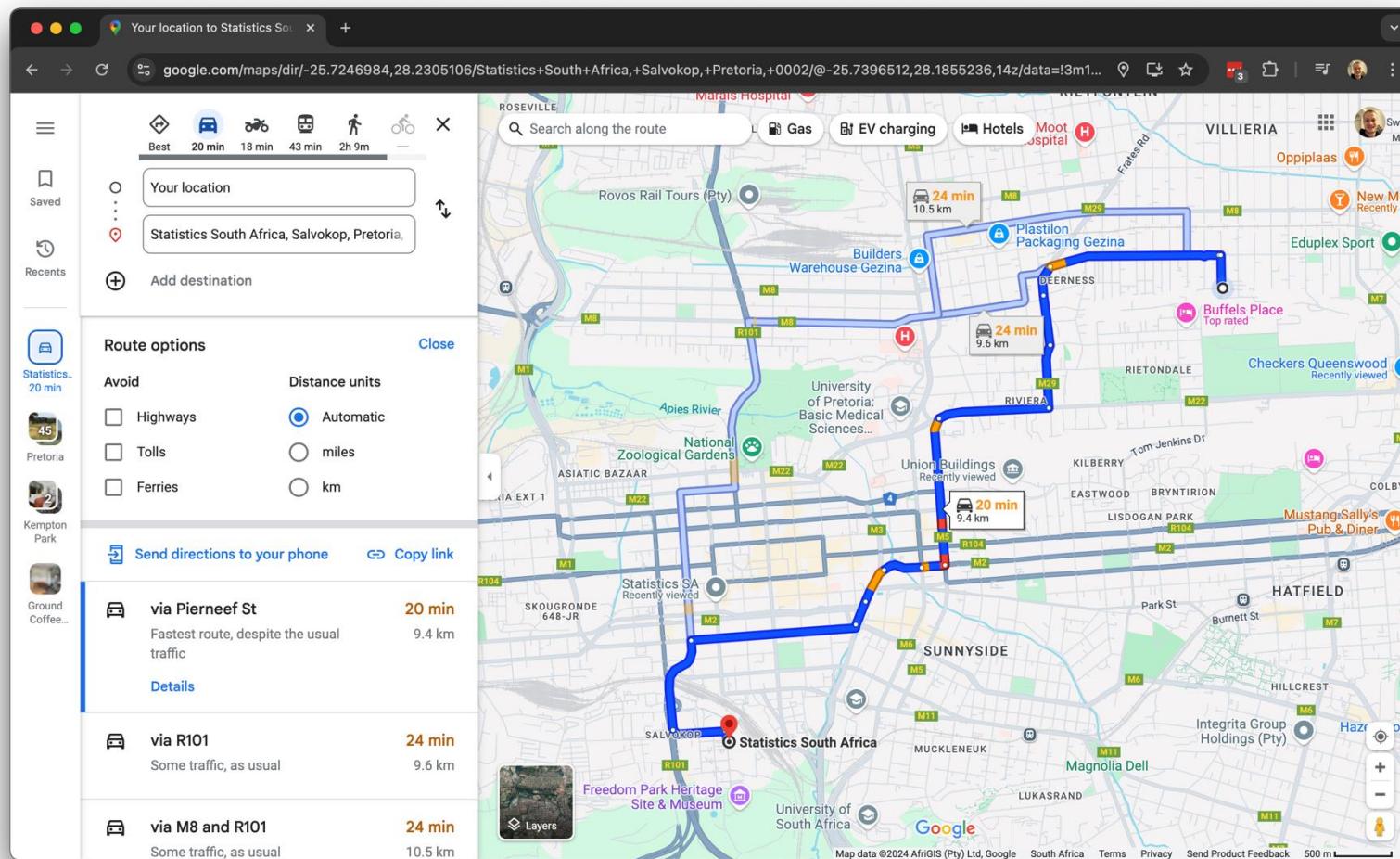
# Why is mapping important?

- Many challenges and problems are inherently geographical.
- Almost all data is tied to specific locations and exhibits spatial variability.
- Helps answer the question: What is this place like compared to other places?
- Translates numerical data into a format that is more intuitively understood by the human brain.

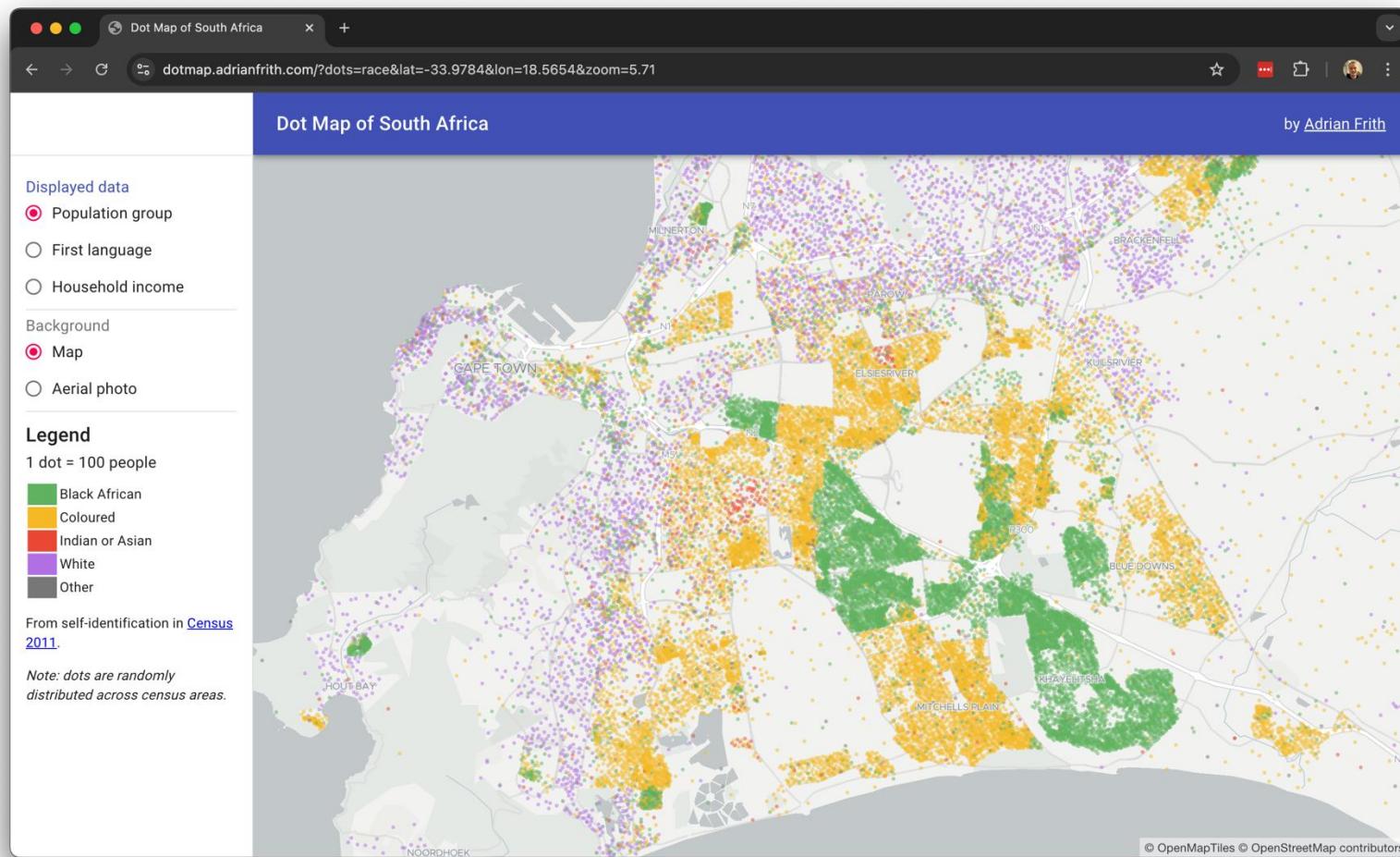
# Bicycle docking stations in London



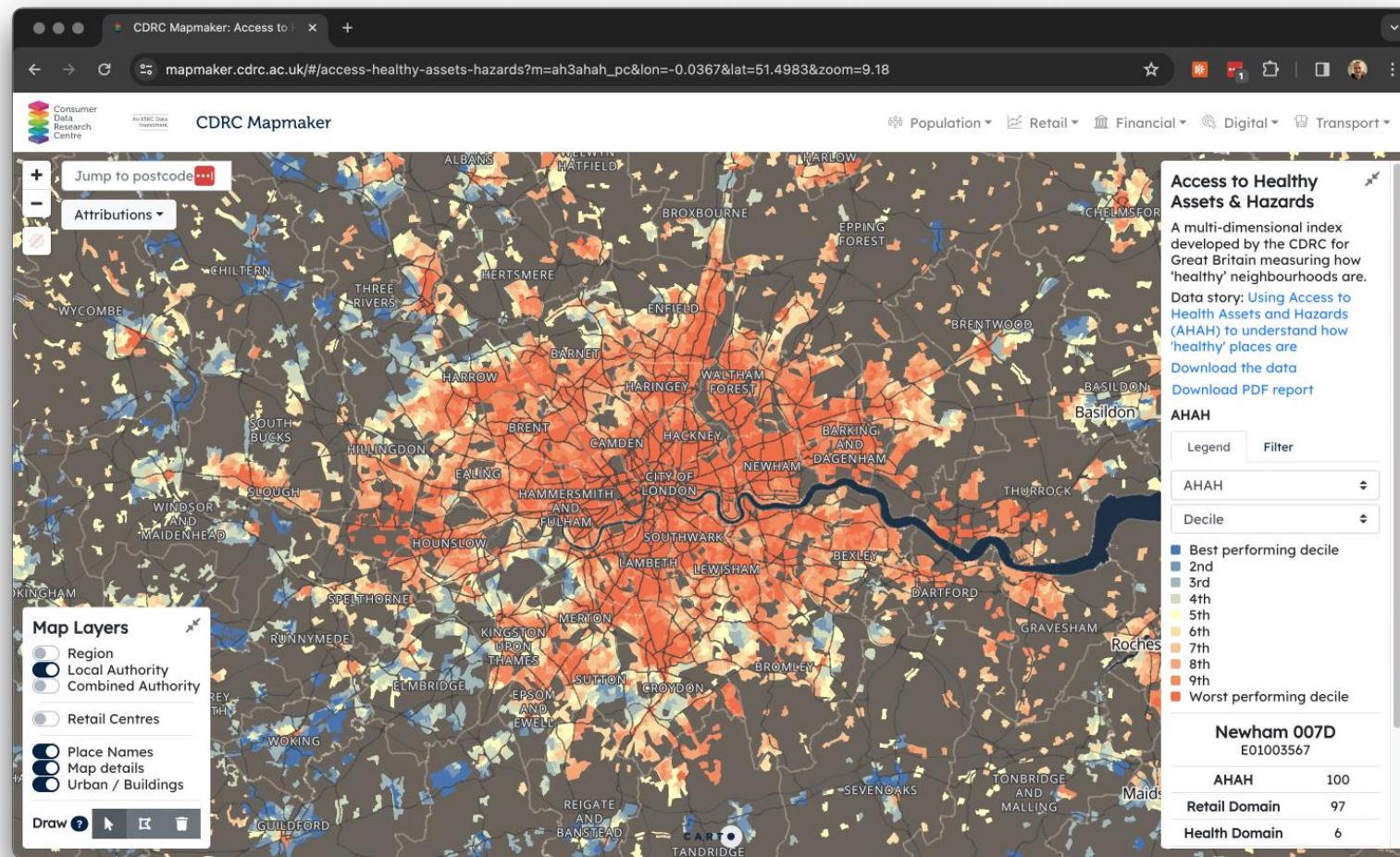
# Finding your way in Pretoria



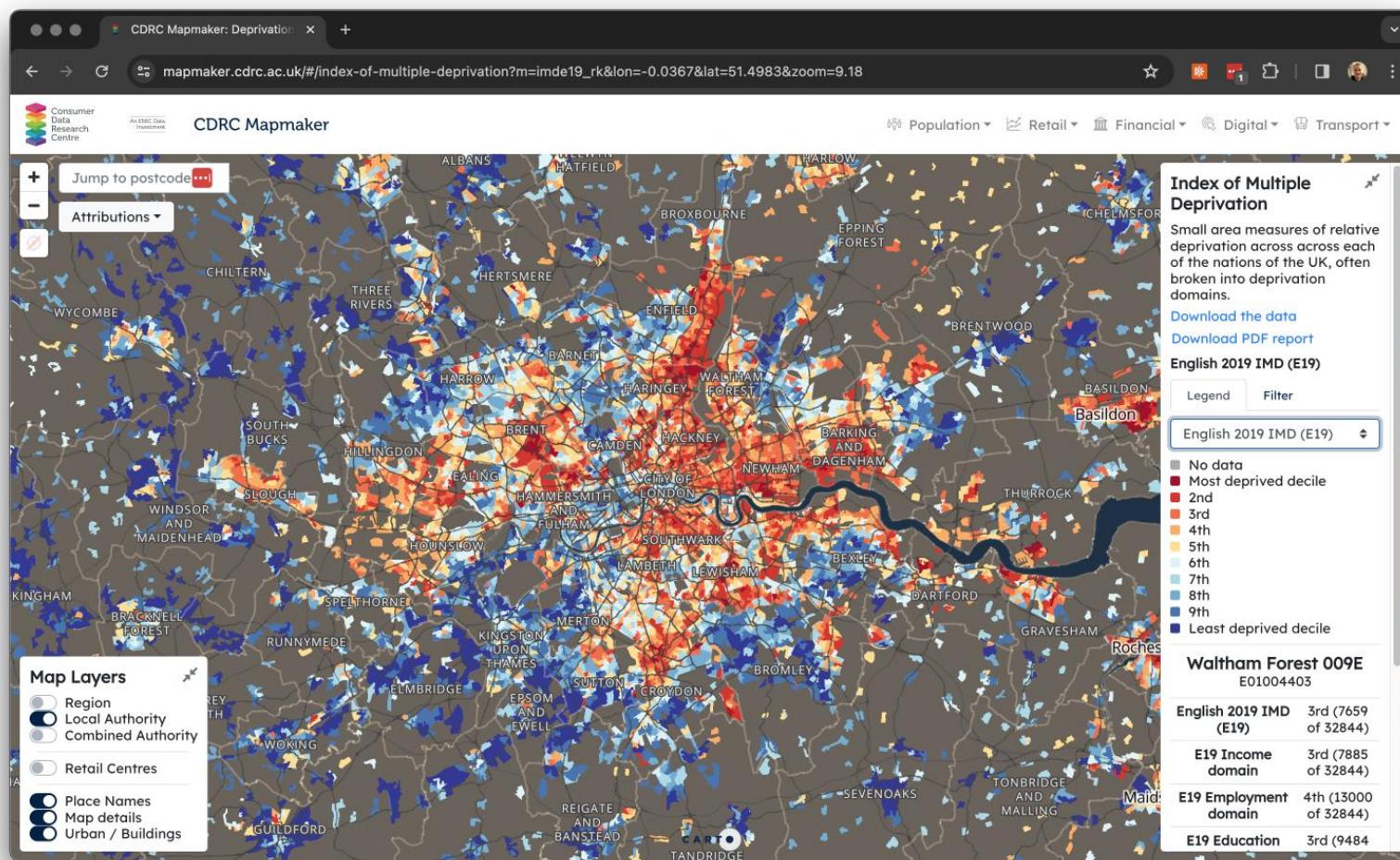
# Segregation in Cape Town



# Access to Healthy Assets and Hazards in the UK



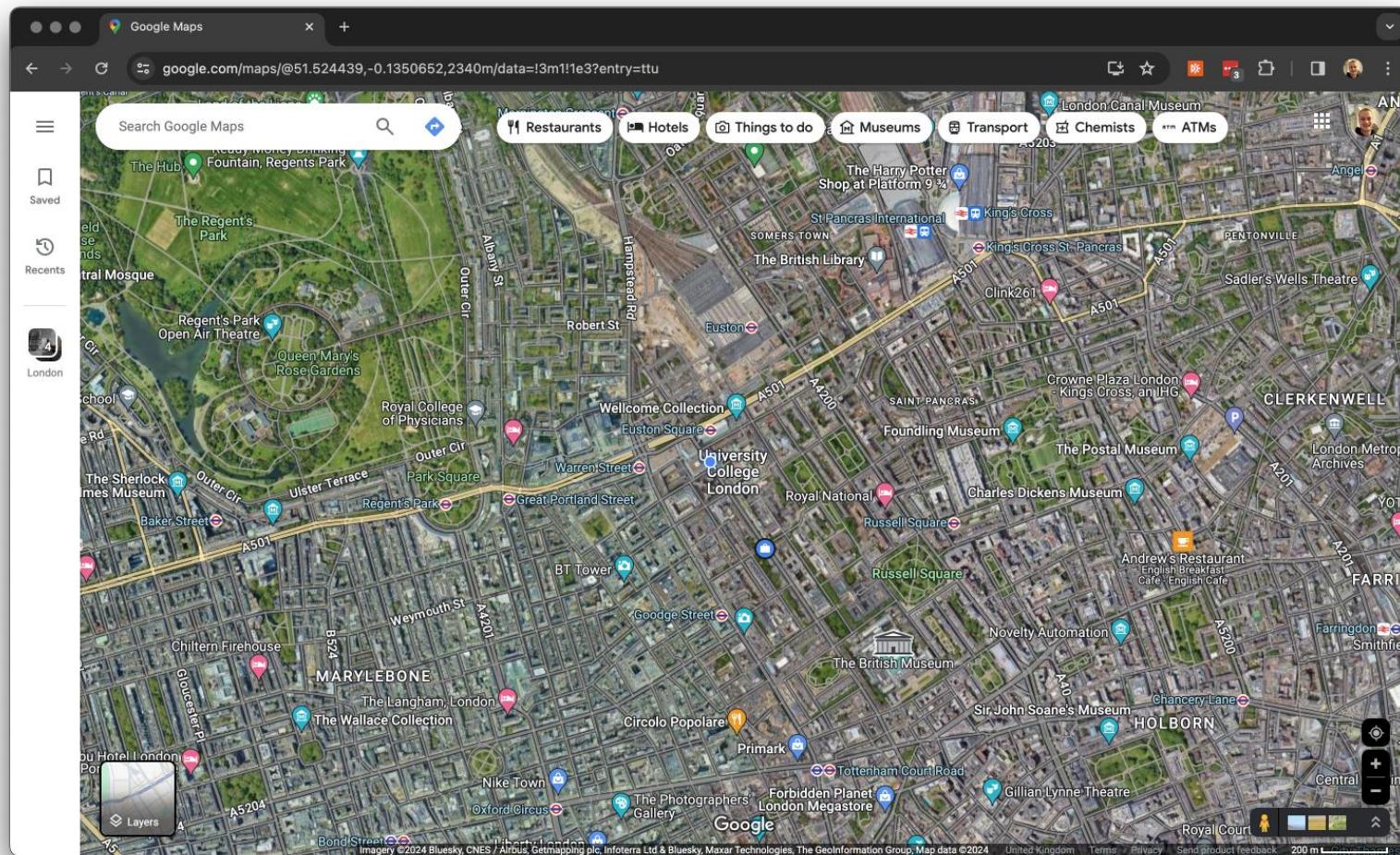
# Index of Multiple Deprivation in England



