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

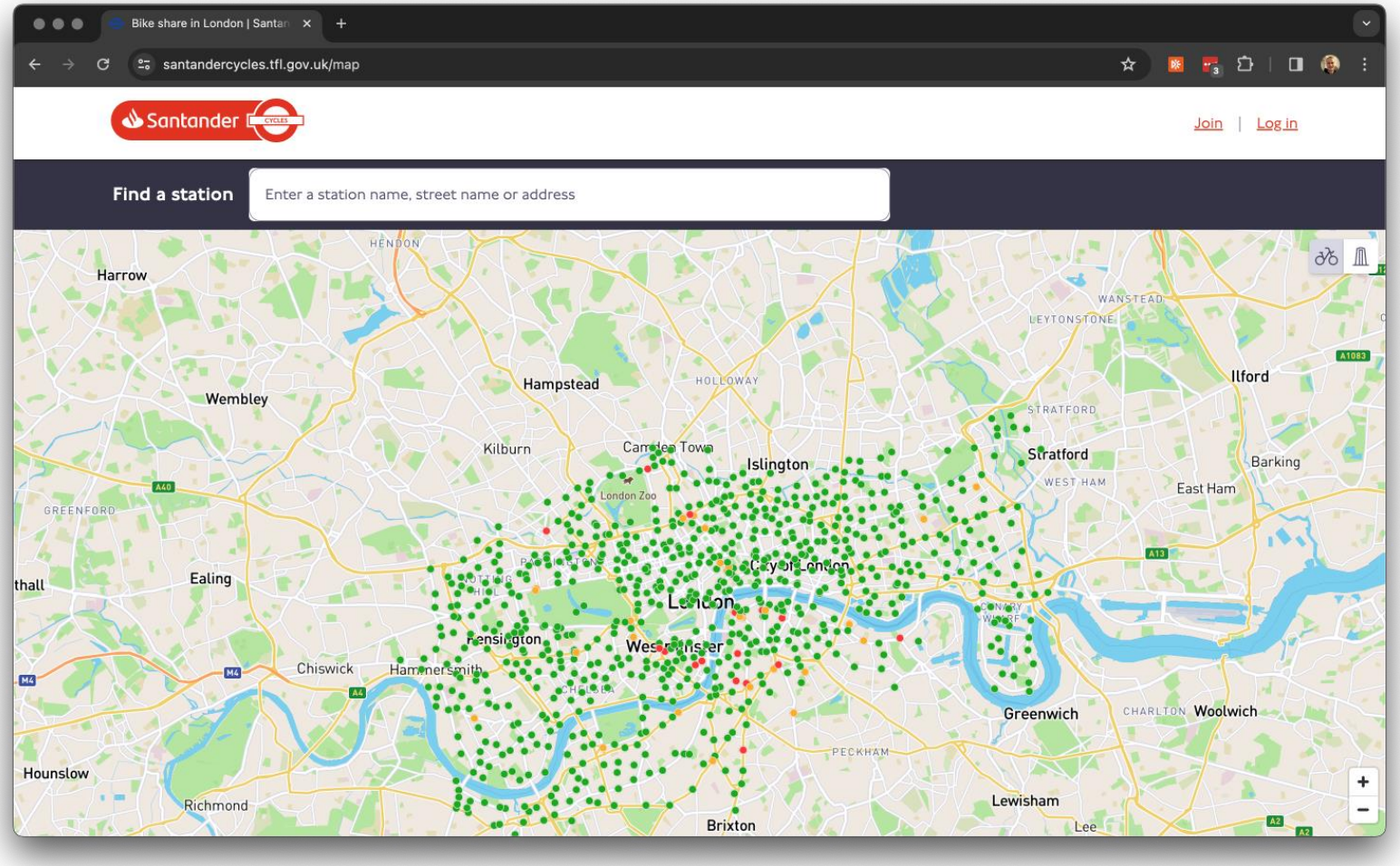
This session

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

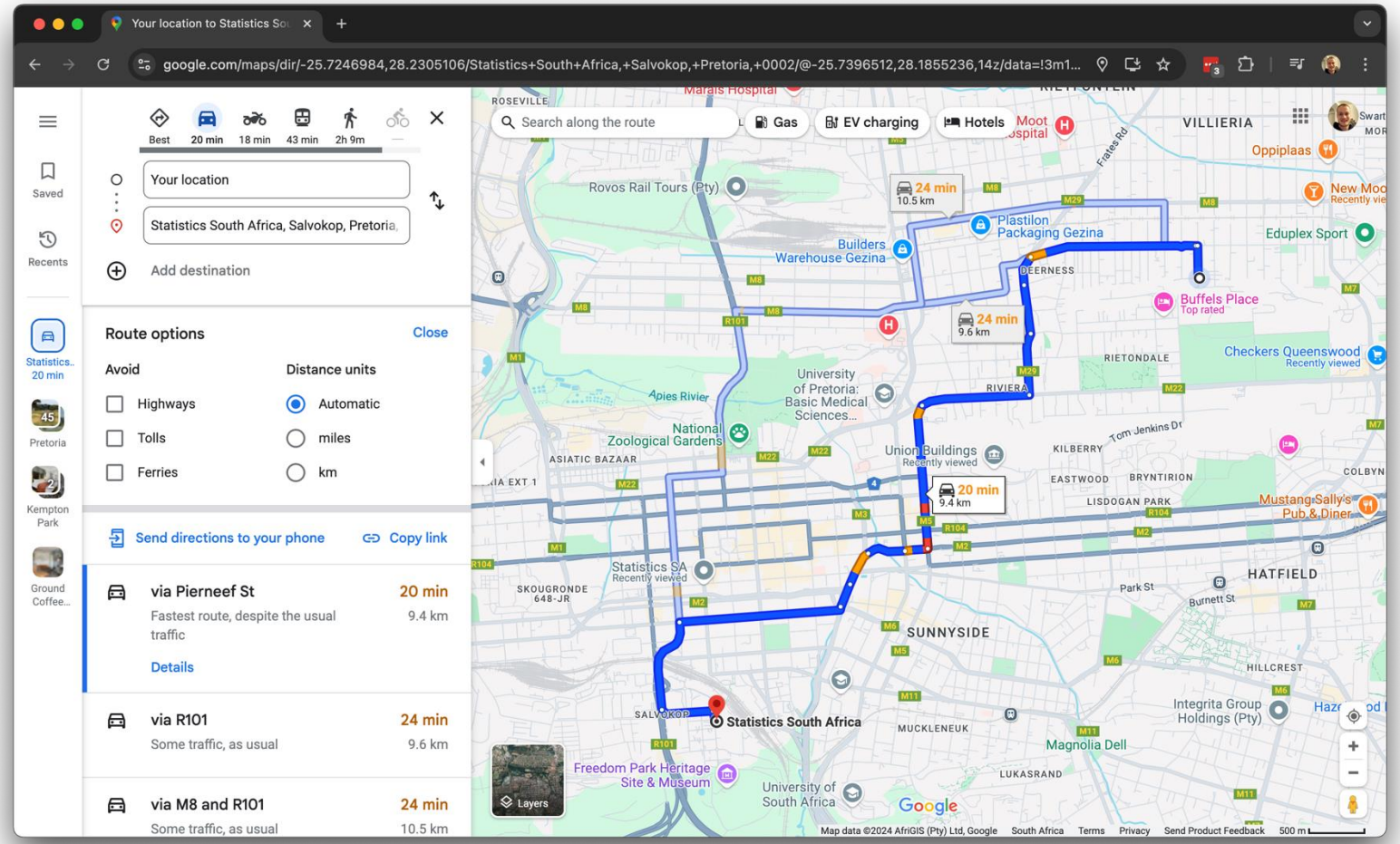
Why is mapping important?

- 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.