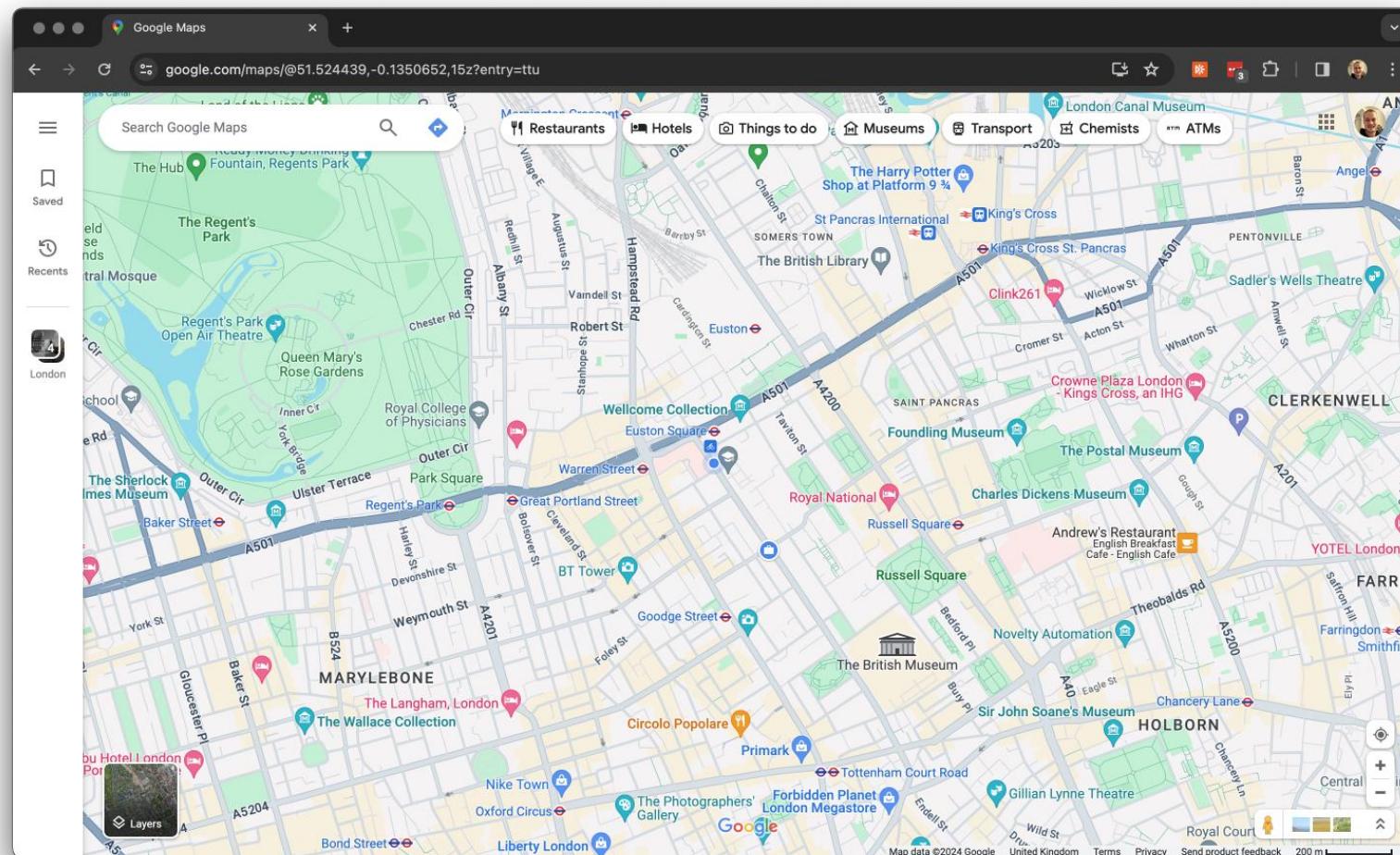
# Requirements

- GIScience: theory
- GISystems: methods and tools

# GIScience



# GIScience



# GIScience

- GIScience relies on representing spatial information in a digital format. Traditionally, geographic information is conveyed in two primary ways:

**Vector** This method uses a finite set of discrete geometric objects, such as points, lines, and polygons, to represent spatial features.

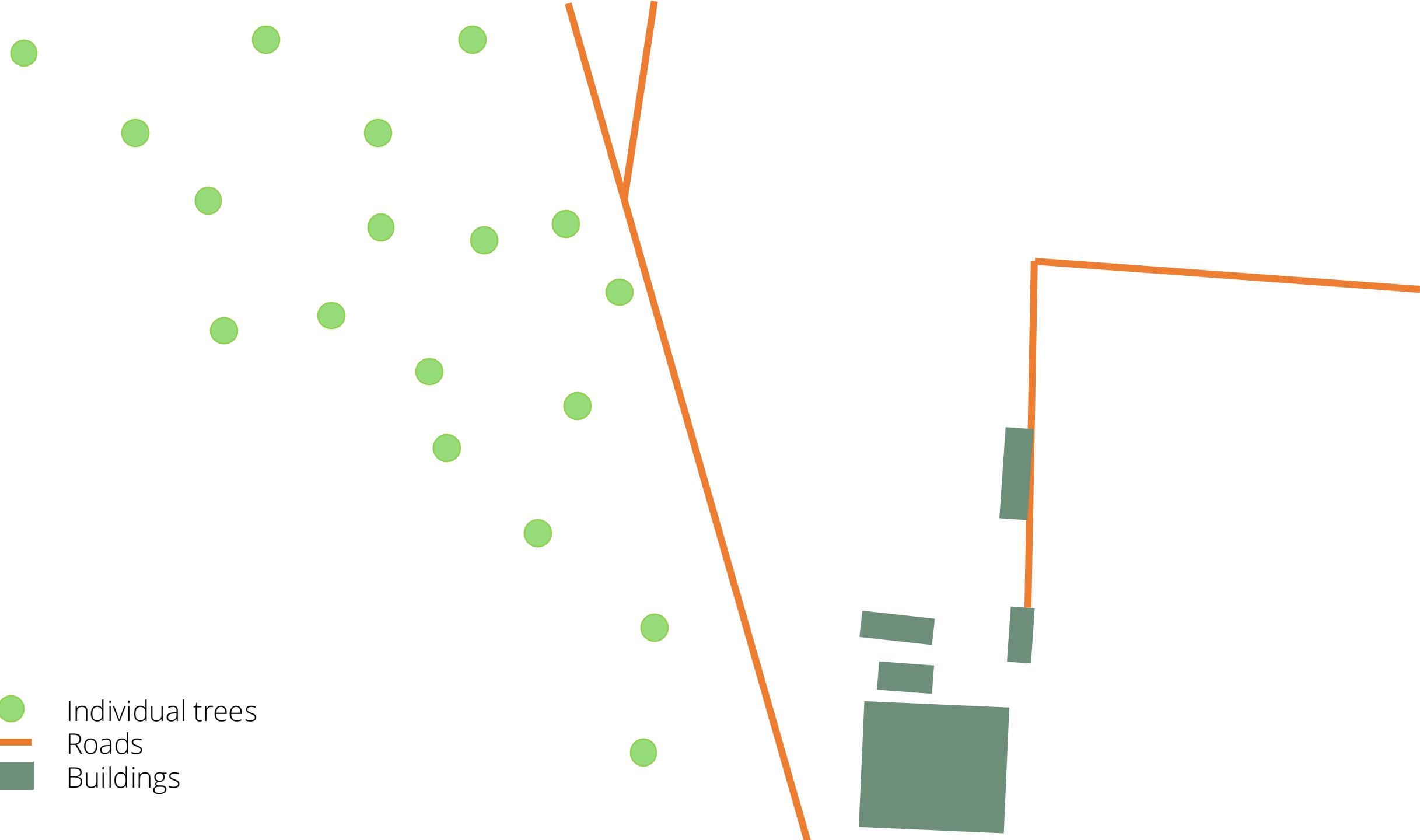
**Raster** This approach employs images or grids to represent surfaces, with each cell or pixel holding a value, often indicating attributes like colors or measurements.





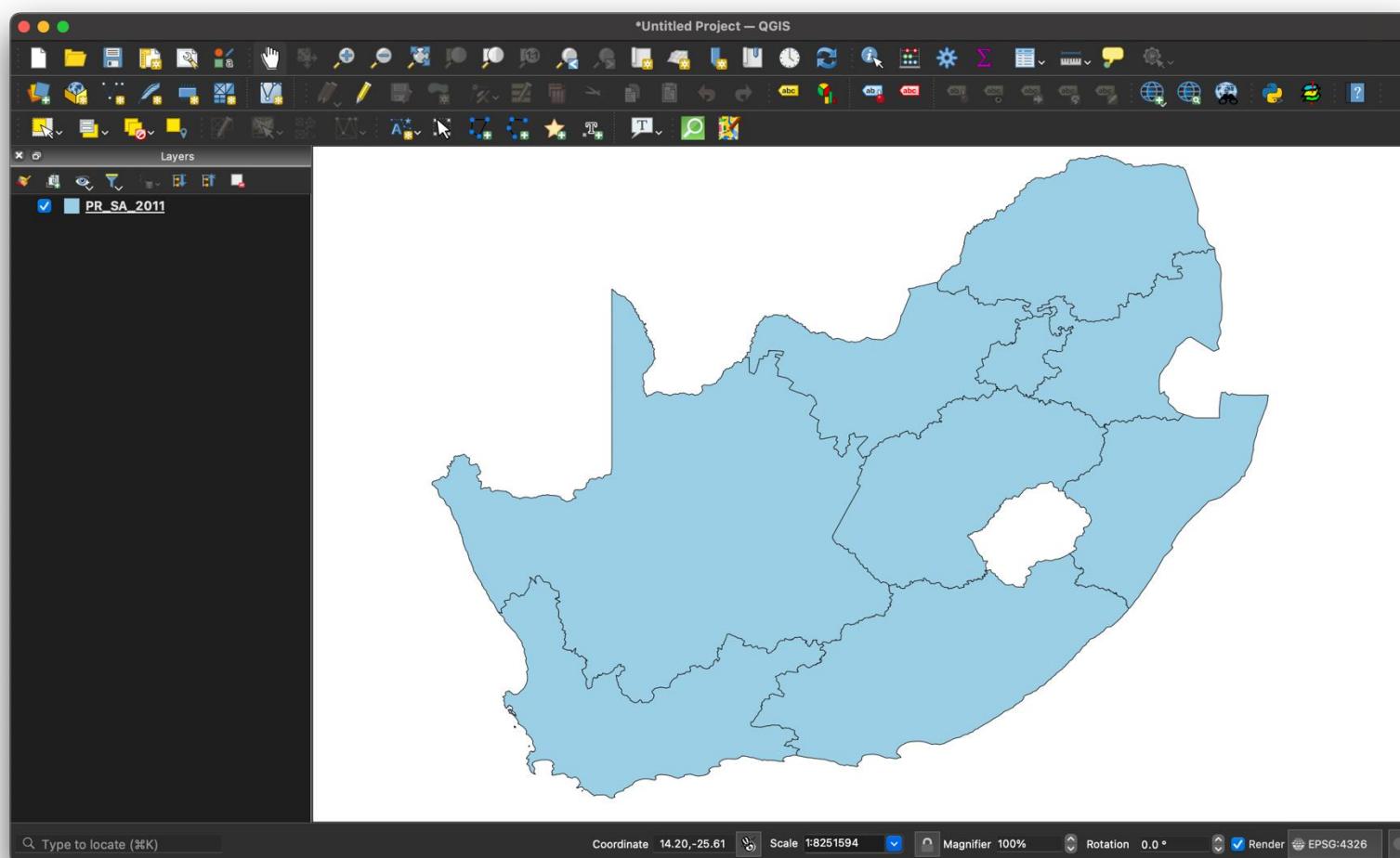




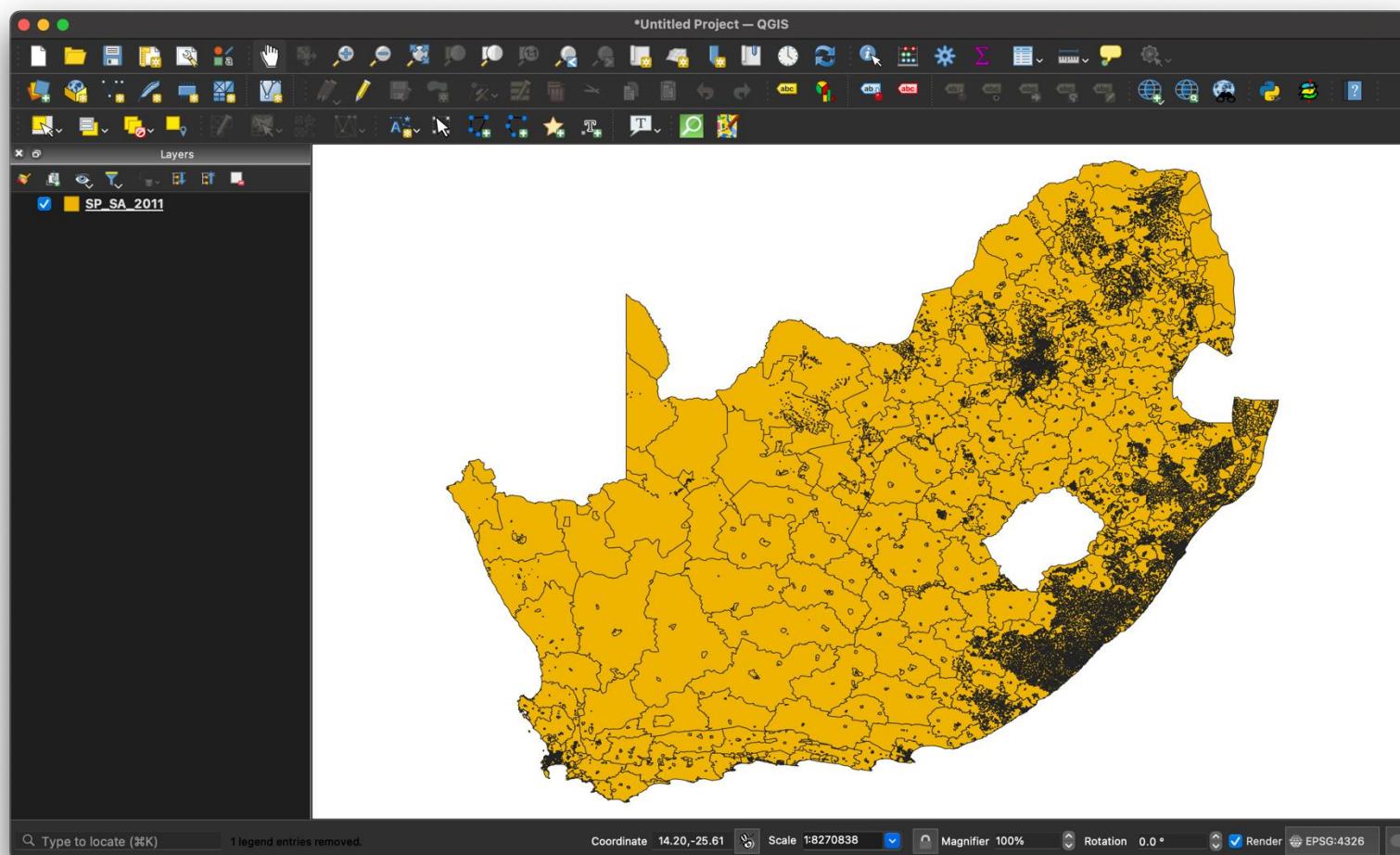


- Individual trees
- Roads
- Buildings

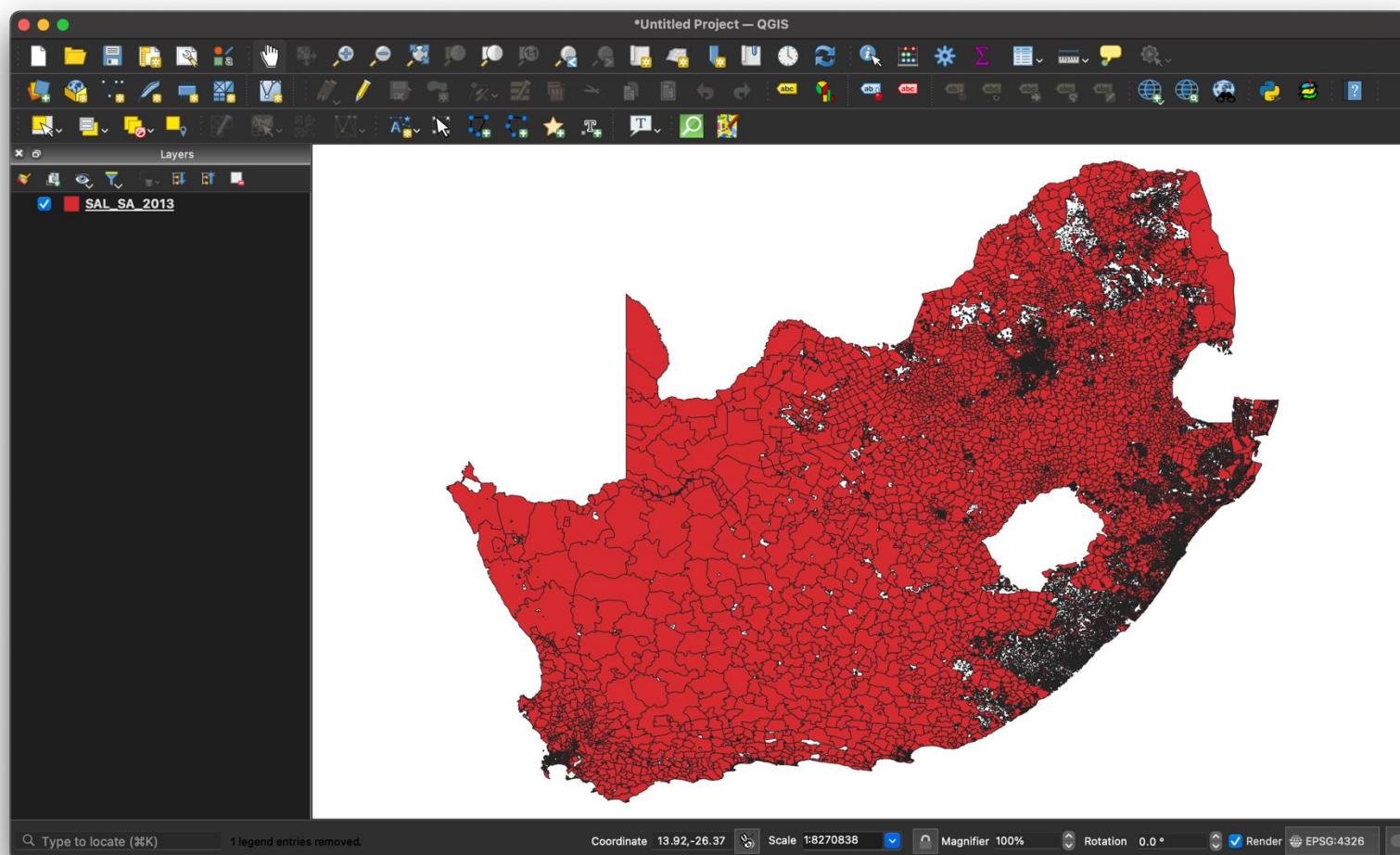
# GIScience



# GIScience



# GIScience



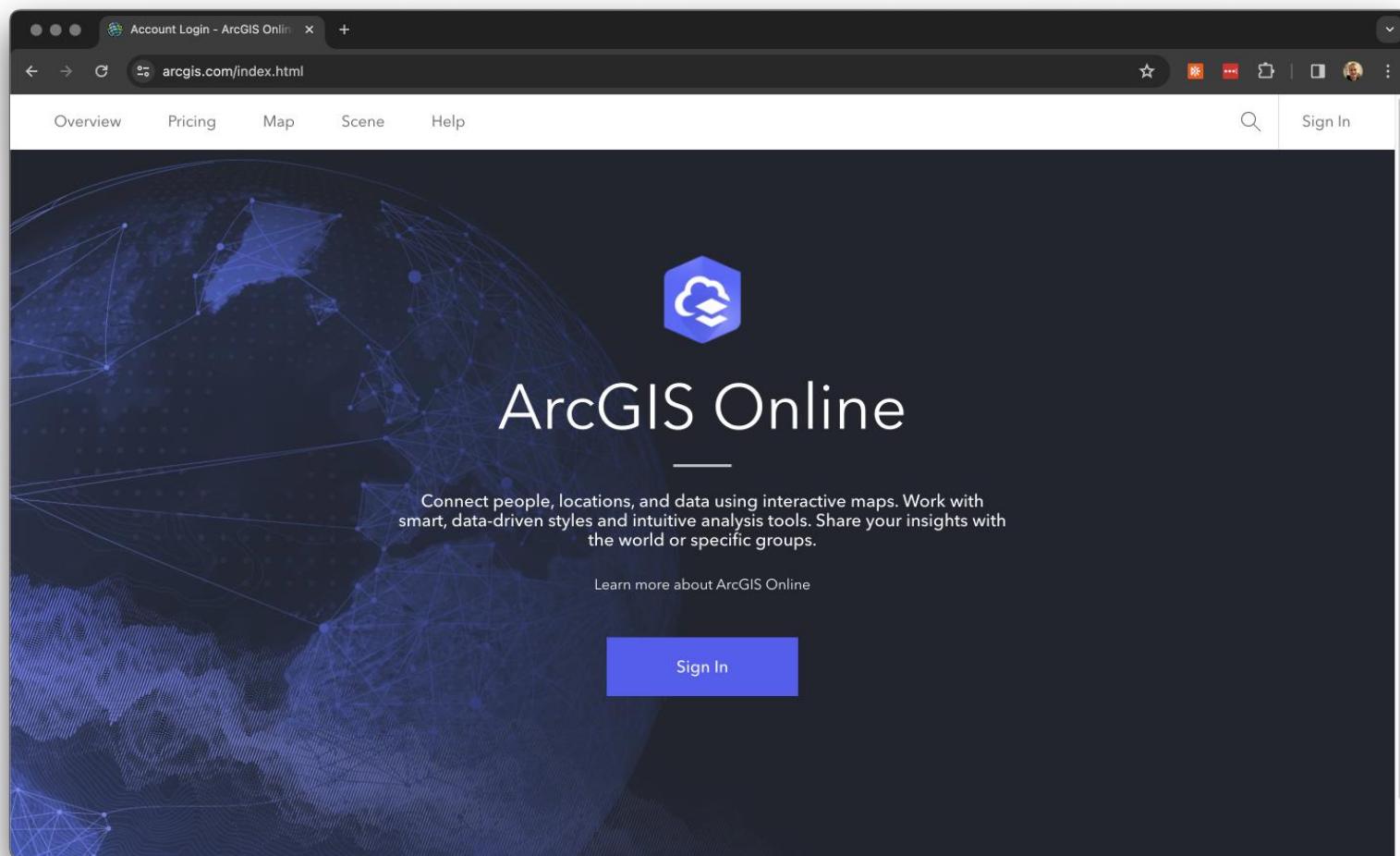
# Father of GIS

- 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.

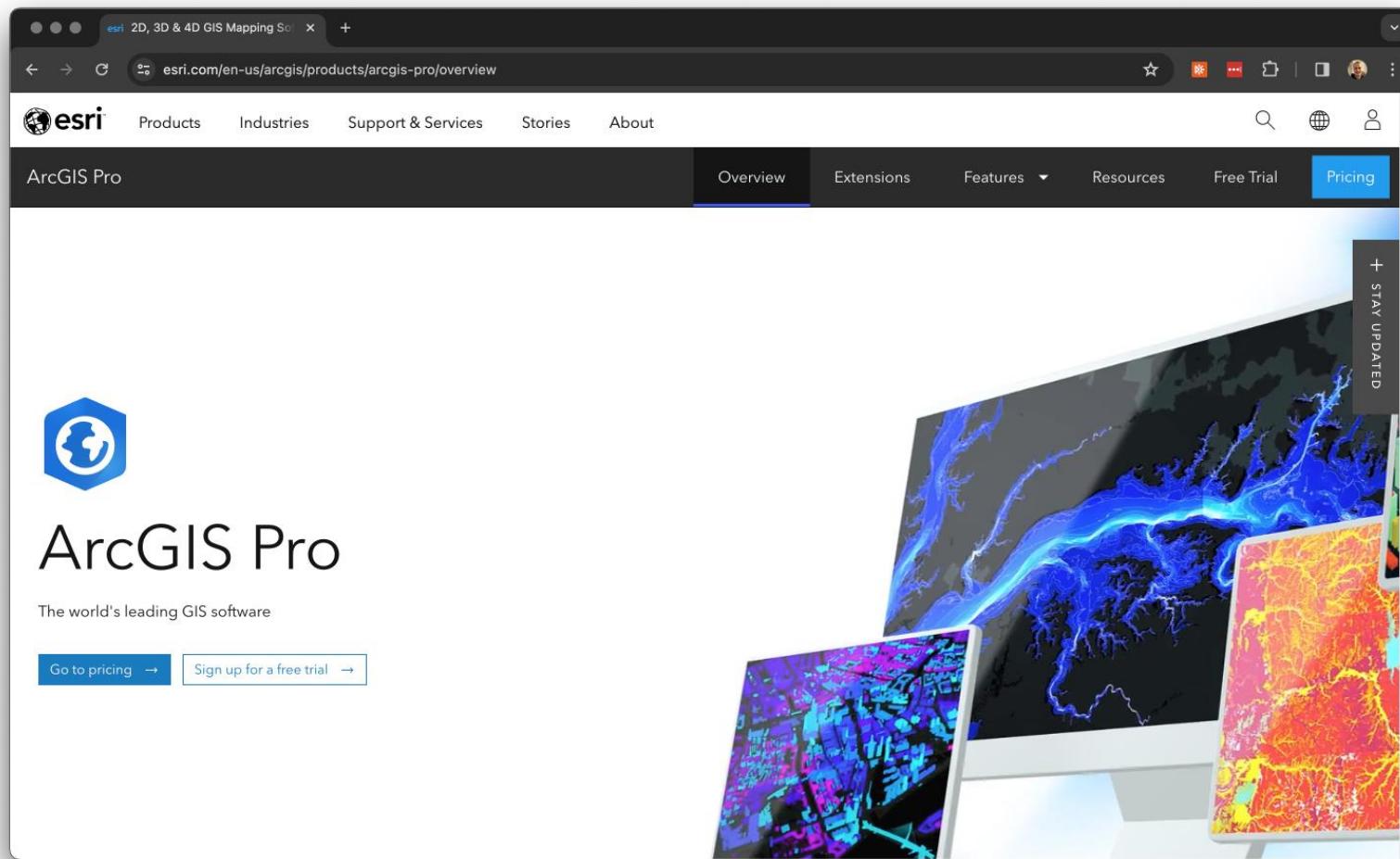
# Geographic Information Systems

- To effectively handle spatial information, we need tools that can:
  - Collect spatial data
  - Store spatial data
  - Analyse spatial data
  - Present spatial data
- Geographic Information Systems help us to manage spatial data: collection, storage, organisation, presentation