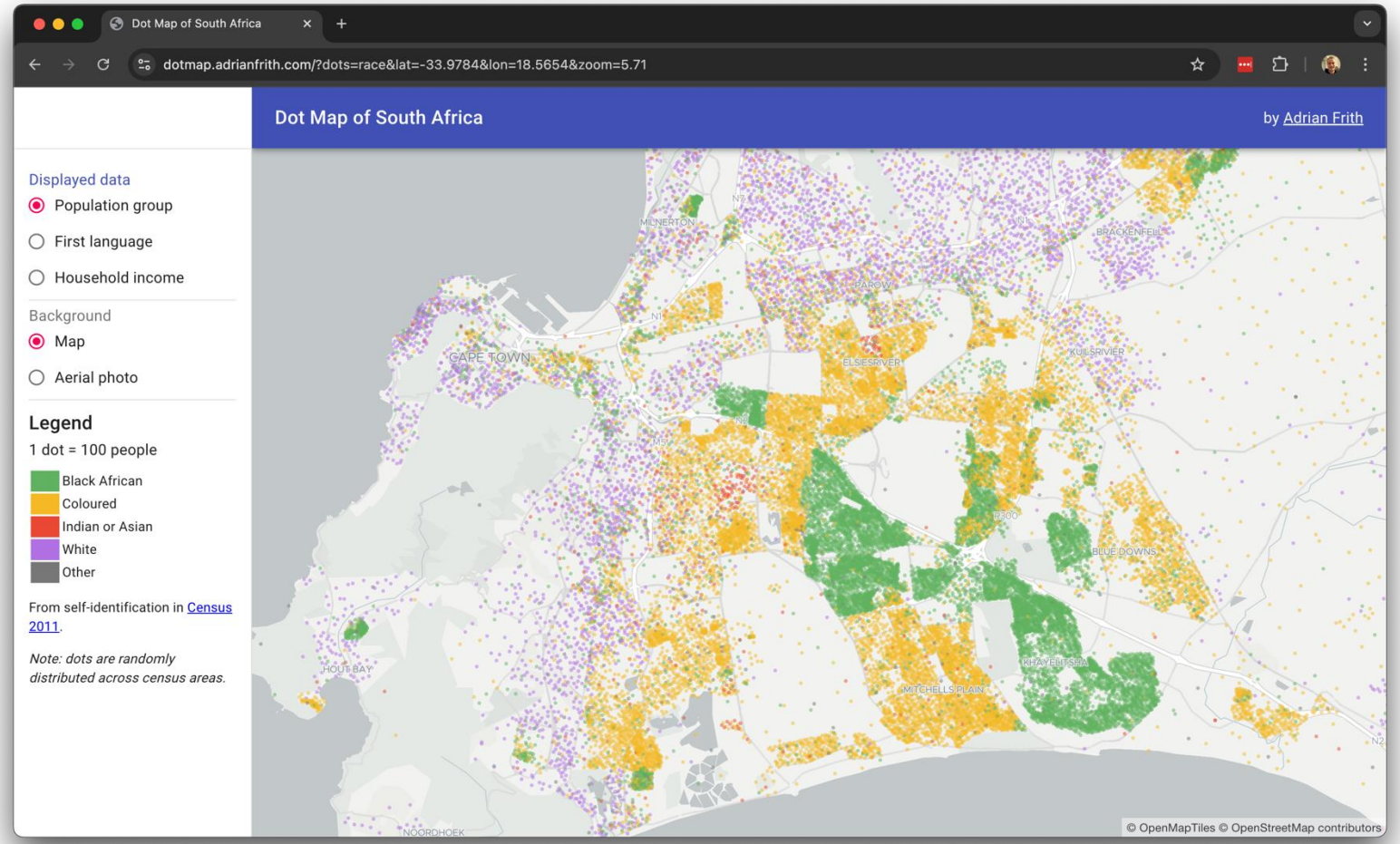
Bicycle docking stations in London



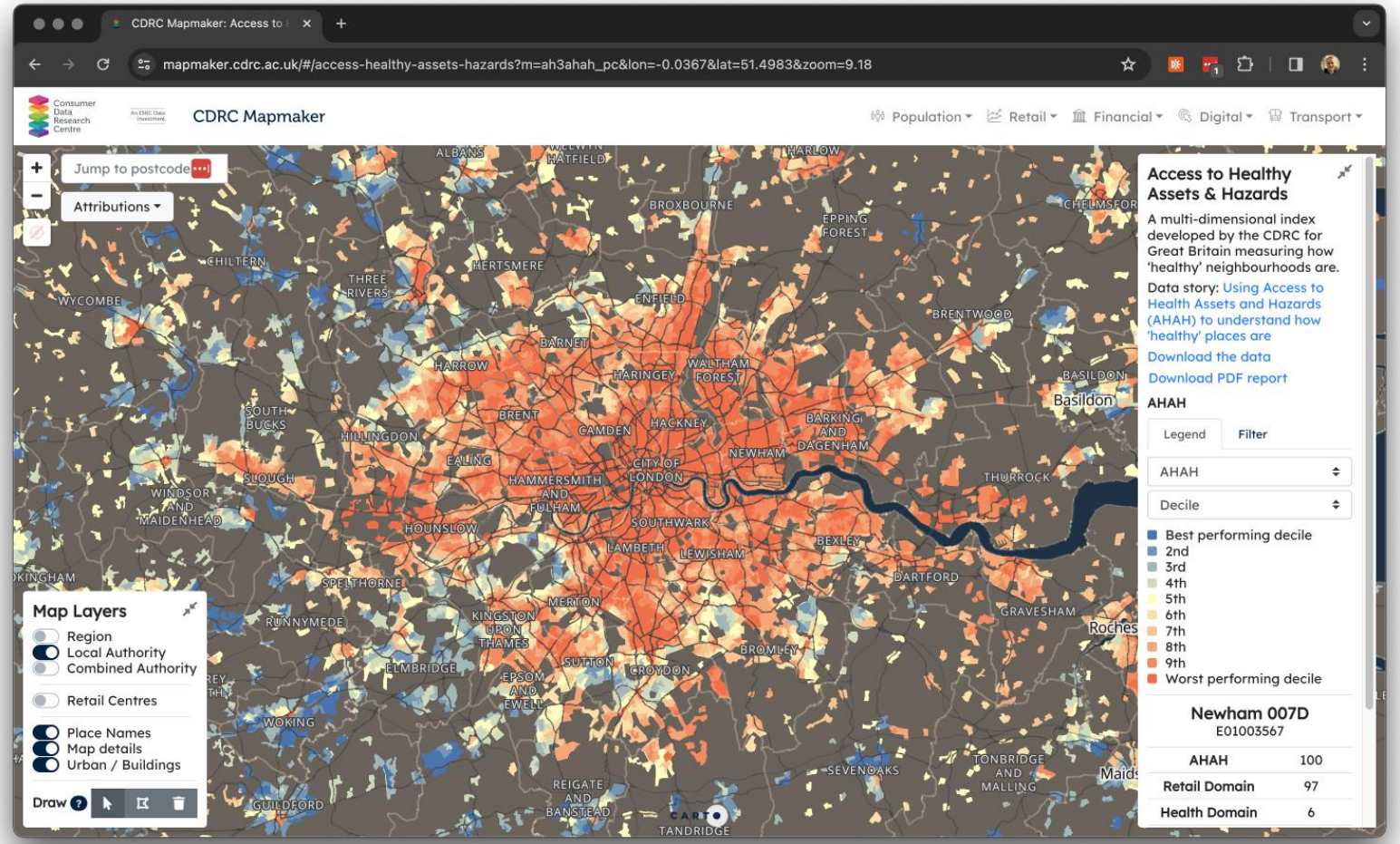
Finding your way in Pretoria



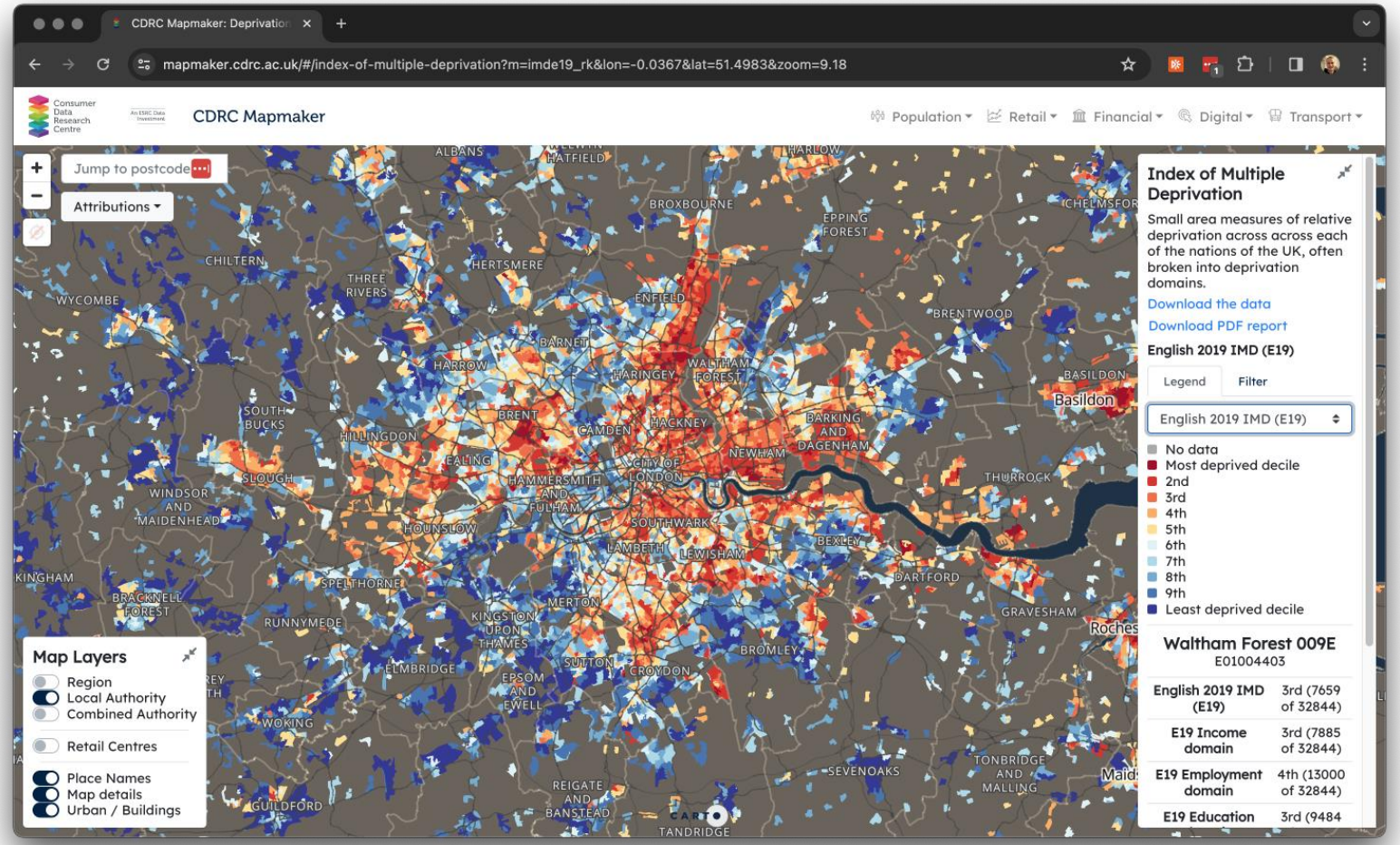