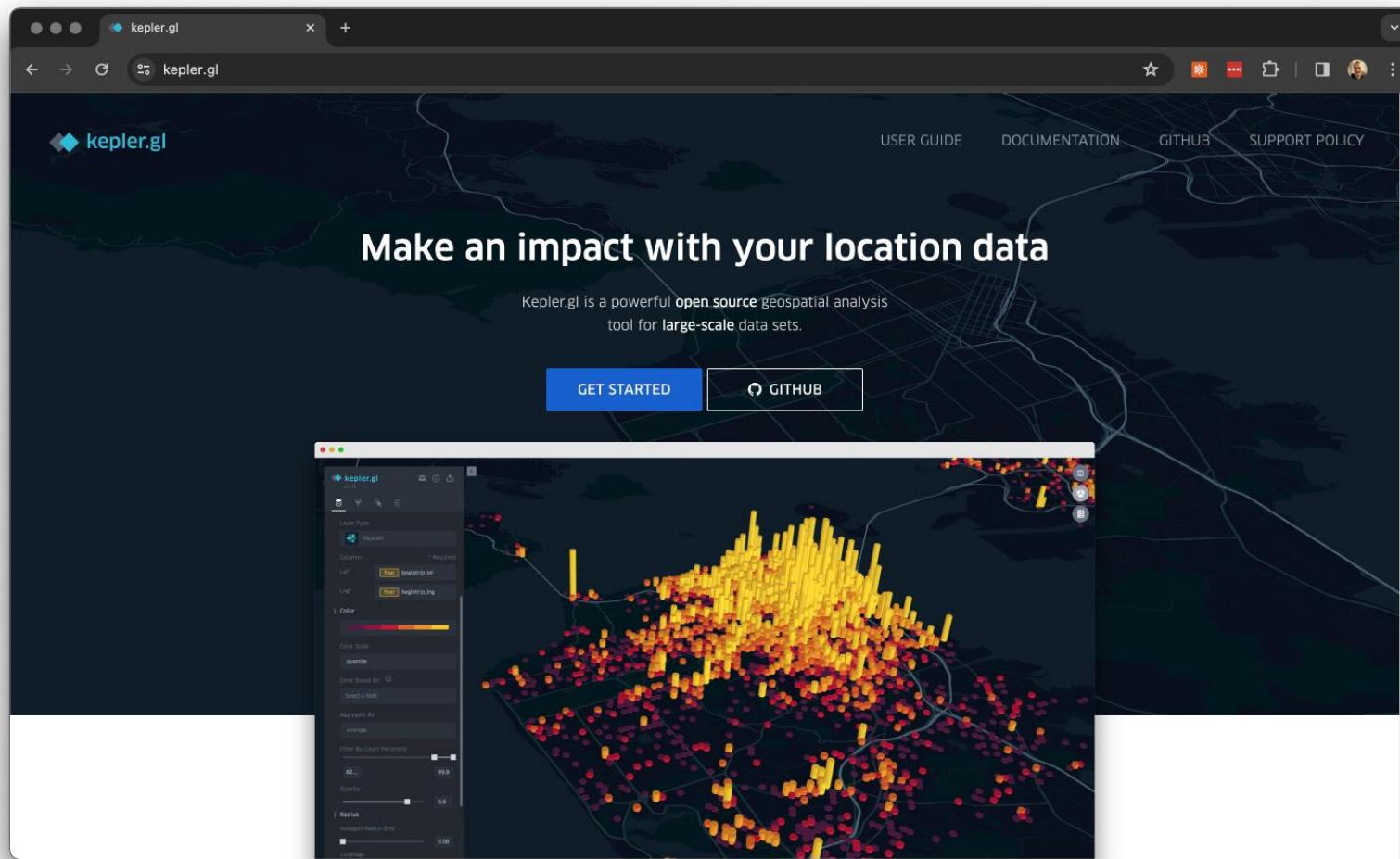
# Geographic Information Systems



# Geographic Information Systems



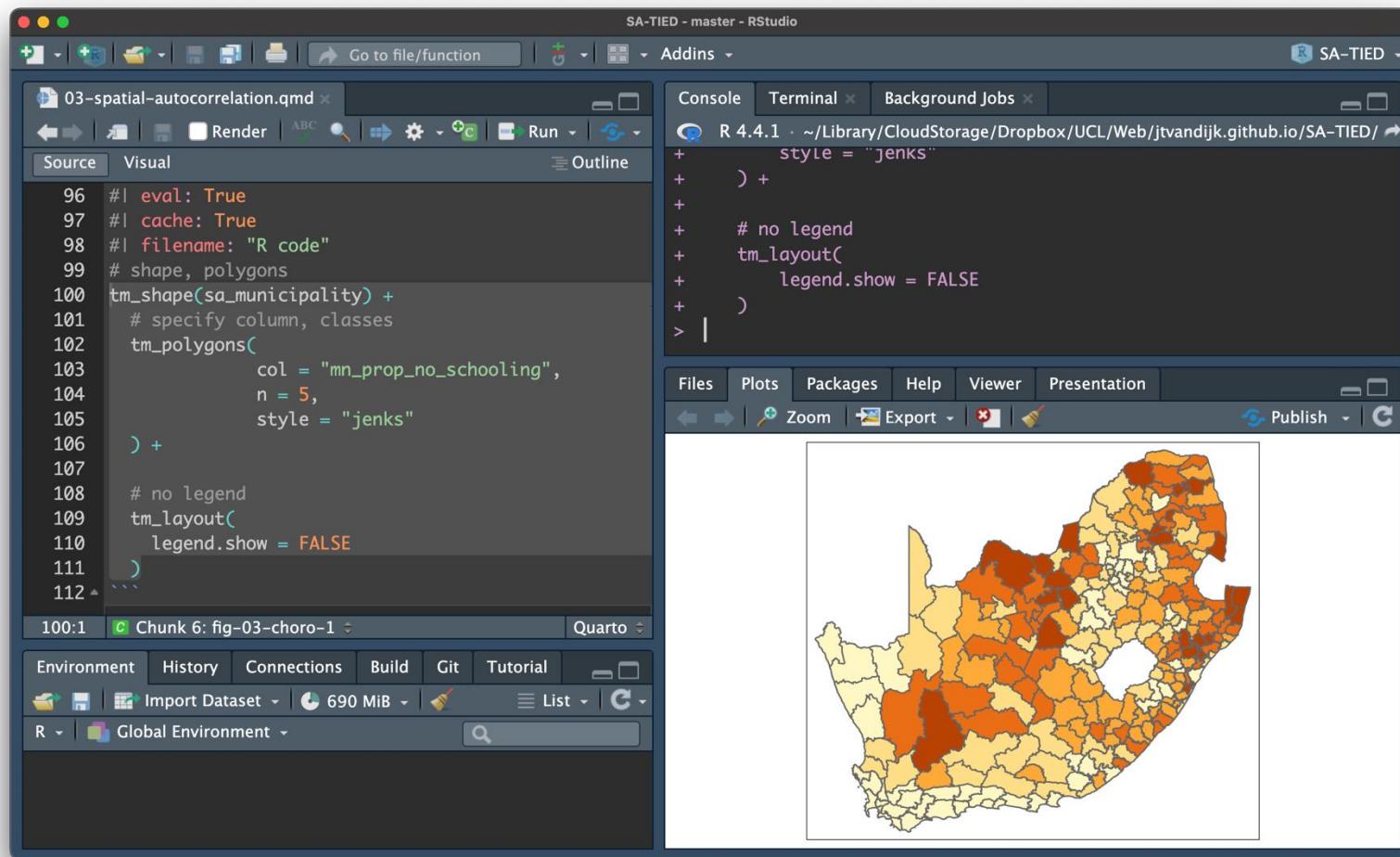
# Geographic Information Systems



# Geographic Information Systems



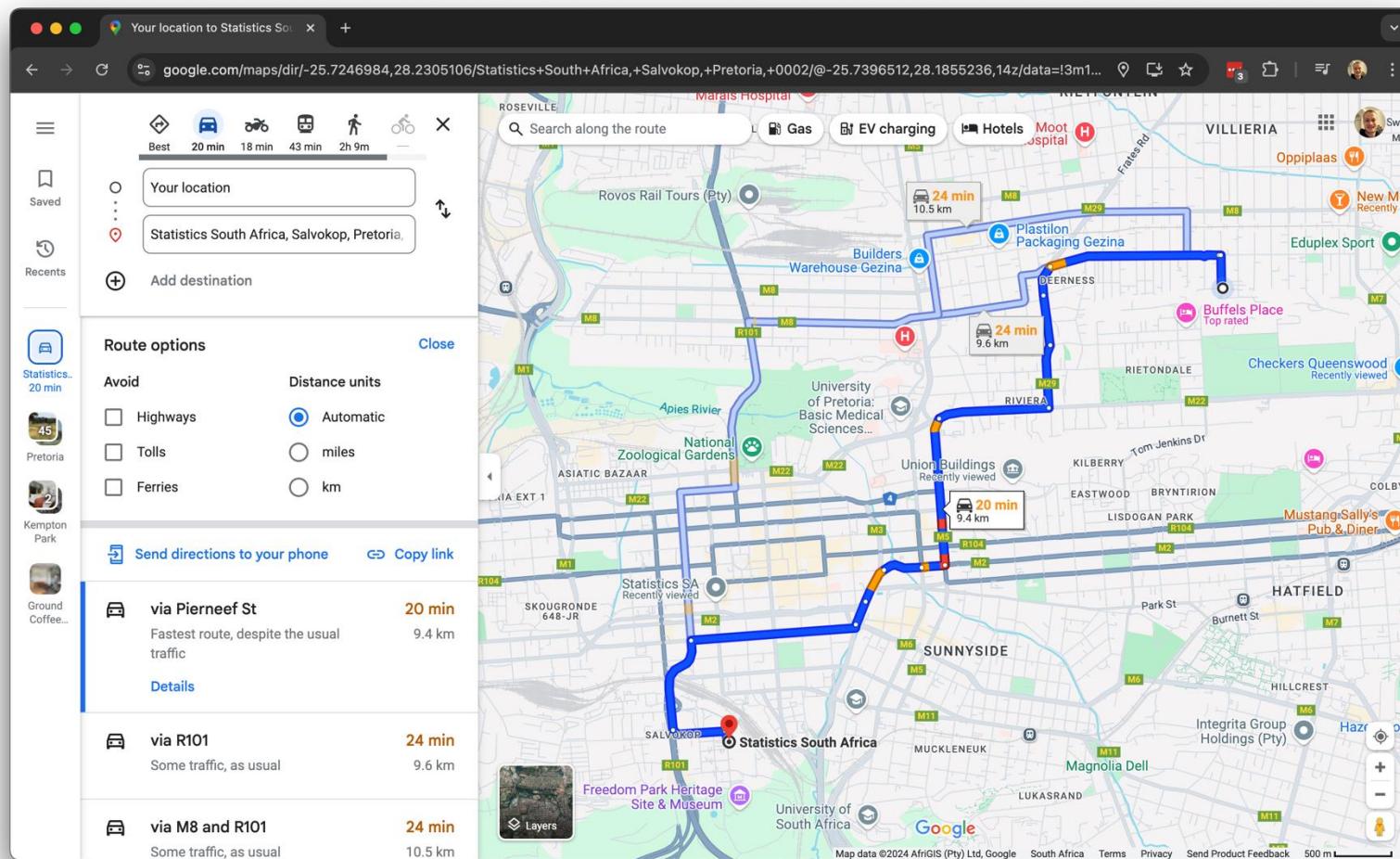
# Geographic Information Systems



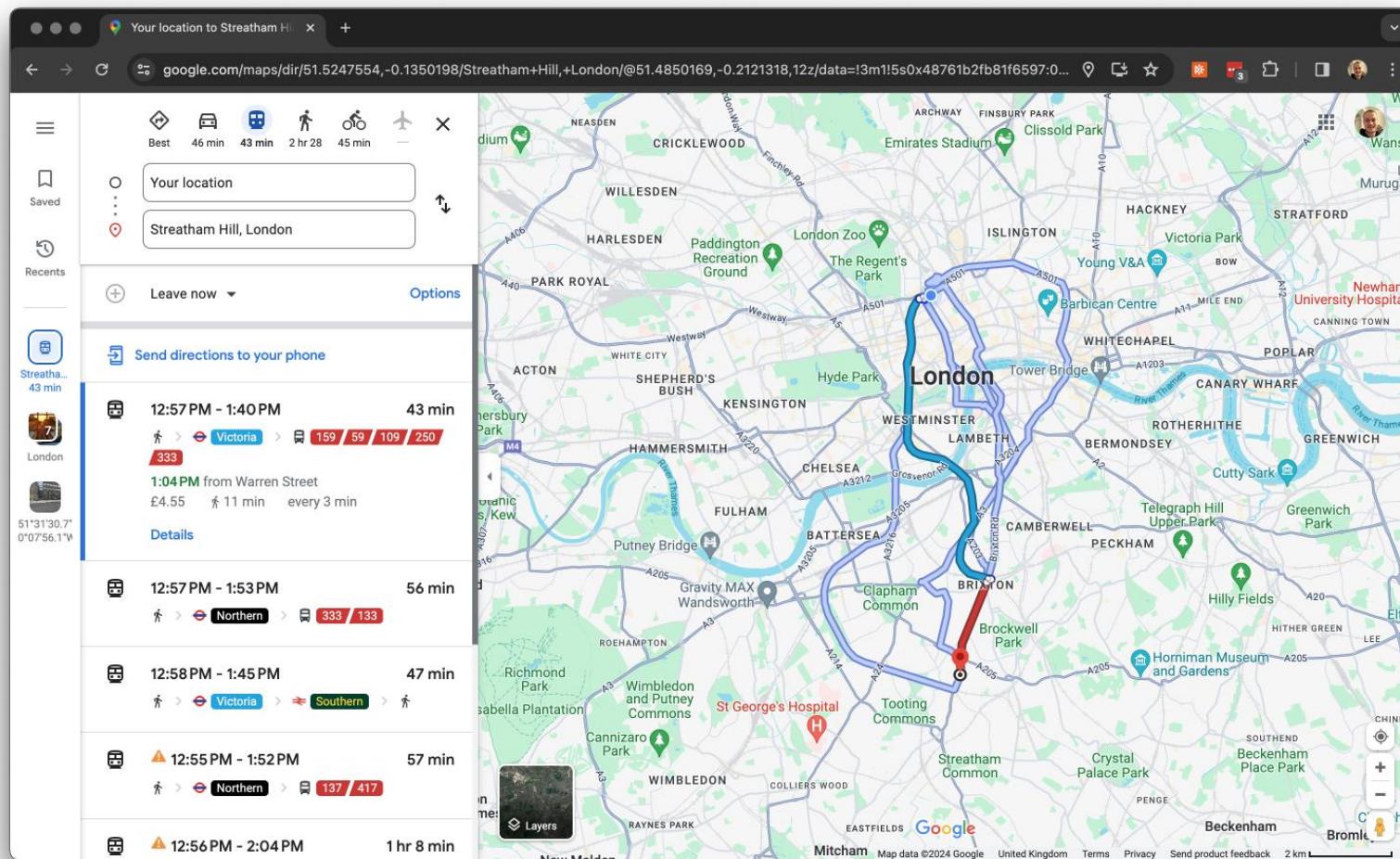
# Spatial analysis

- We can use GIScience and GISystems to convert spatial data into actionable information by quantifying elements such as distributions and spatial processes.