Segregation in Cape Town



Access to Healthy Assets and Hazards in the UK



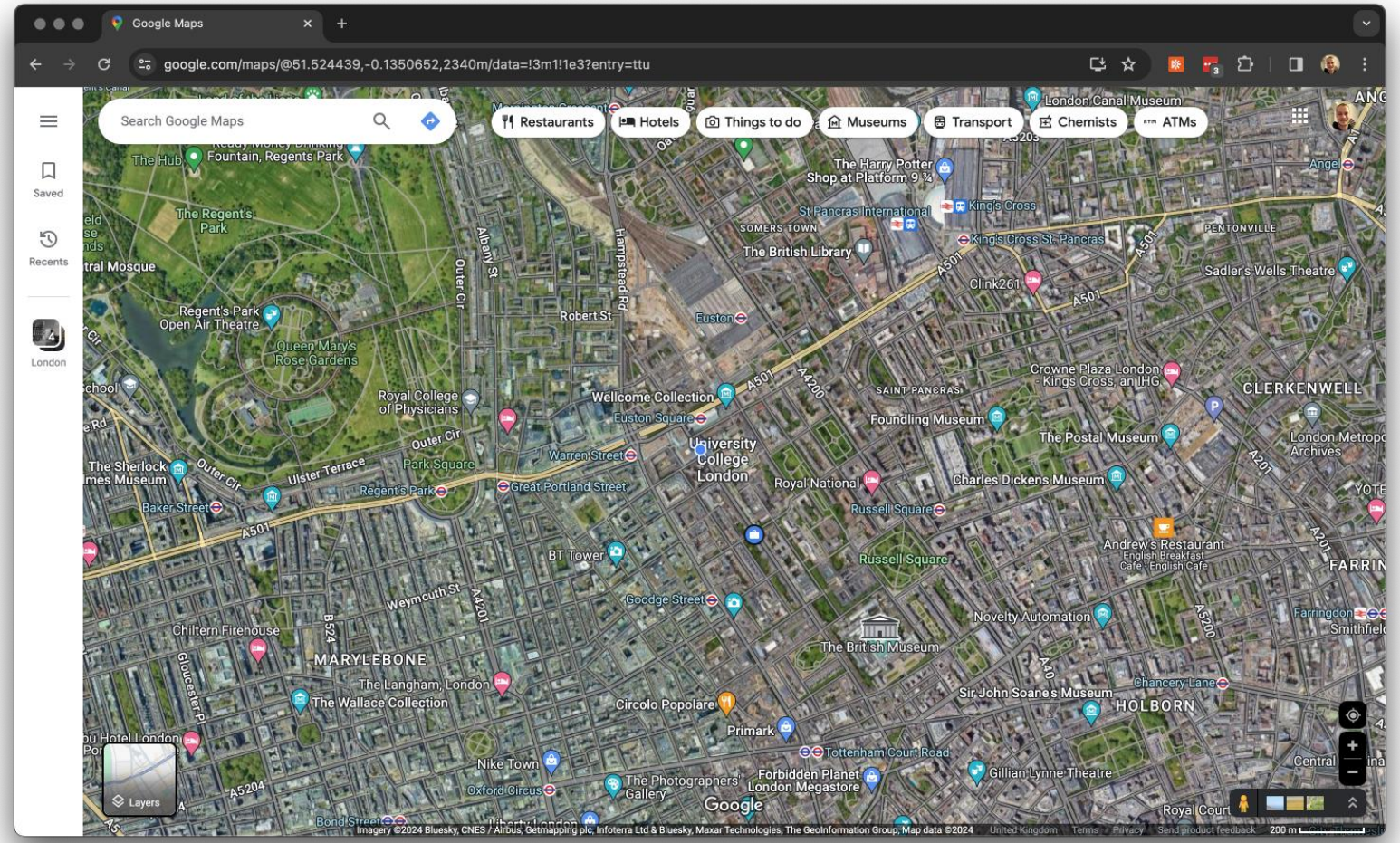
Index of Multiple Deprivation in England