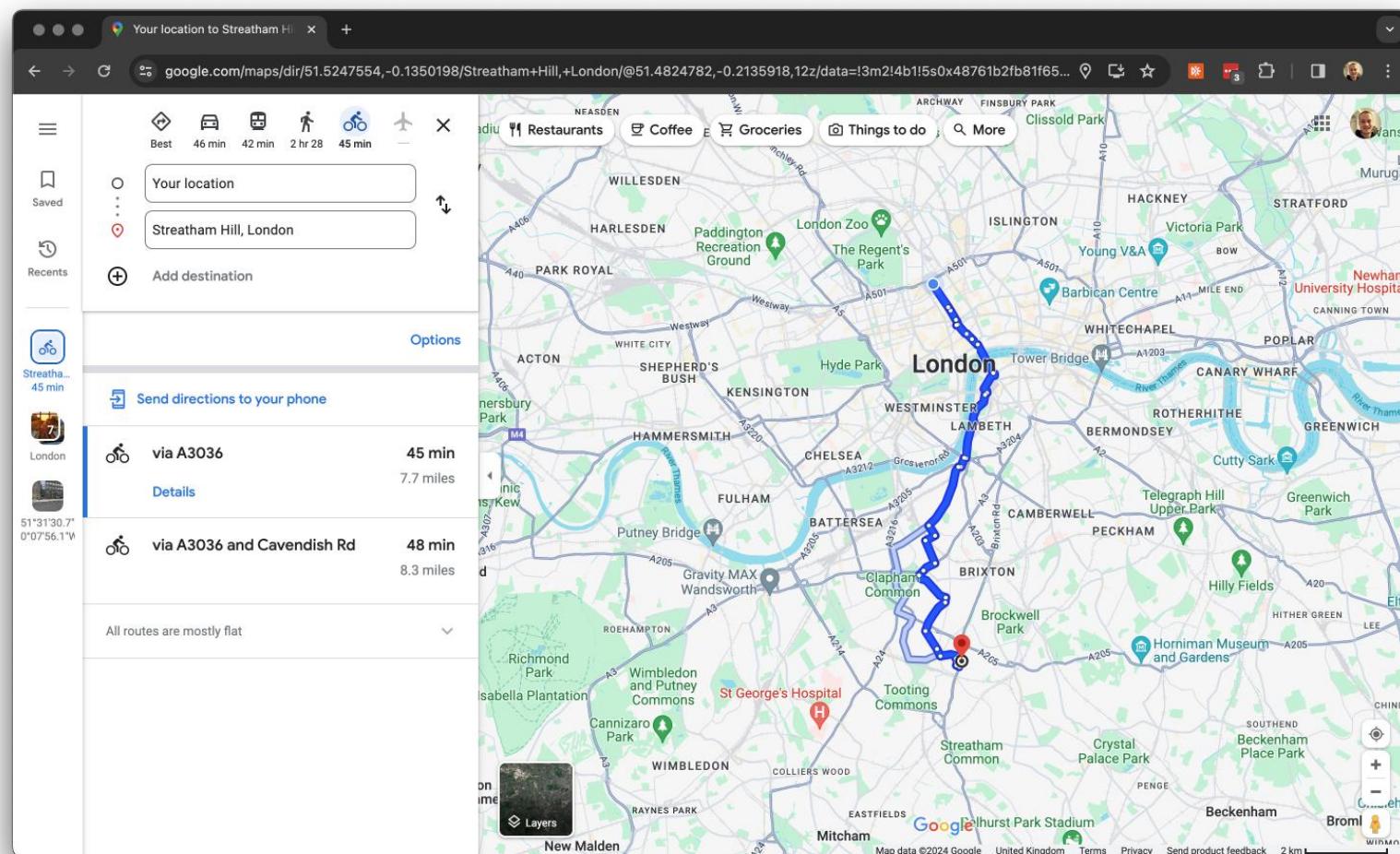
# Spatial analysis



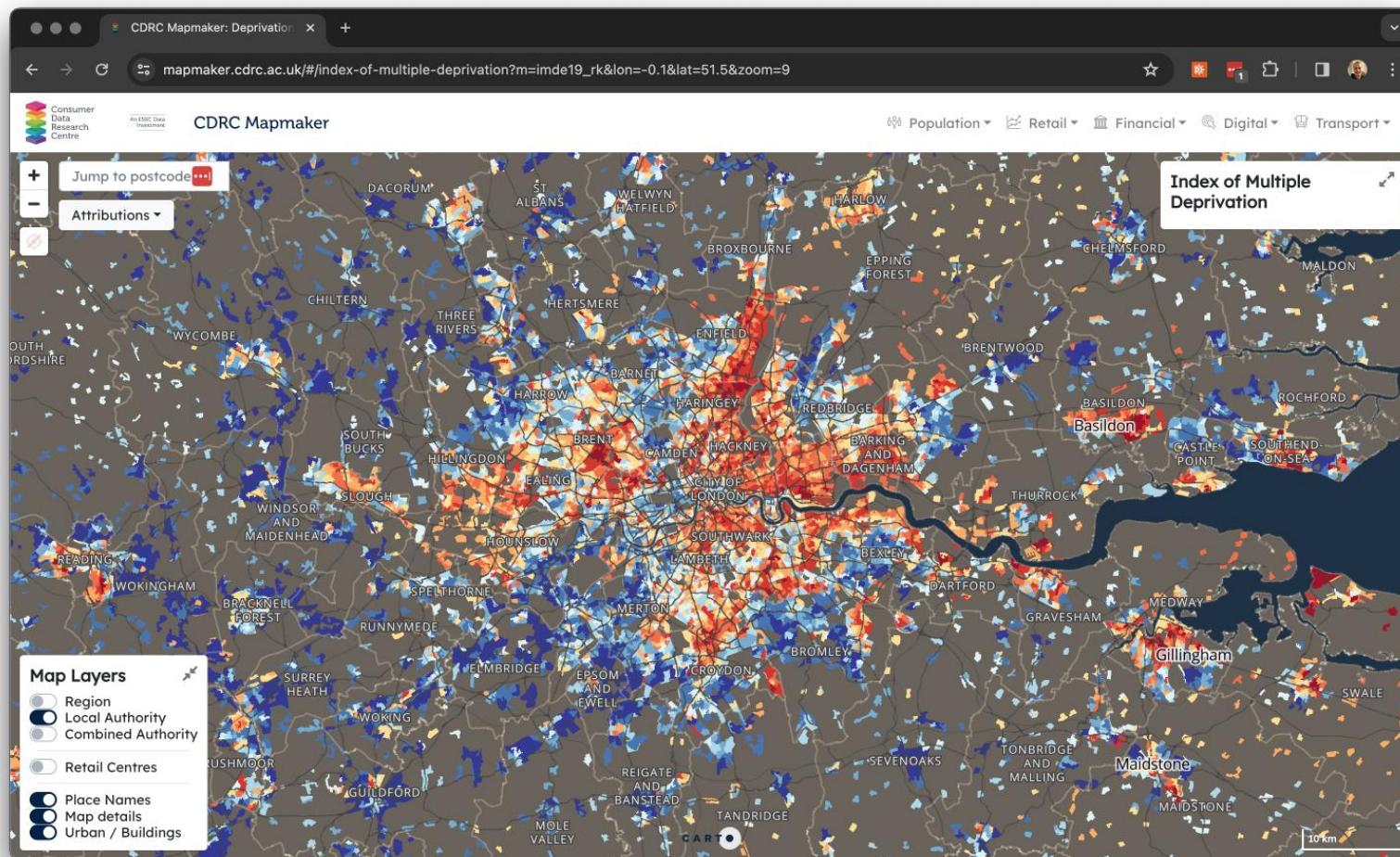
# Spatial analysis



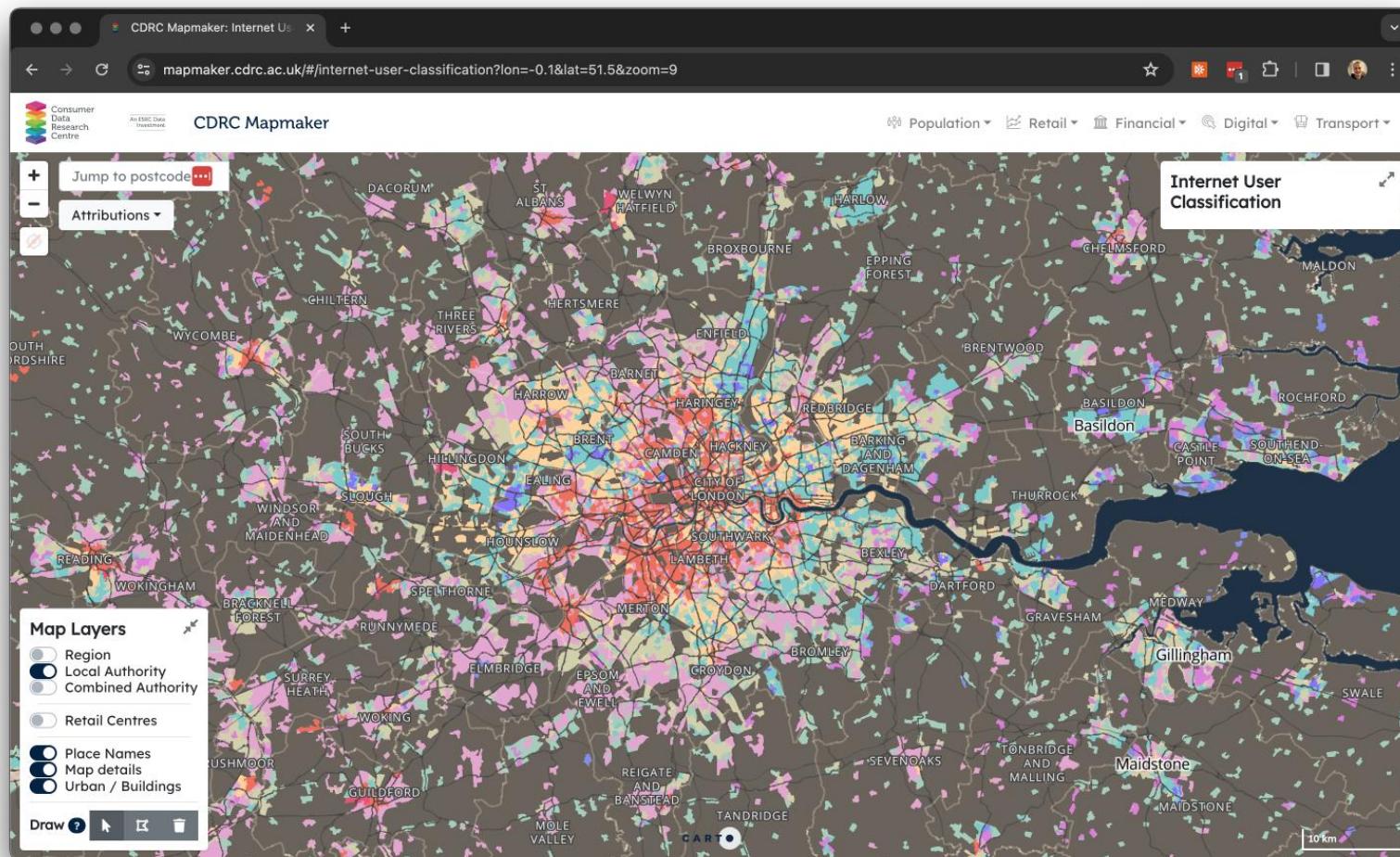
# Spatial analysis



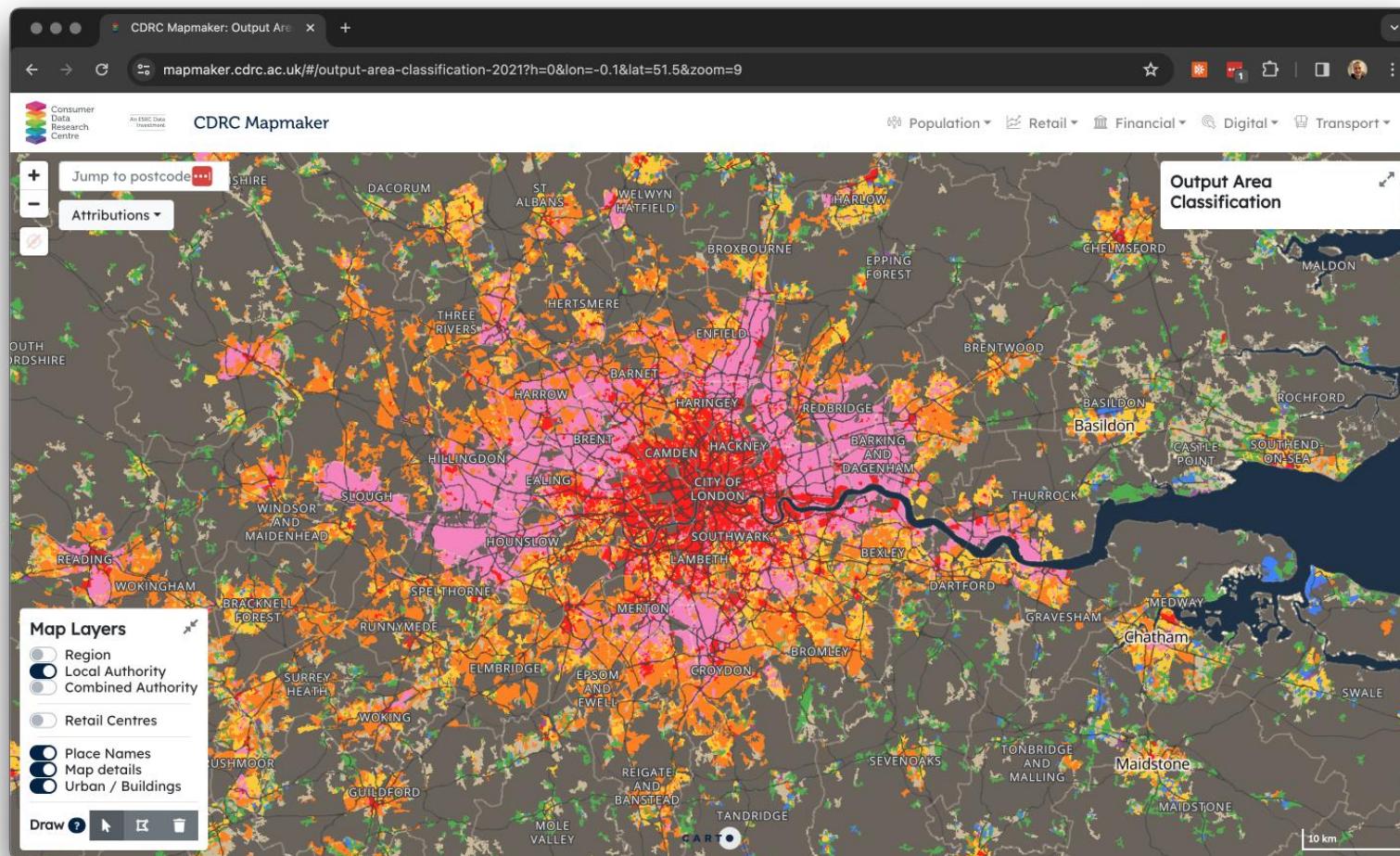
# Spatial analysis



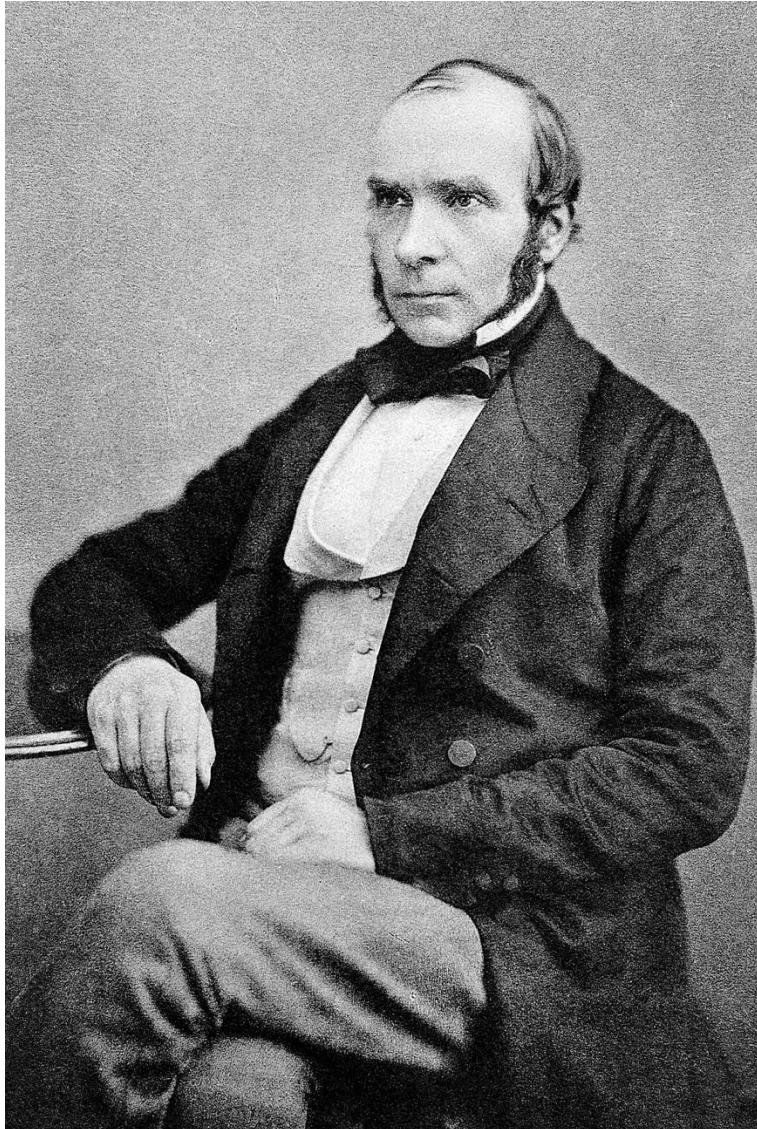
# Spatial analysis



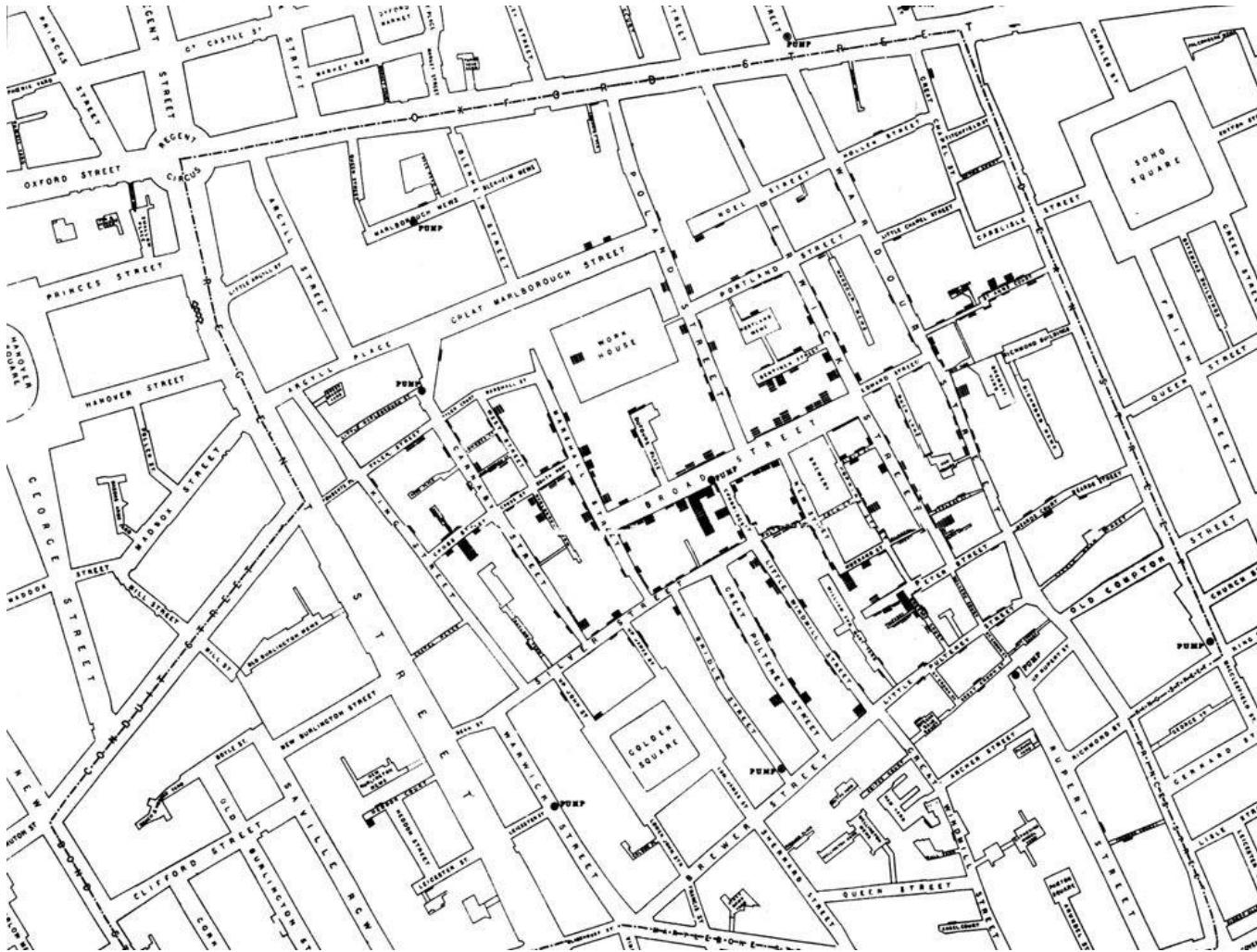
# Spatial analysis



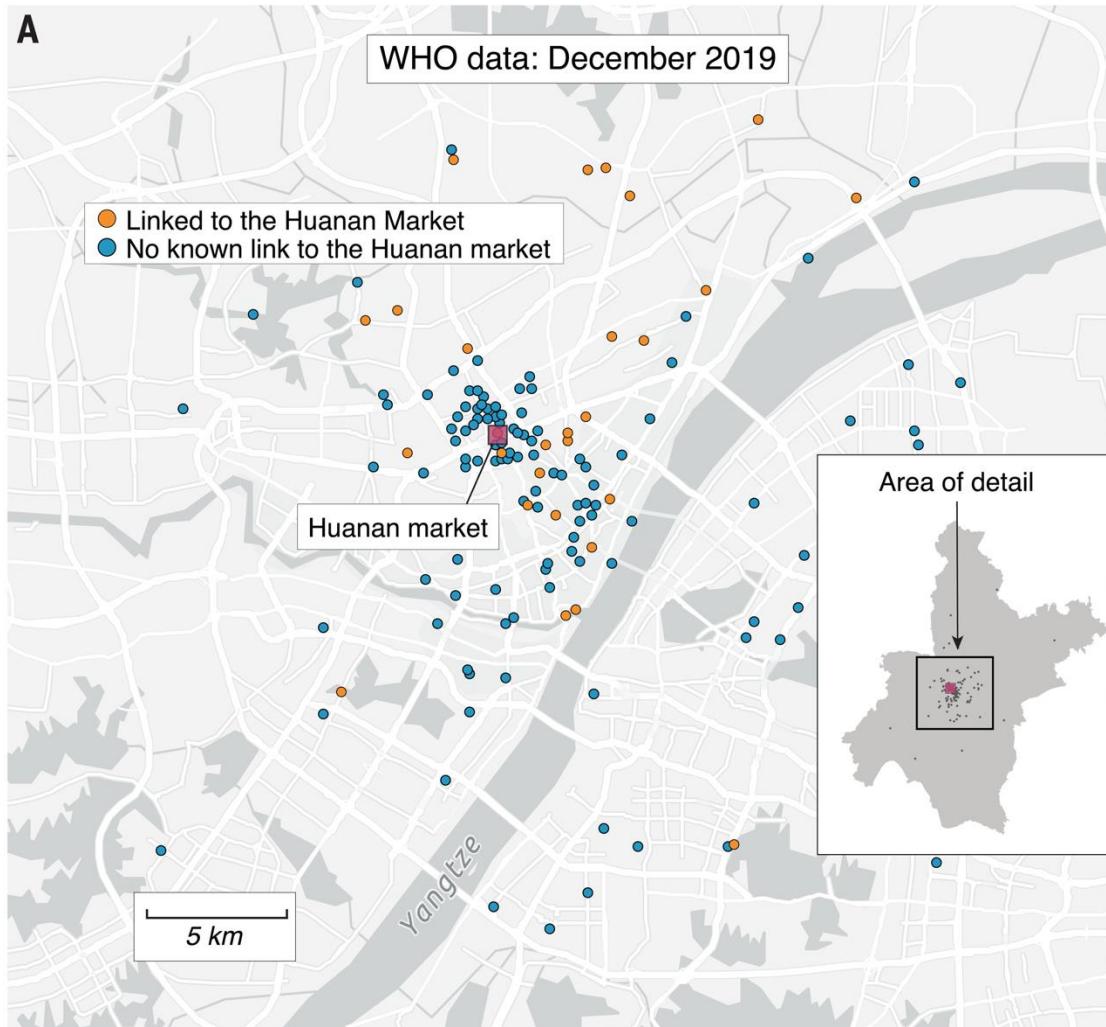
# Spatial analysis



# Spatial analysis



# Spatial analysis

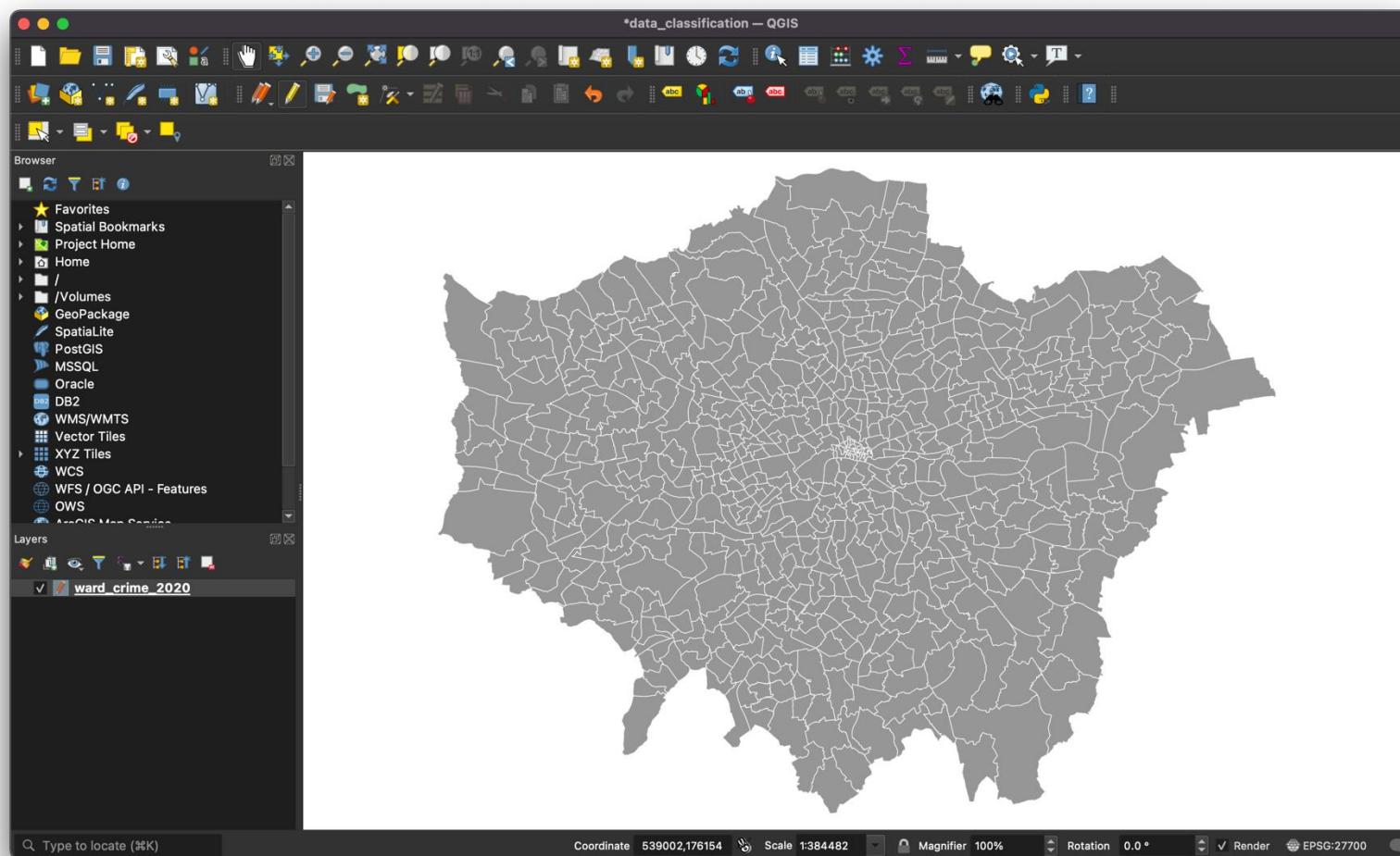


Worobey *et al.* 2022

# Mapping 101

- Continuous values can be divided into classes, serving as the spatial equivalent of a histogram.
- To determine the optimal classification method and where to set the breaks, it is essential to analyse the data distribution using a histogram.
- The choice of classification approach is guided by this distribution, balancing the trade-off between **minimising information loss** and maintaining readability and simplicity.

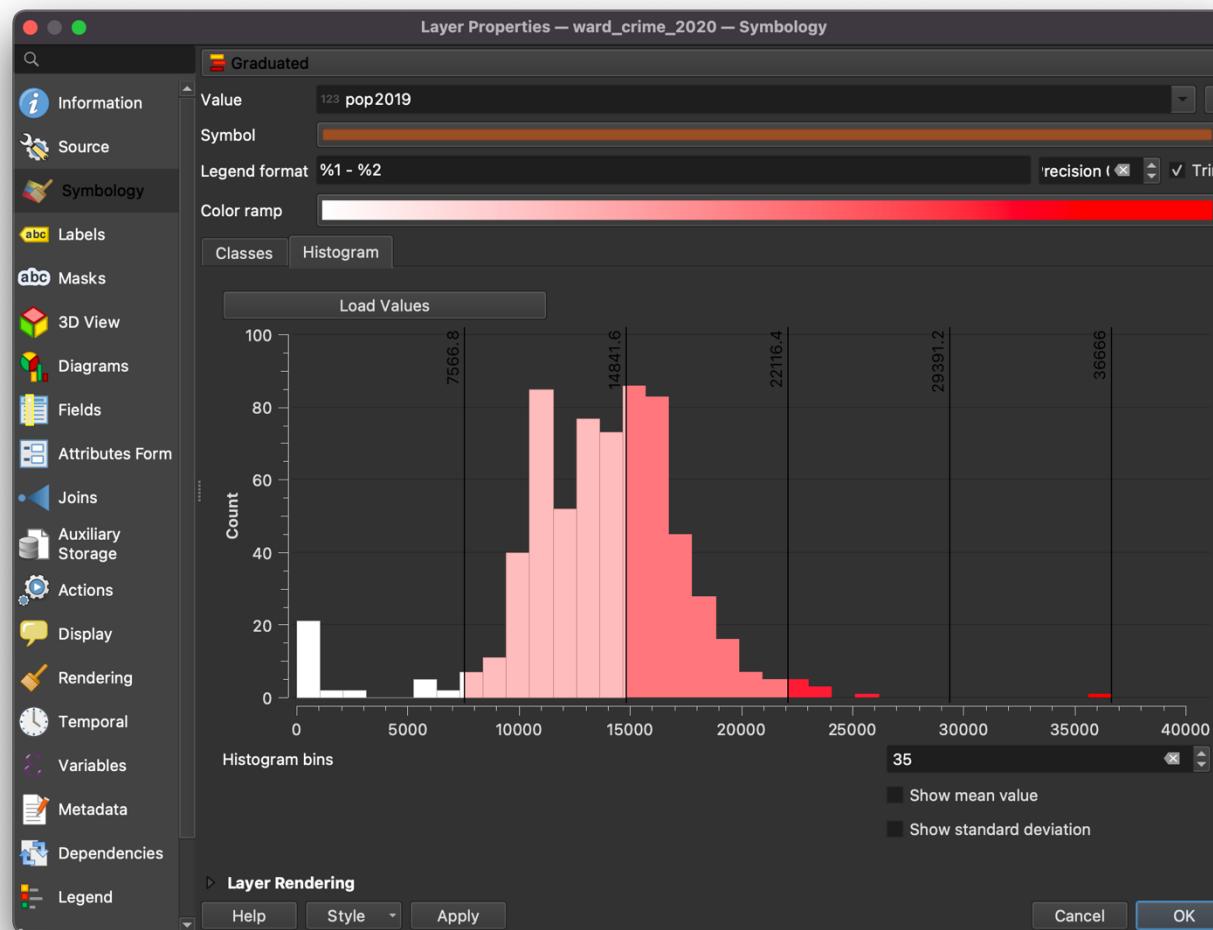
# Mapping 101



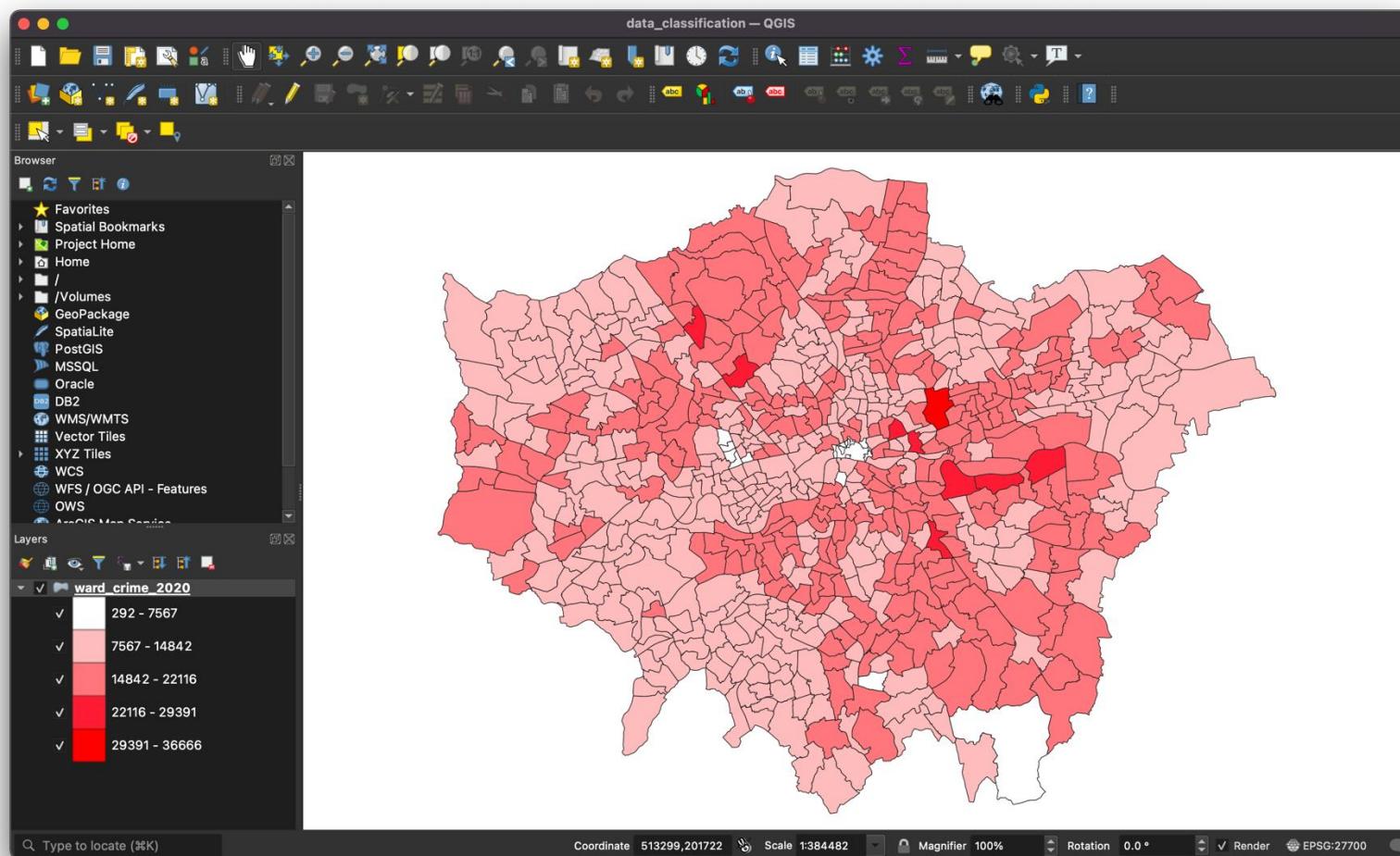
# Mapping 101

**Equal Interval:** This method divides the range of data values into equal intervals based on numerical value. If the data distribution is skewed, this approach can give more weight to outliers, potentially leading to less balanced representation across the classes.