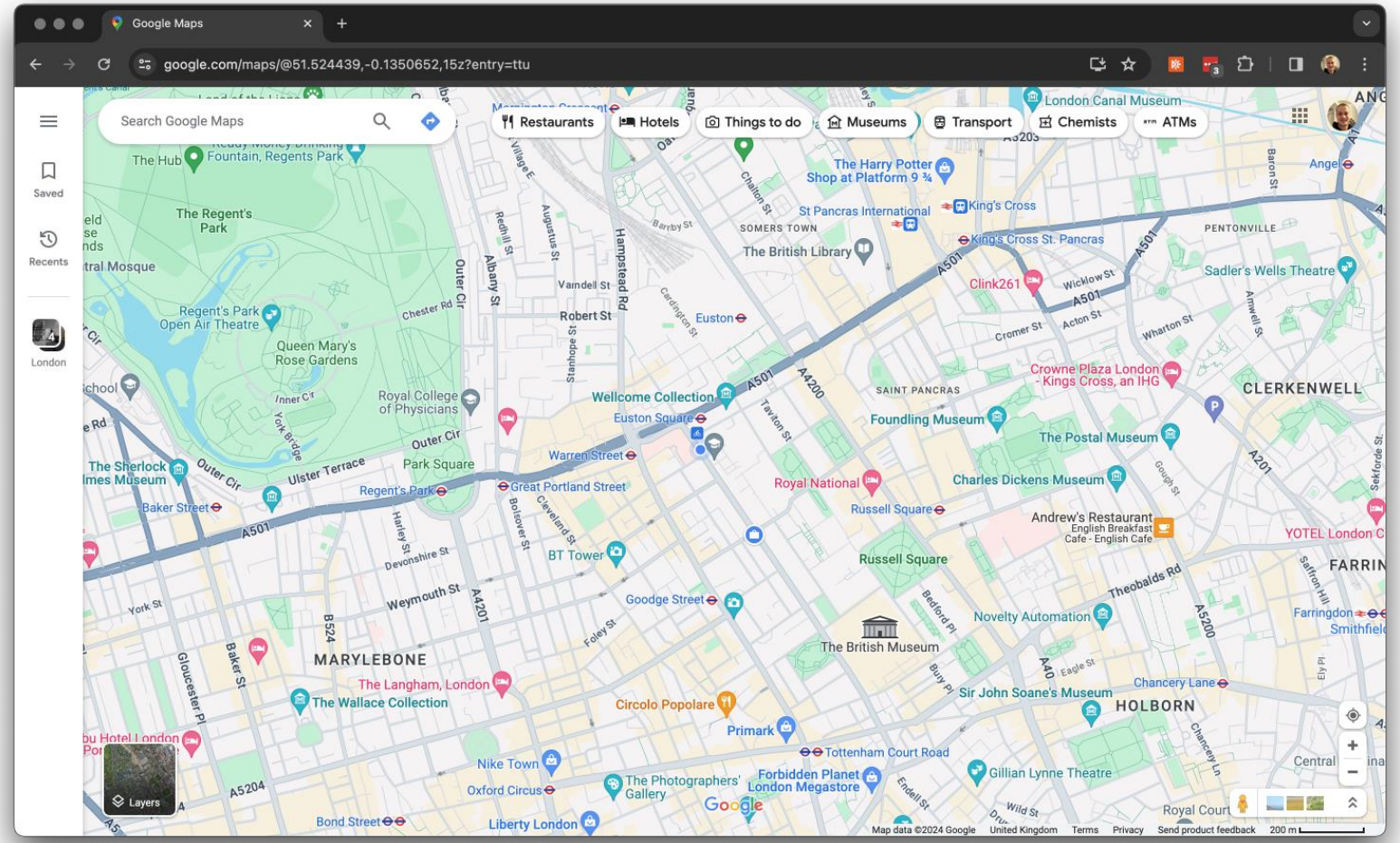
Requirements

- GIScience: theory
- GISystems: methods and tools

GIScience


Google Maps
google.com/maps/@51.524439,-0.1350652,2340m/data=!3m1!1e3?entry=ttu
Search Google Maps
Restaurants
Hotels
Things to do
Museums
Transport
Chemists
ATMs
Saved
Recents
London
Layers
The Regent's Park
Regent's Park Open Air Theatre
Queen Mary's Rose Gardens
Royal College of Physicians
Wellcome Collection
Euston Square
University College London
Warren Street
Great Portland Street
Regent's Park
Baker Street
Park Square
Outer Cir
Ulster Terrace
The Sherlock Holmes Museum
Robert St
Euston
Albany St
Hampstead Rd
A501
A4200
A4201
A5200
A40
A5204
B524
The British Library
St Pancras International
King's Cross
King's Cross St. Pancras
SOMERS TOWN
PENTONVILLE
SAINT PANCRAS
CLERKENWELL
MARYLEBONE
HOLBORN
FARRIN
The Harry Potter Shop at Platform 9 ¾
London Canal Museum
Sadler's Wells Theatre
Clink261
Crowne Plaza London Kings Cross, an IHG
Foundling Museum
The Postal Museum
Charles Dickens Museum
Royal National
Russell Square
Andrew's Restaurant
English Breakfast Cafe · English Cafe
BT Tower
Goodge Street
The British Museum
Novelty Automation
Sir John Soane's Museum
Chancery Lane
Tottenham Court Road
Primark
Circolo Popolare
Nike Town
Oxford Circus
The Photographers' Gallery
Forbidden Planet London Megastore
Gillian Lynne Theatre
Royal Court
Chiltern Firehouse
The Langham, London
The Wallace Collection
Weymouth St
Gloucester Pl
Google
Imagery ©2024 Bluesky, CNES / Airbus, Getmapping plc, Infoterra Ltd & Bluesky, Maxar Technologies, The GeoInformation Group, Map data ©2024
United Kingdom
Terms
Privacy
Send product feedback
200 m

GIScience

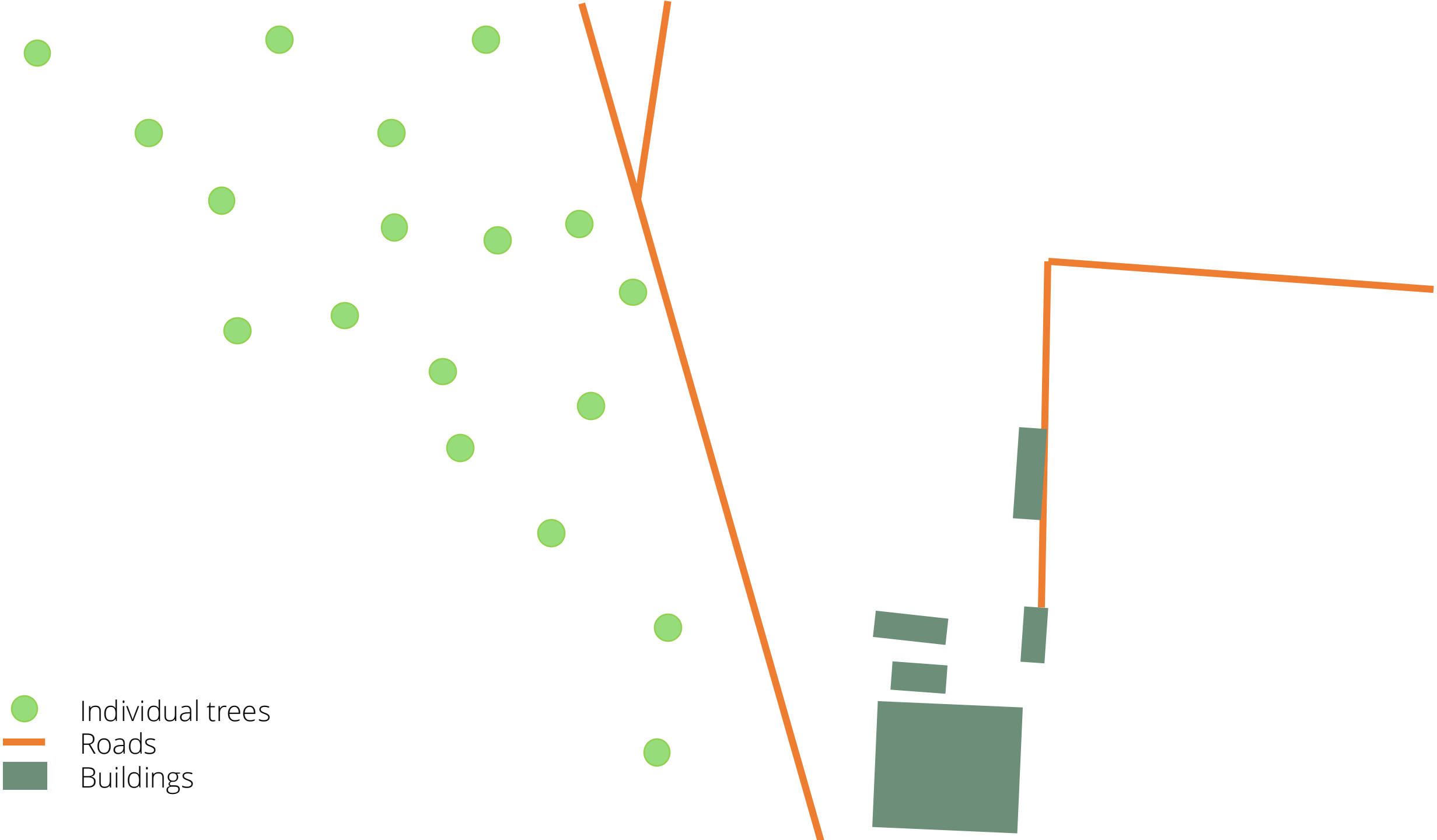

Google Maps
google.com/maps/@51.524439,-0.1350652,15z?entry=ttu
Search Google Maps
Restaurants
Hotels
Things to do
Museums
Transport
Chemists
ATMs
Saved
Recents
London
The Hub
Fountain, Regents Park
The Regent's Park
Regent's Park Open Air Theatre
Queen Mary's Rose Gardens
Royal College of Physicians
The Sherlock Holmes Museum
Baker Street
Regent's Park
Great Portland Street
Warren Street
Euston Square
Wellcome Collection
Euston
Robert St
Park Square
Outer Cir
Ulster Terrace
A501
A4200
A4201
A5204
B524
Harley St
Devonshire St
Weymouth St
Goodge Street
BT Tower
MARYLEBONE
The Langham, London
The Wallace Collection
Circolo Popolare
Primark
Nike Town
Oxford Circus
The Photographers' Gallery
Bond Street
Liberty London
Tottenham Court Road
Forbidden Planet London Megastore
Gillian Lynne Theatre
Google
The British Museum
Russell Square
Royal National
Charles Dickens Museum
Foundling Museum
The Postal Museum
Crowne Plaza London - Kings Cross, an IHG
SAINT PANCRAS
CLERKENWELL
Andrew's Restaurant
English Breakfast Cafe - English Cafe
Novelty Automation
Sir John Soane's Museum
HOLBORN
Chancery Lane
Royal Court
Theobalds Rd
A40
A5200
A201
Farringdon
Smithfield
FARRIN
YOTEL London C
The British Library
St Pancras International
King's Cross
King's Cross St. Pancras
SOMERS TOWN
PENTONVILLE
Clink261
Sadler's Wells Theatre
London Canal Museum
The Harry Potter Shop at Platform 9 ¾
Angel
Layers
Map data ©2024 Google
United Kingdom
Terms
Privacy
Send product feedback
200 m

GIScience

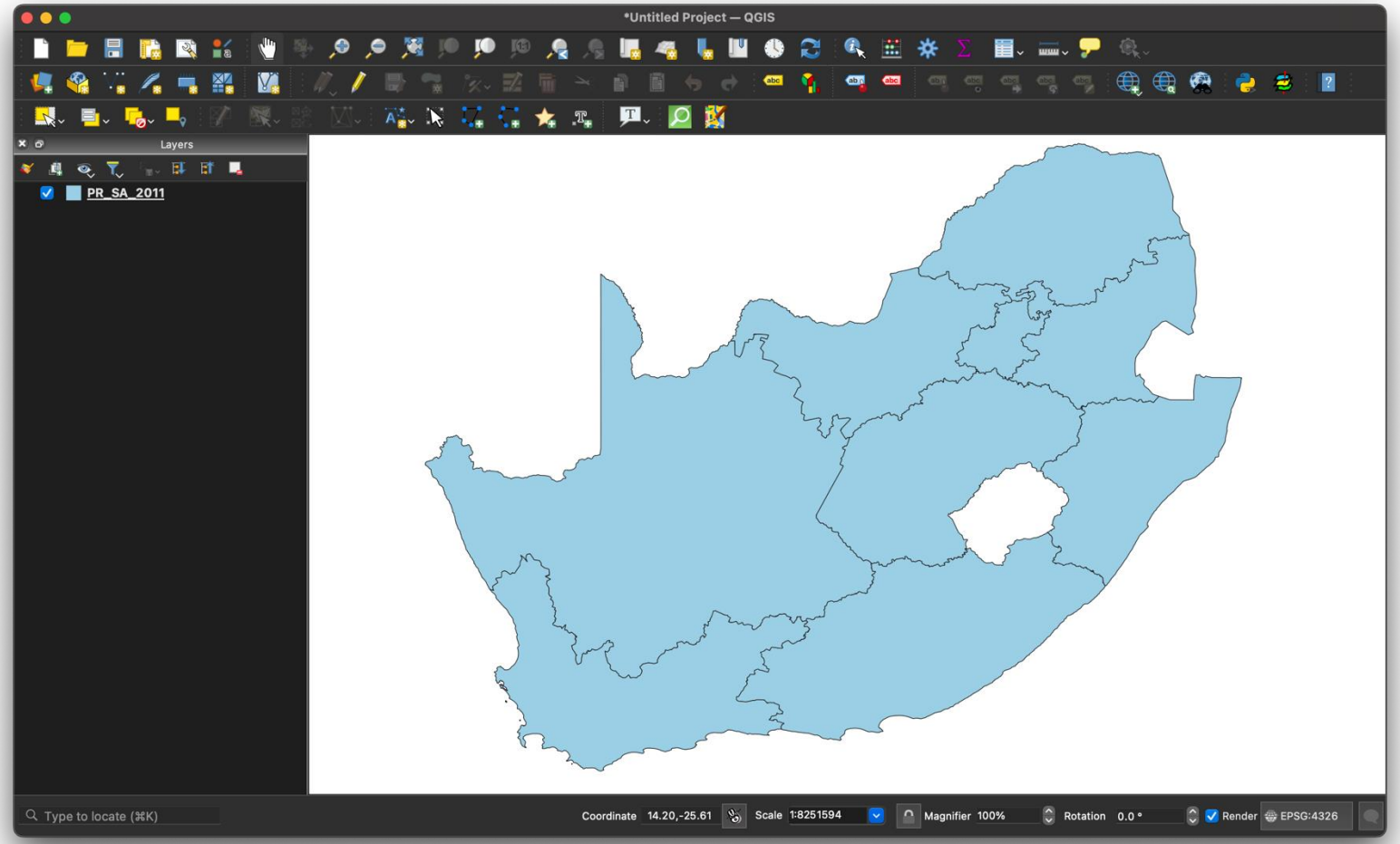
- GIScience relies on representing spatial information in a digital format. Traditionally, geographic information is conveyed in two primary ways:.

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

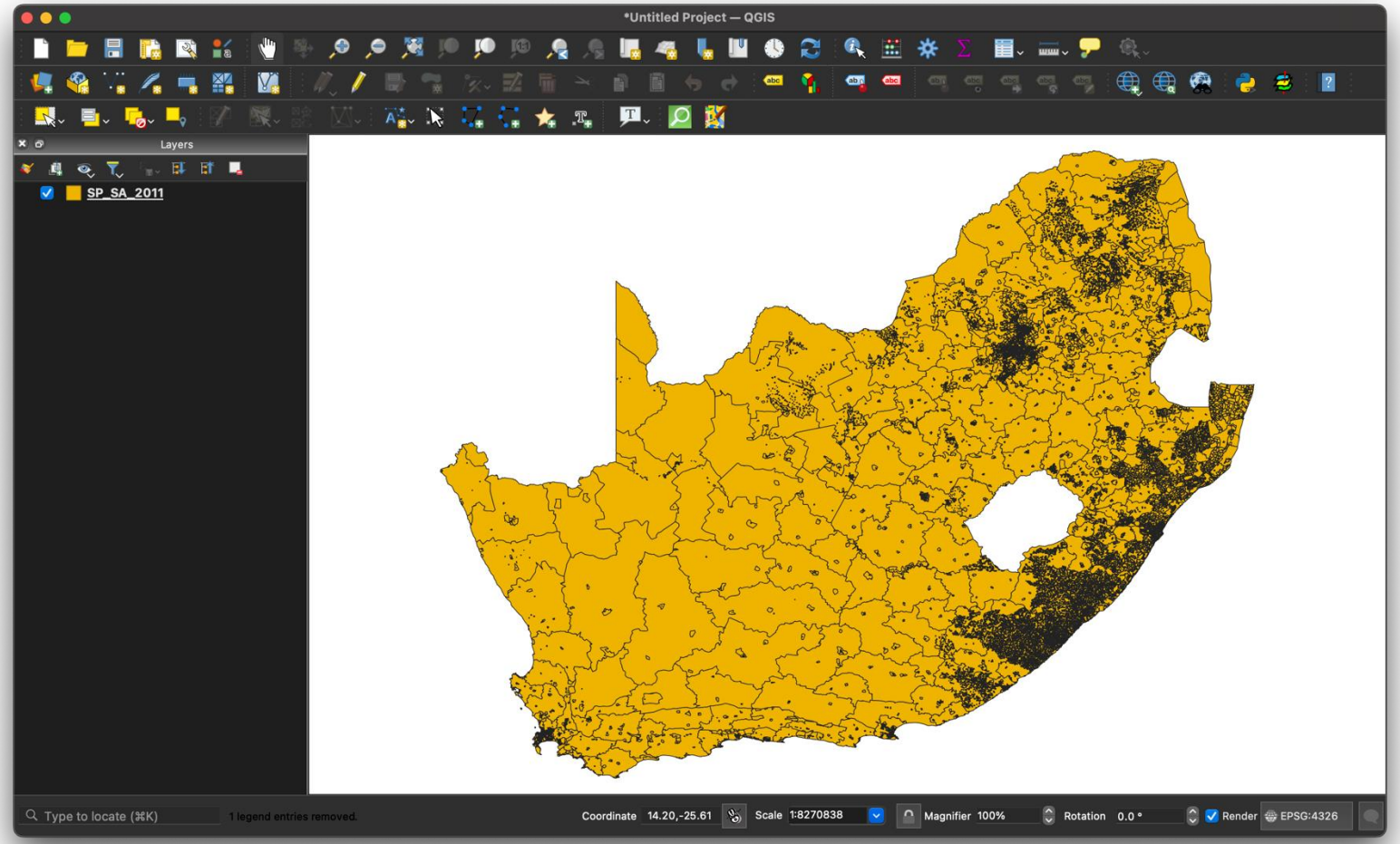
  Raster This approach employs images or grids to represent surfaces, with each cell or pixel holding a value, often indicating attributes like colors or measurements.