# Mapping 101



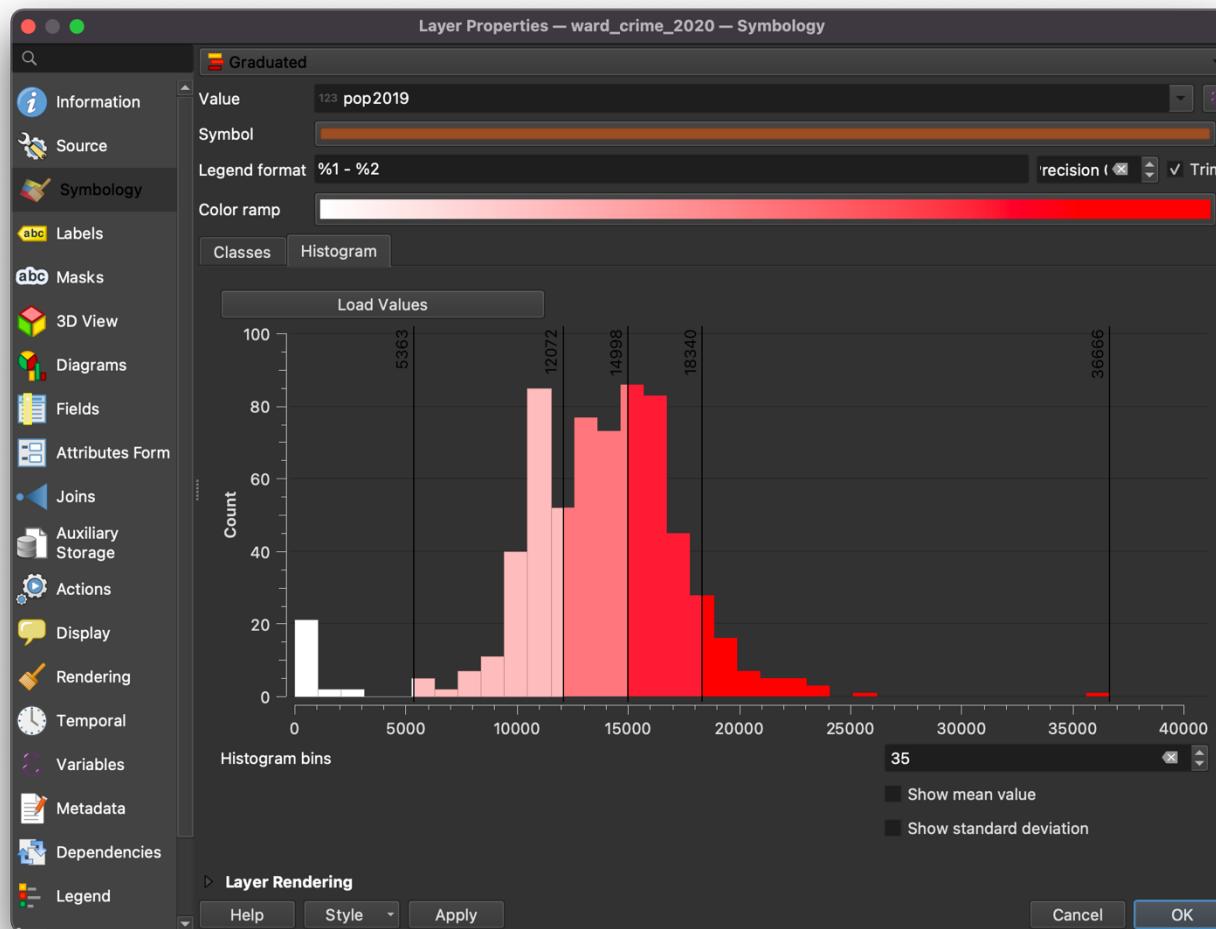
# Mapping 101



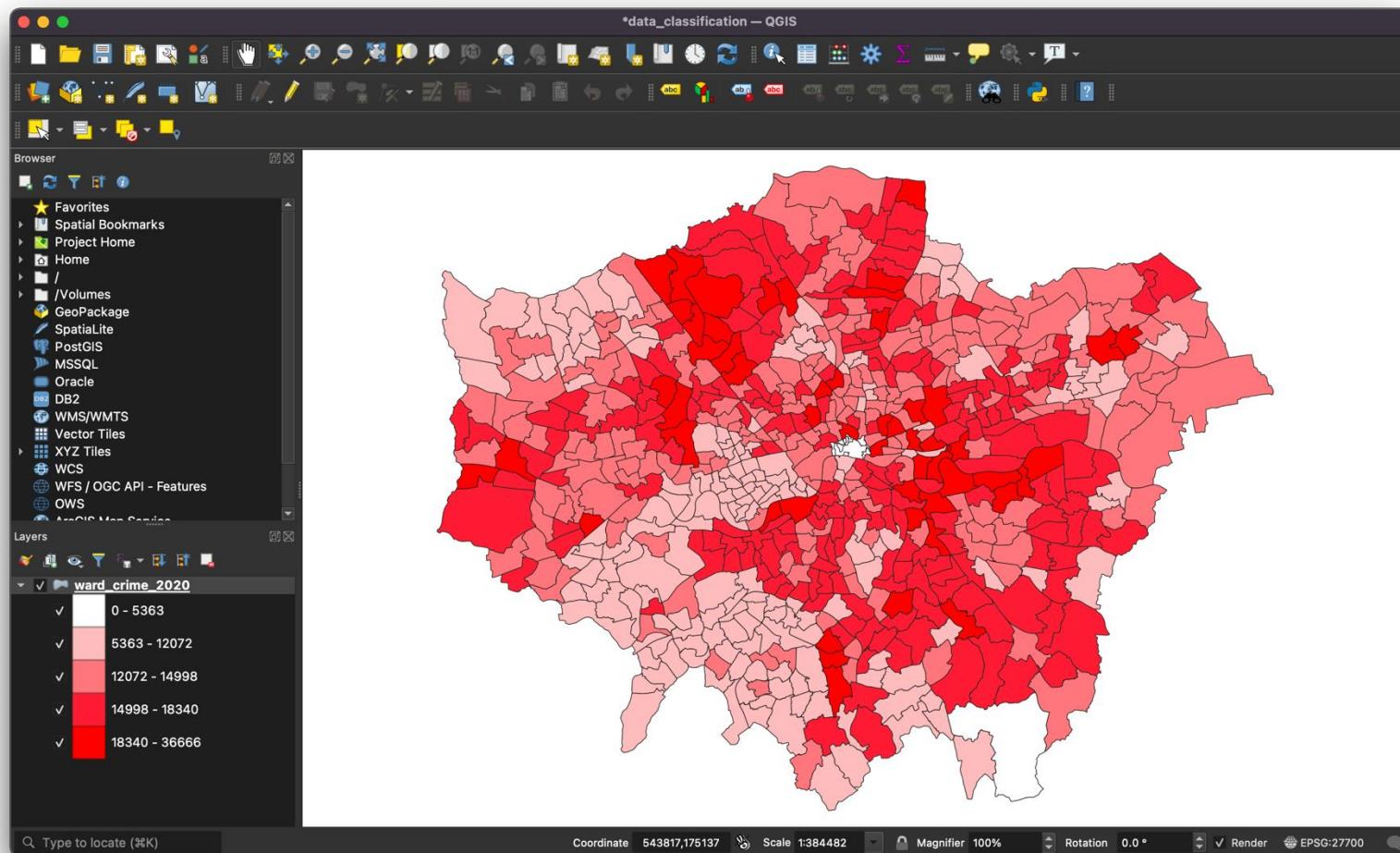
# Mapping 101

**Natural Breaks/Jenks:** This method aims to minimise the average deviation within each class from the class mean while maximising the deviation between the means of different classes. In the case of a skewed distribution, outliers often end up in their own class, although this can vary depending on the number of classes chosen.

# Mapping 101



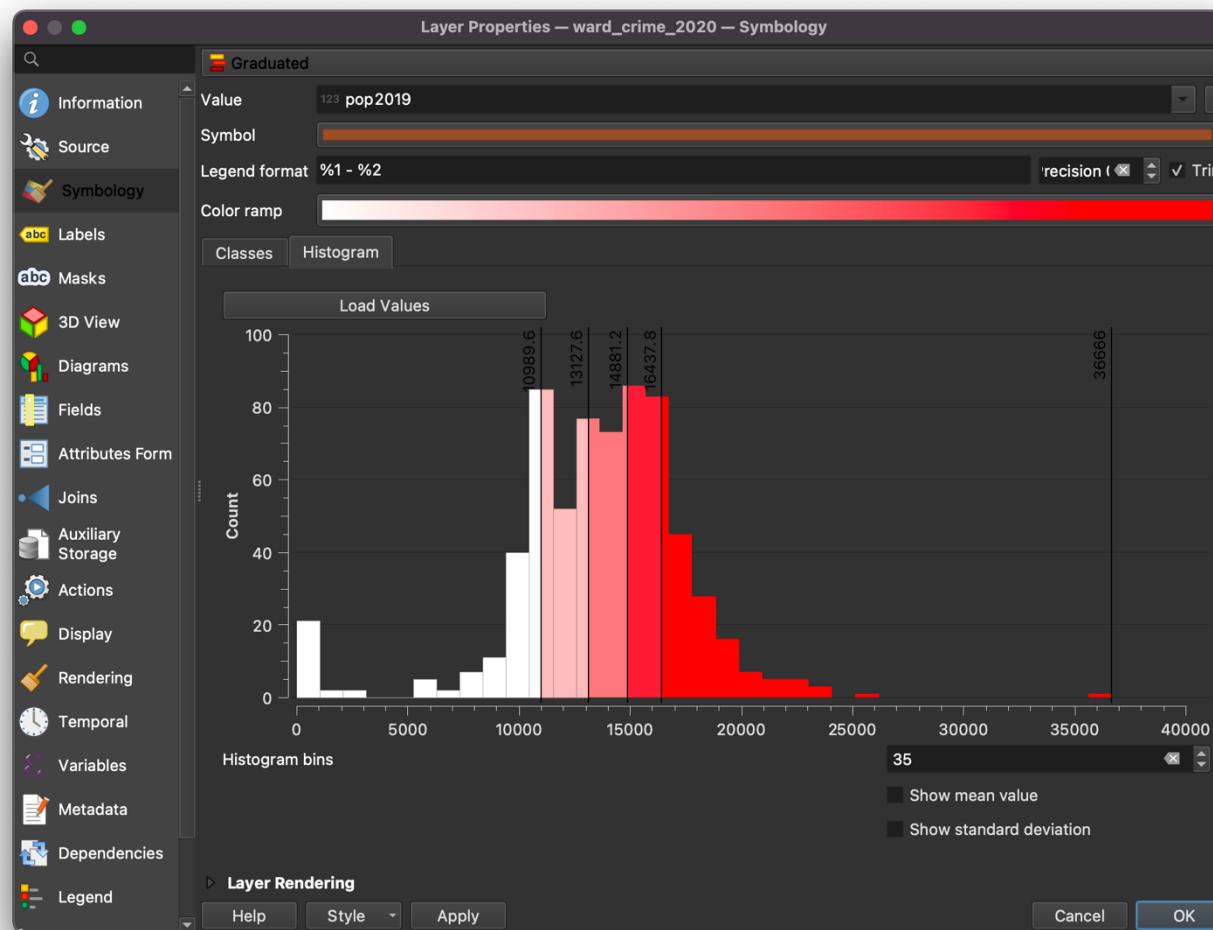
# Mapping 101



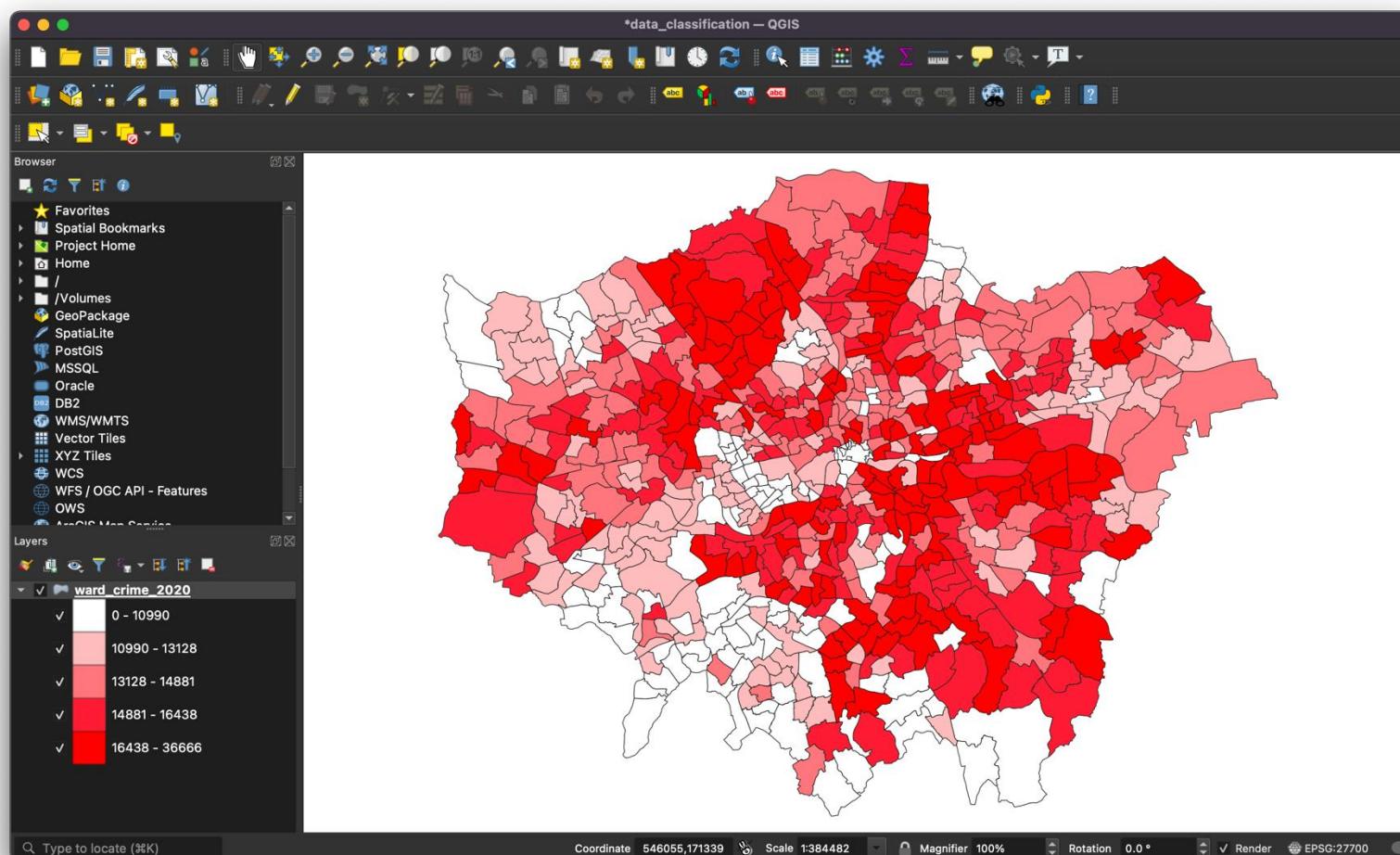
# Mapping 101

**Quantiles:** This method divides the distribution into classes so that each bin contains the same number of values. The splits are based on the rank order of the data rather than the actual values. When the distribution is skewed, this approach can result in very different values being grouped together in the same bin.