Individual trees
Roads
Buildings

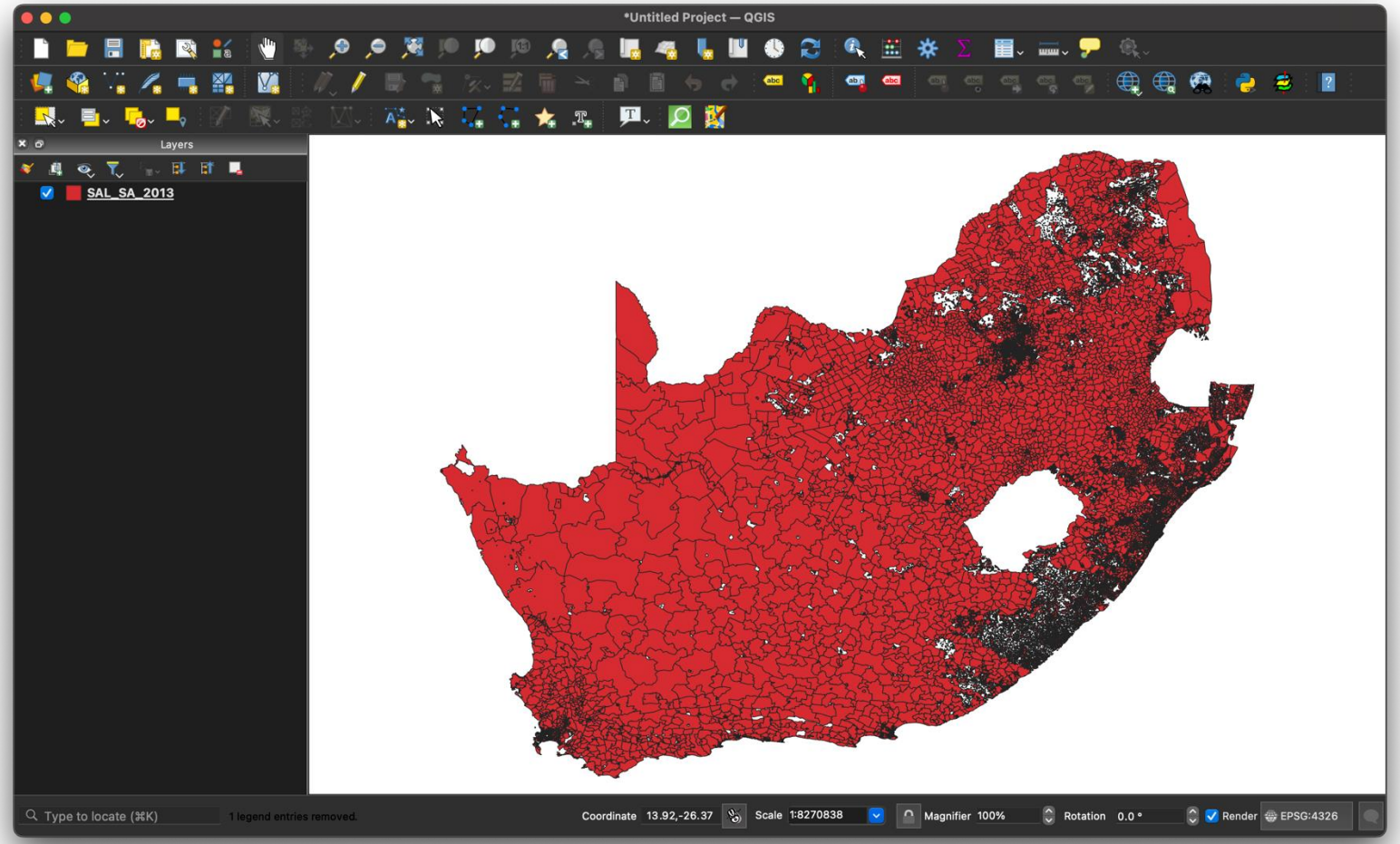
GIScience


*Untitled Project — QGIS
Layers
PR_SA_2011
Type to locate (⌘K)
Coordinate 14.20,-25.61
Scale 1:8251594
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326

GIScience


*Untitled Project — QGIS
Layers
SP_SA_2011
Type to locate (⌘K)
1 legend entries removed.
Coordinate 14.20,-25.61
Scale 1:8270838
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326

GIScience


*Untitled Project — QGIS
Layers
SAL_SA_2013
Type to locate (⌘K)
1 legend entries removed.
Coordinate 13.92,-26.37
Scale 1:8270838
Magnifier 100%
Rotation 0.0 °
Render
EPSG:4326

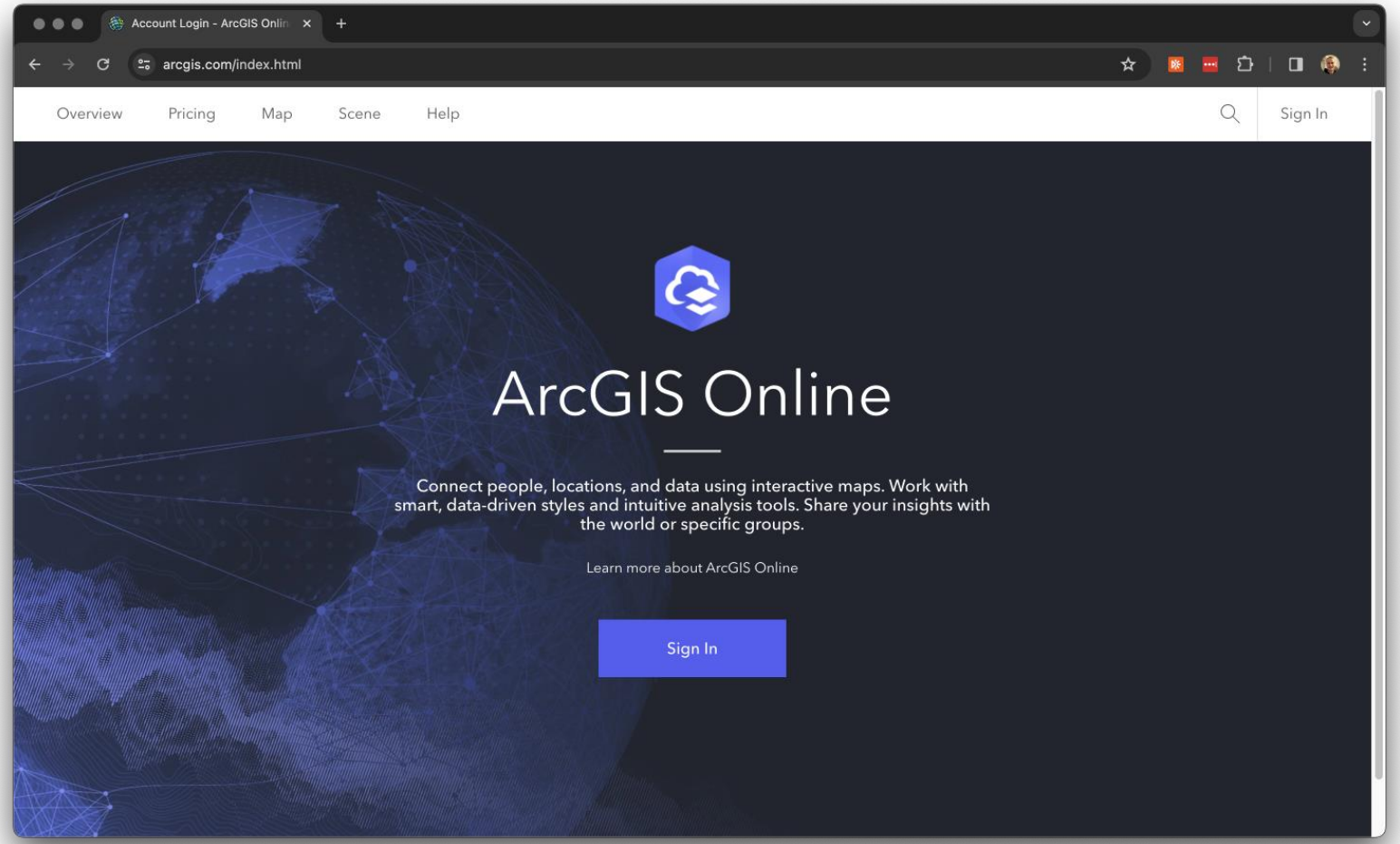
Father of GIS

- 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.

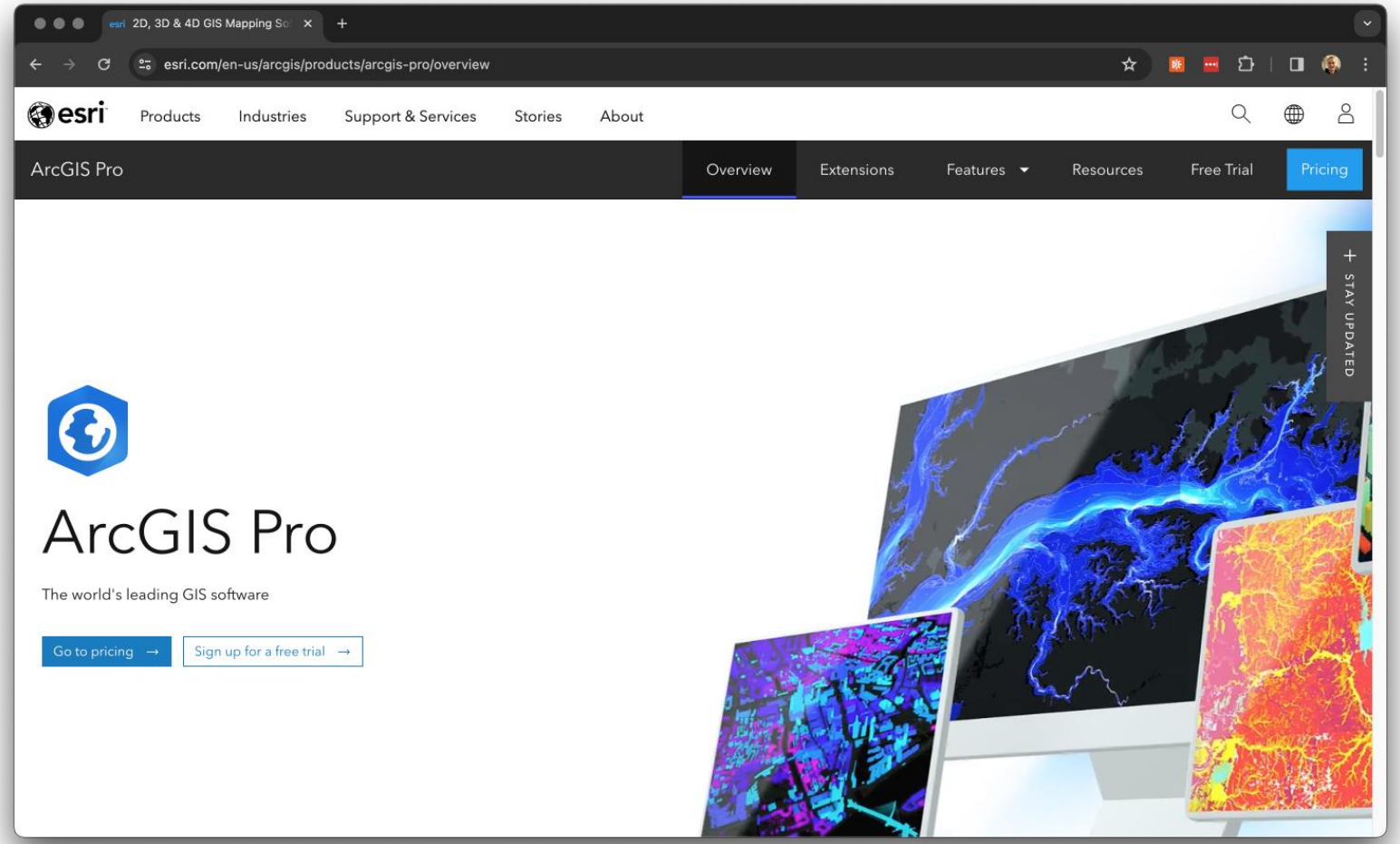
Geographic Information Systems

- 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

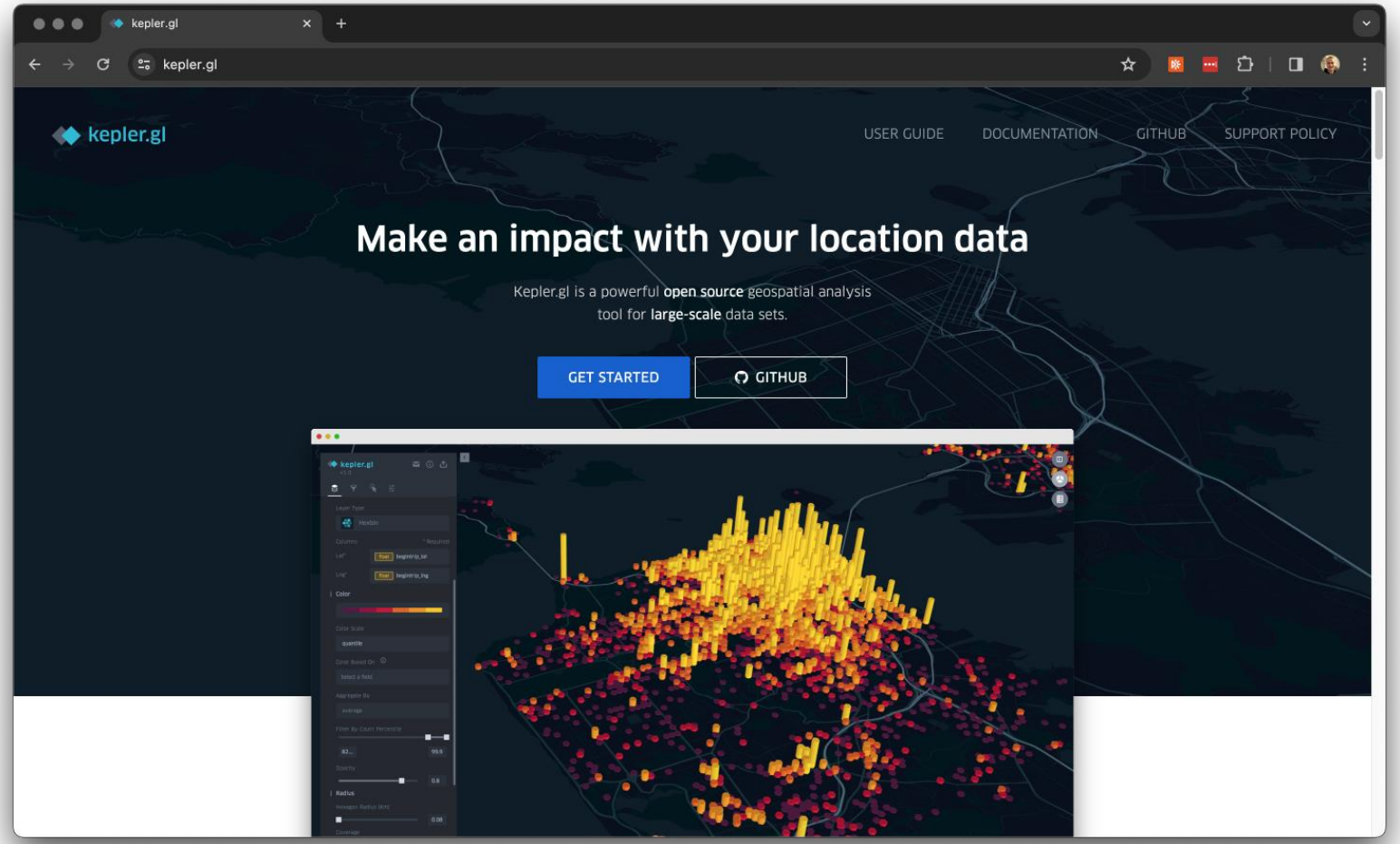
Geographic Information Systems



Geographic Information Systems

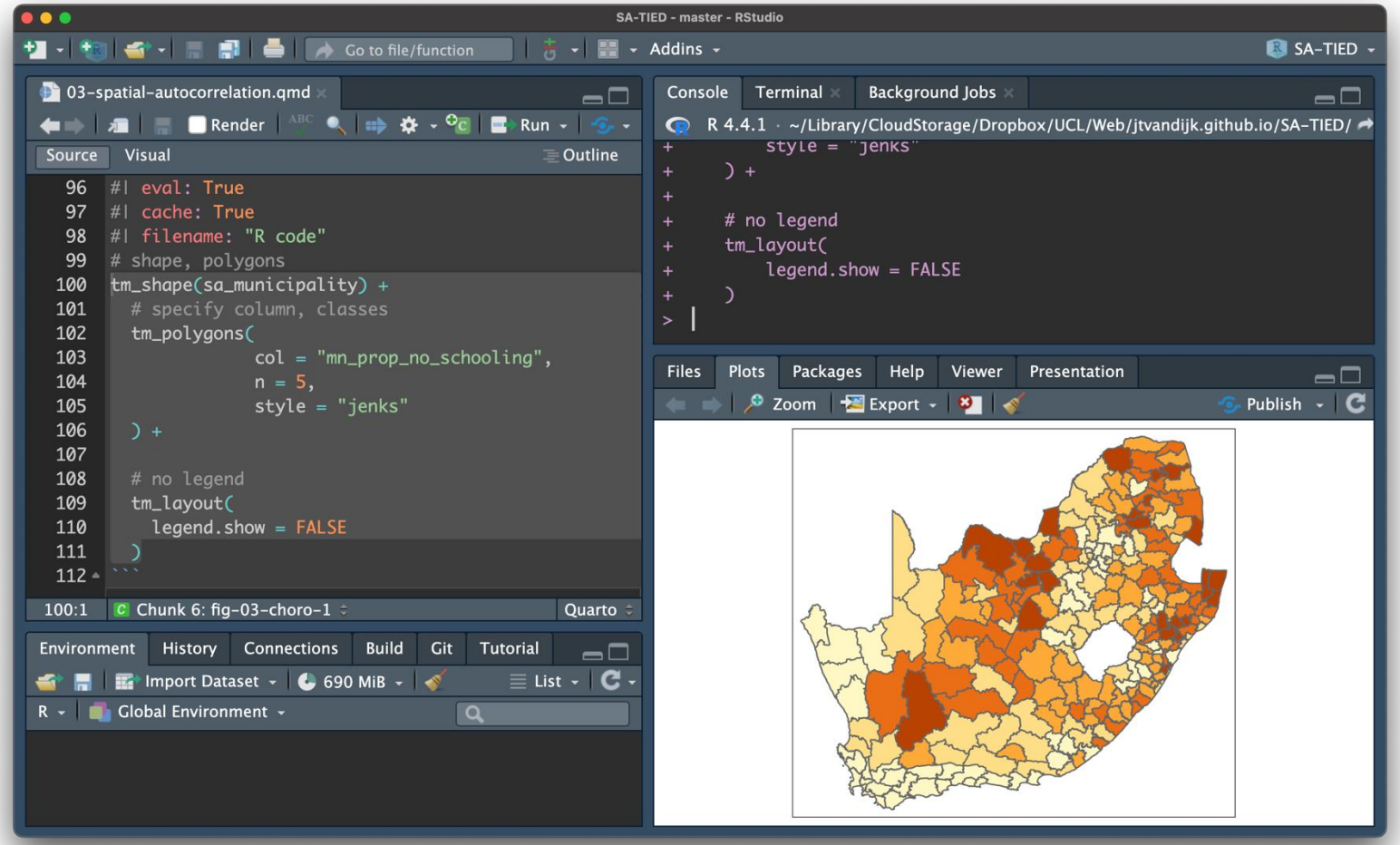