# Mapping 101



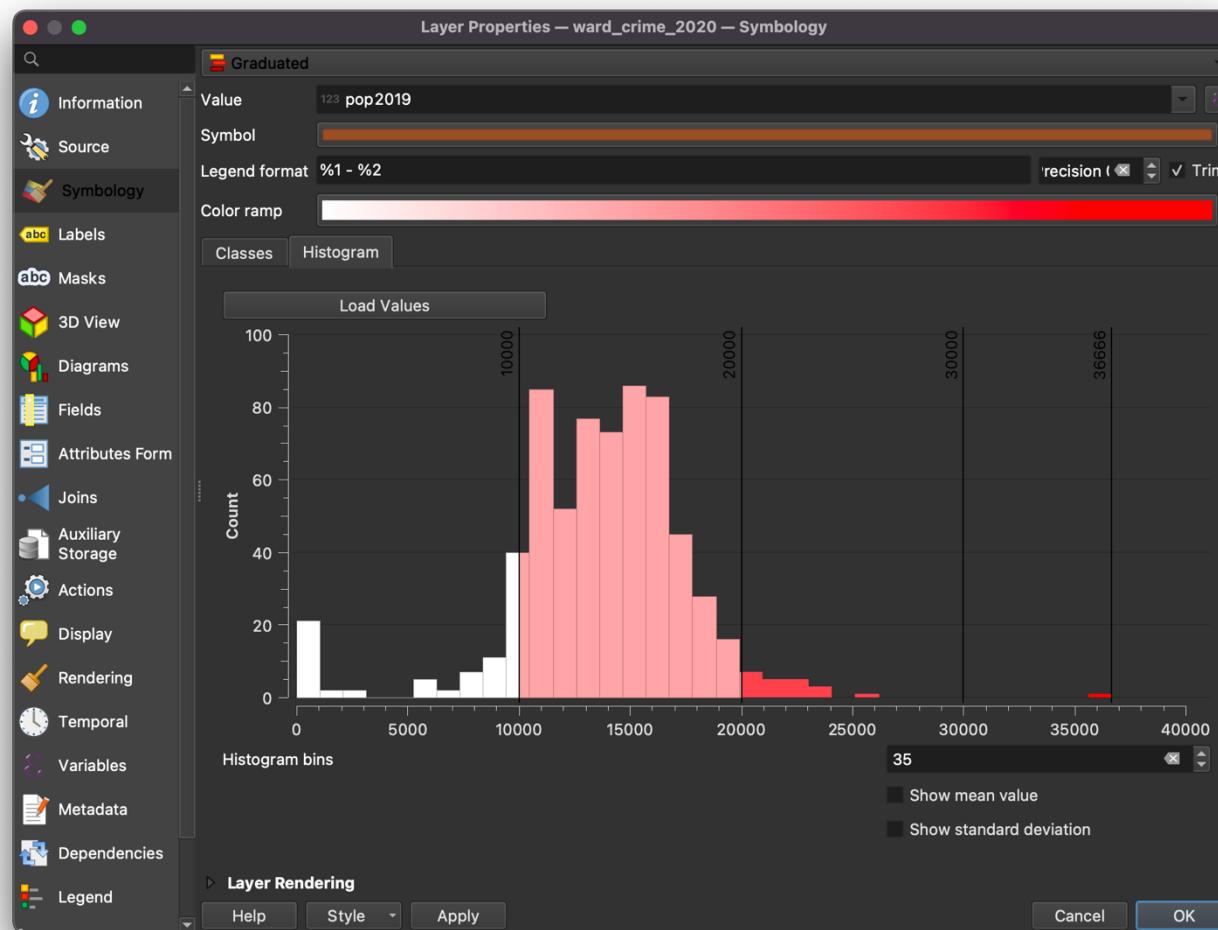
# Mapping 101



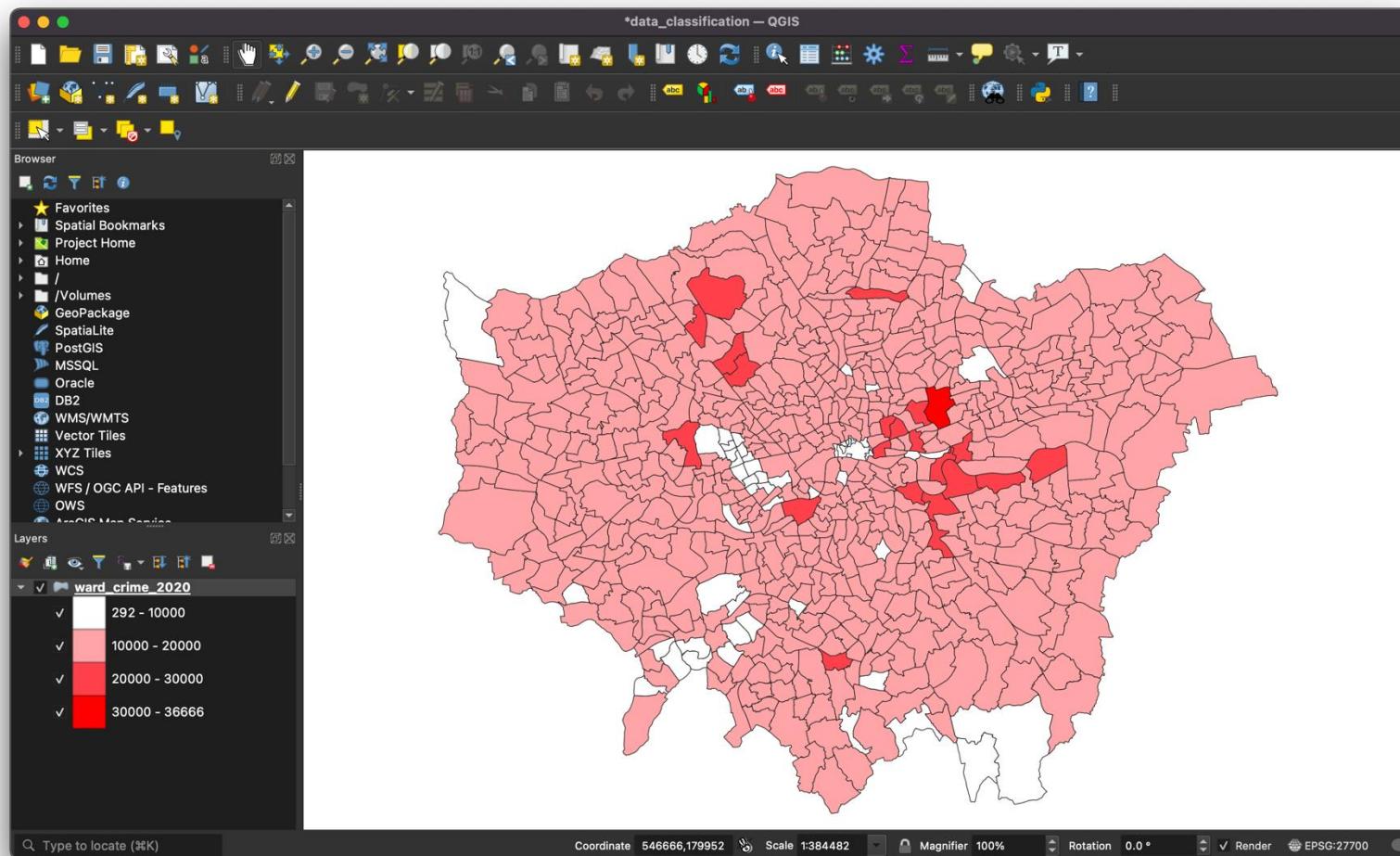
# Mapping 101

Pretty: This method divides the distribution using a series of aesthetically pleasing, rounded values. The splits are not influenced by the data or its distribution. In cases of skewed distribution, this approach can lead to bins containing a low number of values, potentially overlooking important data patterns.

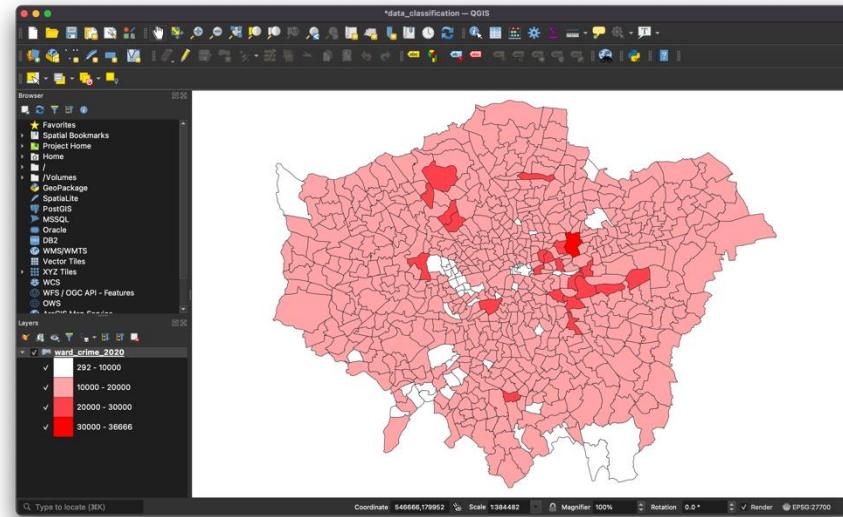
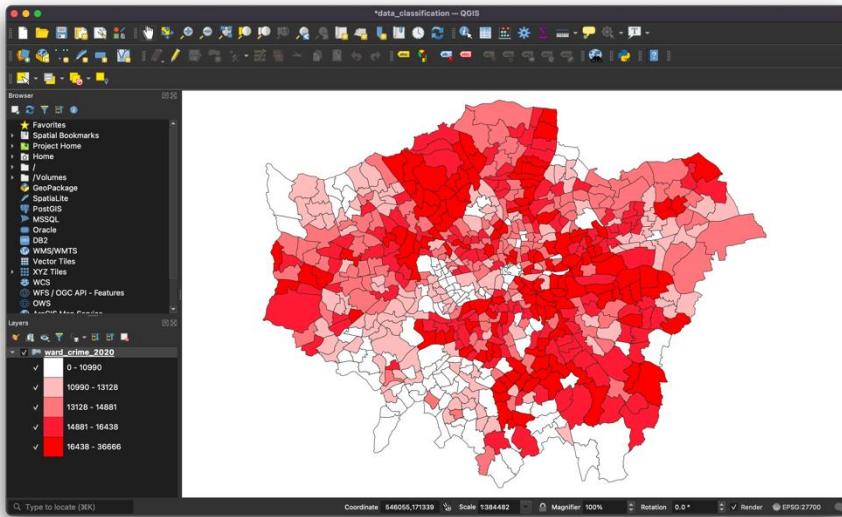
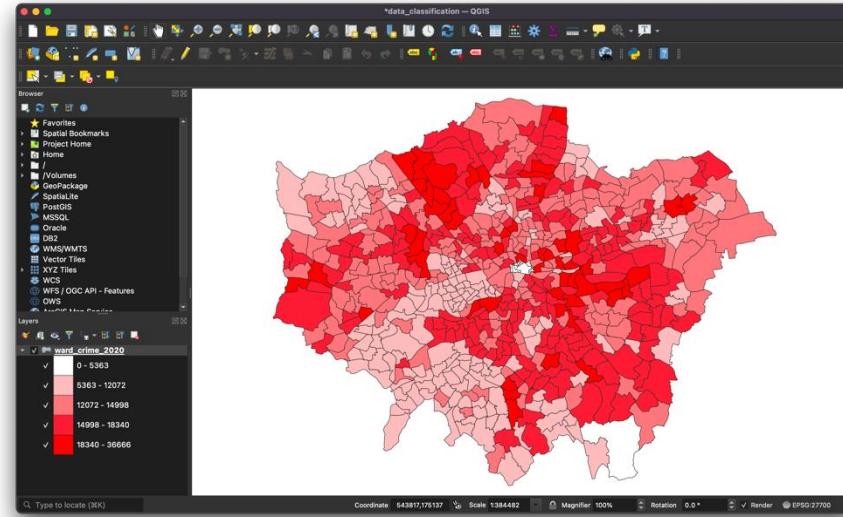
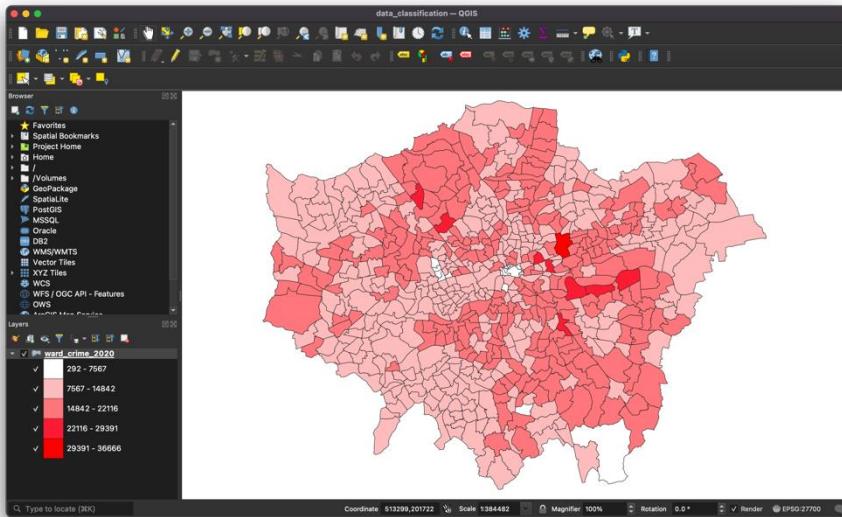
# Mapping 101



# Mapping 101



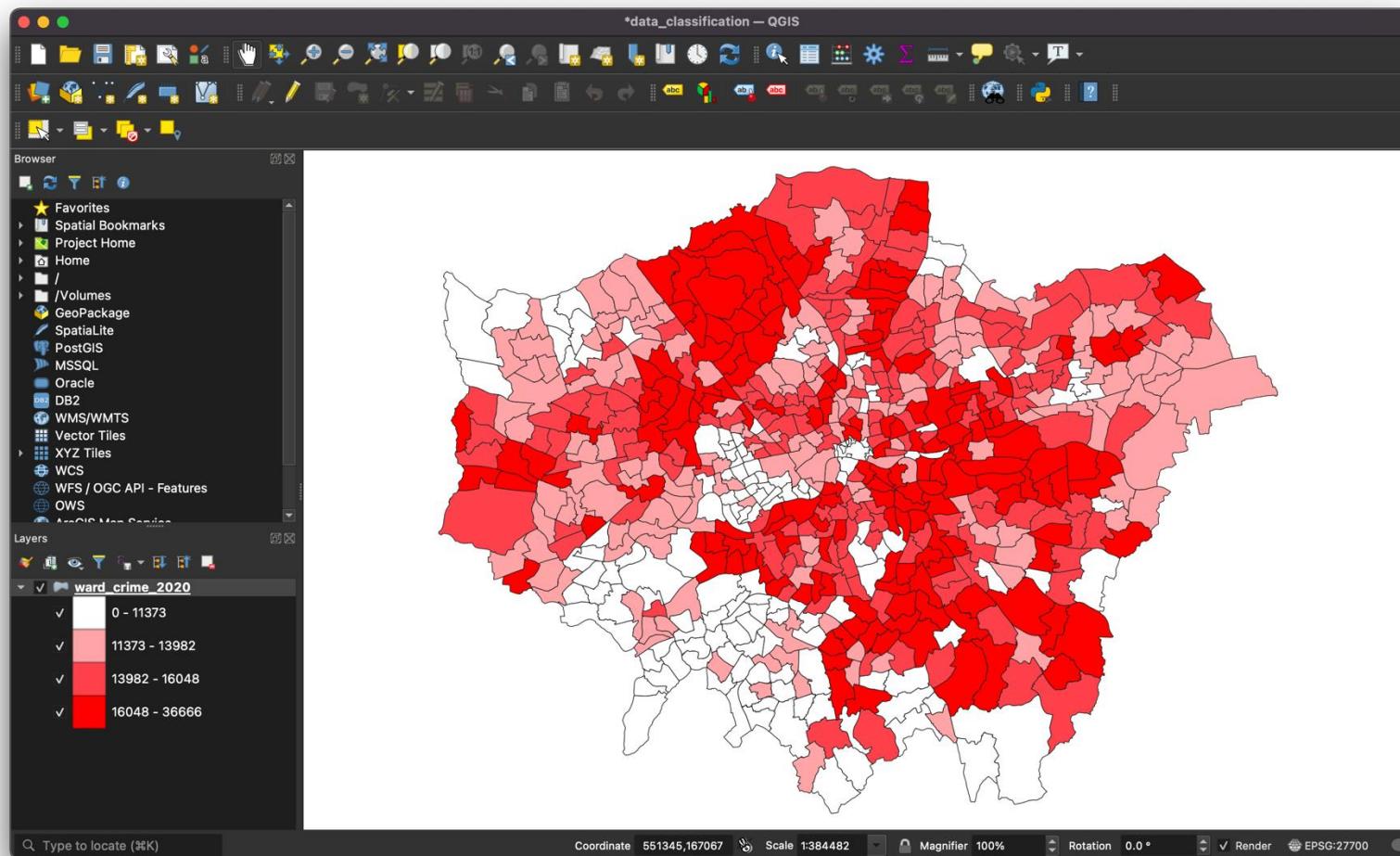
# Different classifications, different maps



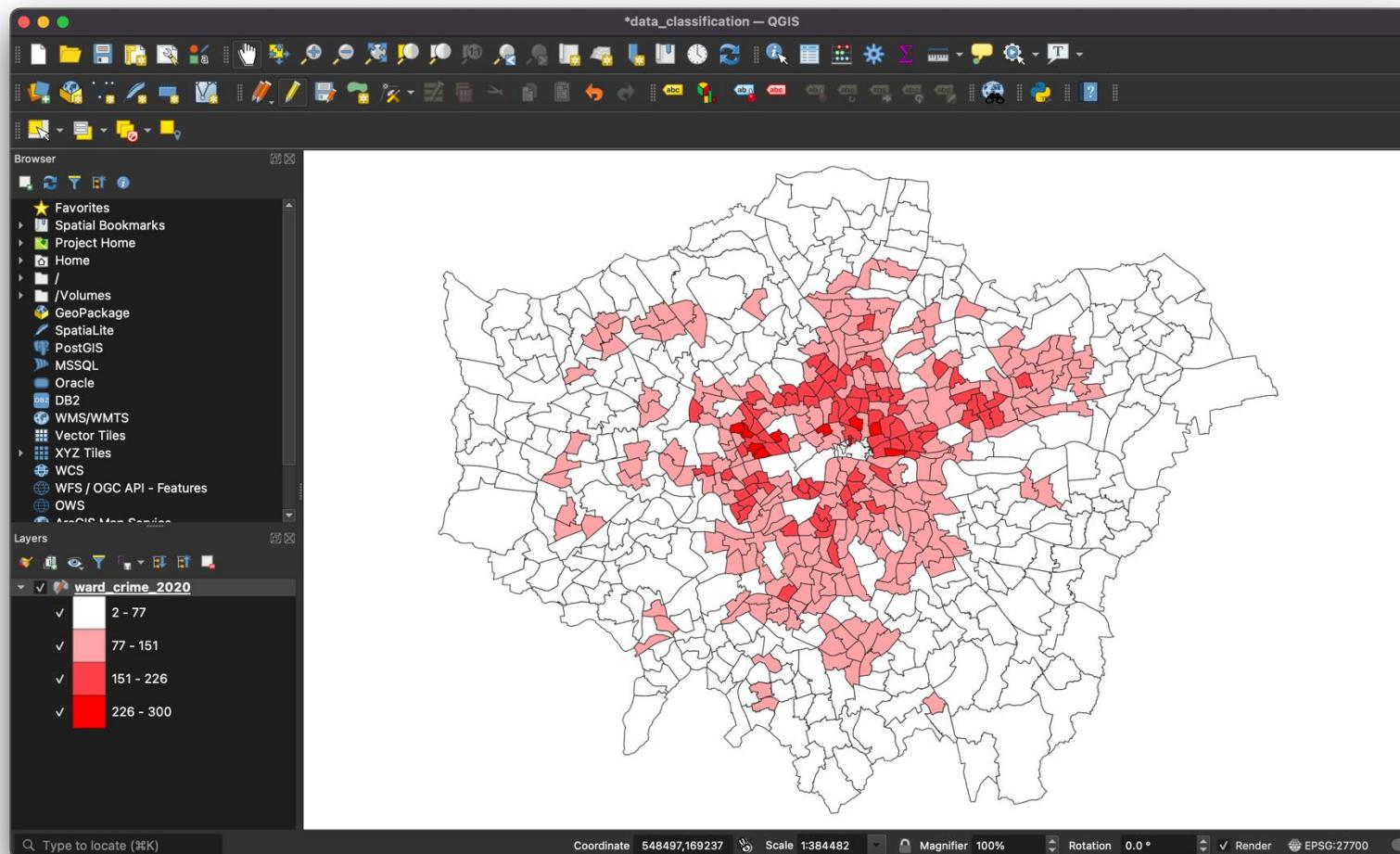
# Mapping 101

**Standardisation:** Plotting raw numbers on choropleth maps should generally be avoided, as it can be misleading. Standardisation is achieved by dividing the variable of interest (numerator) by a standardising variable (denominator), such as total population, area, or another relevant metric. This approach allows for more accurate and meaningful comparisons across different regions or units.

# Mapping 101



# Mapping 101



# Mapping 101

Publishable maps should include the following essential elements:

- Title
- Legend
- Scale bar / scale text
- North arrow
- And: attribution and data sources

Optional:

- Text and labels can be important for providing context, but may sometimes be distracting, depending on the map's purpose.

# Conclusion

- Maps are important tools to effectively convey spatial heterogenous information.
- At least two things are required: GIScience and GISystems.
- There is no one map to rule them all.

# Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk



# Practical

[jtvdijk.github.io/SA-TIED](https://jtvdijk.github.io/SA-TIED)