Geographic Information Systems



Geographic Information Systems

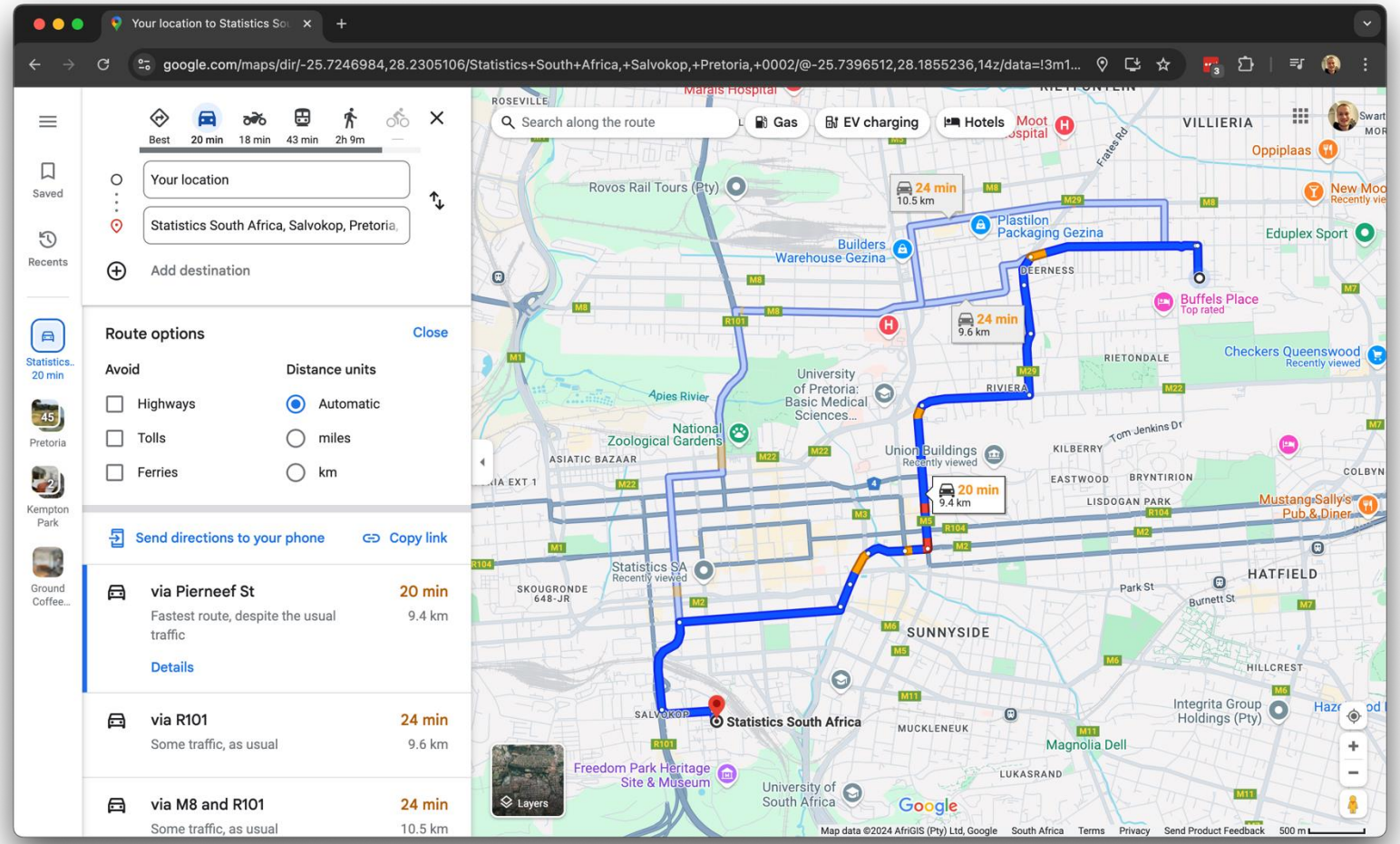
Geographic Information Systems



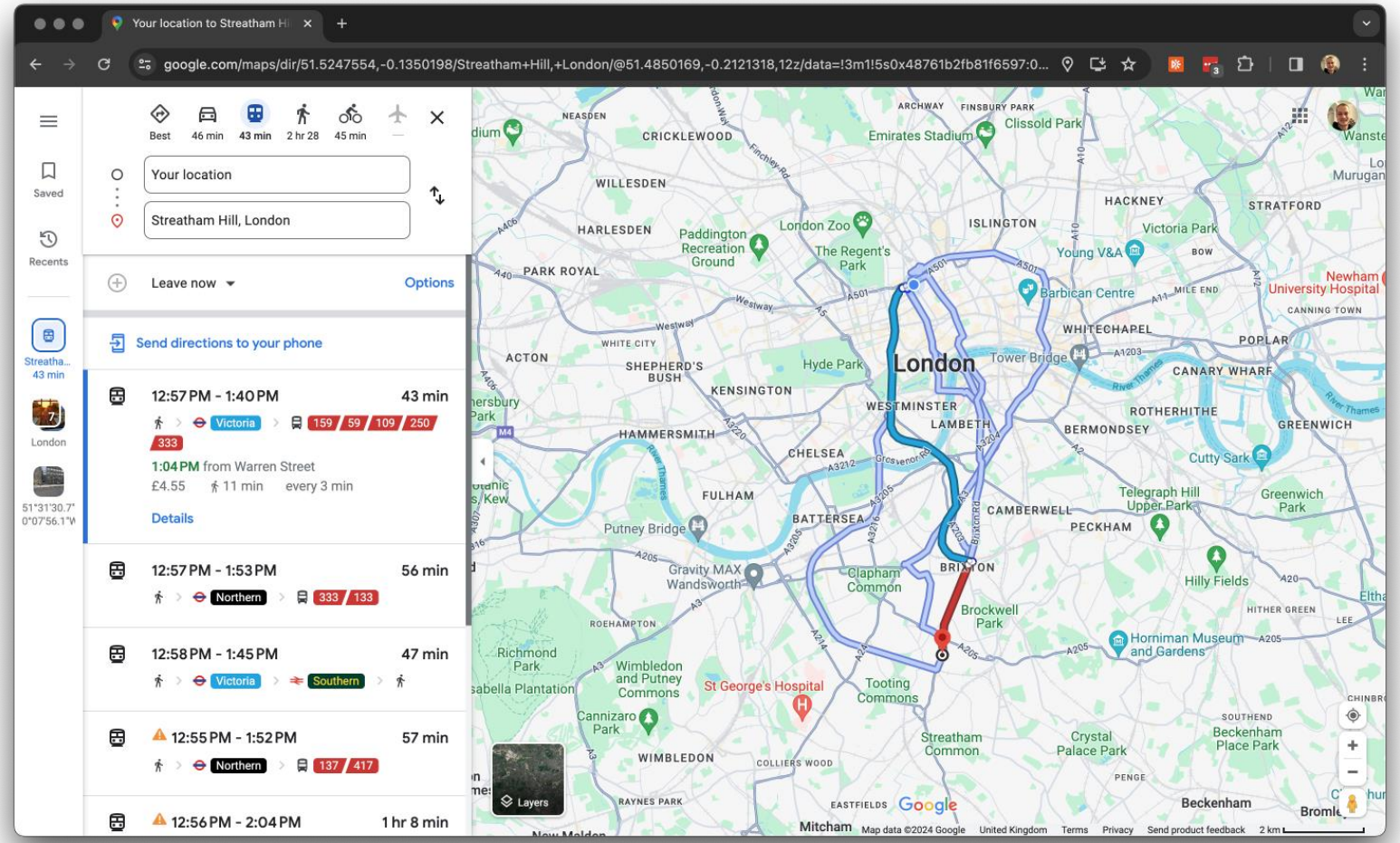
Spatial analysis

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

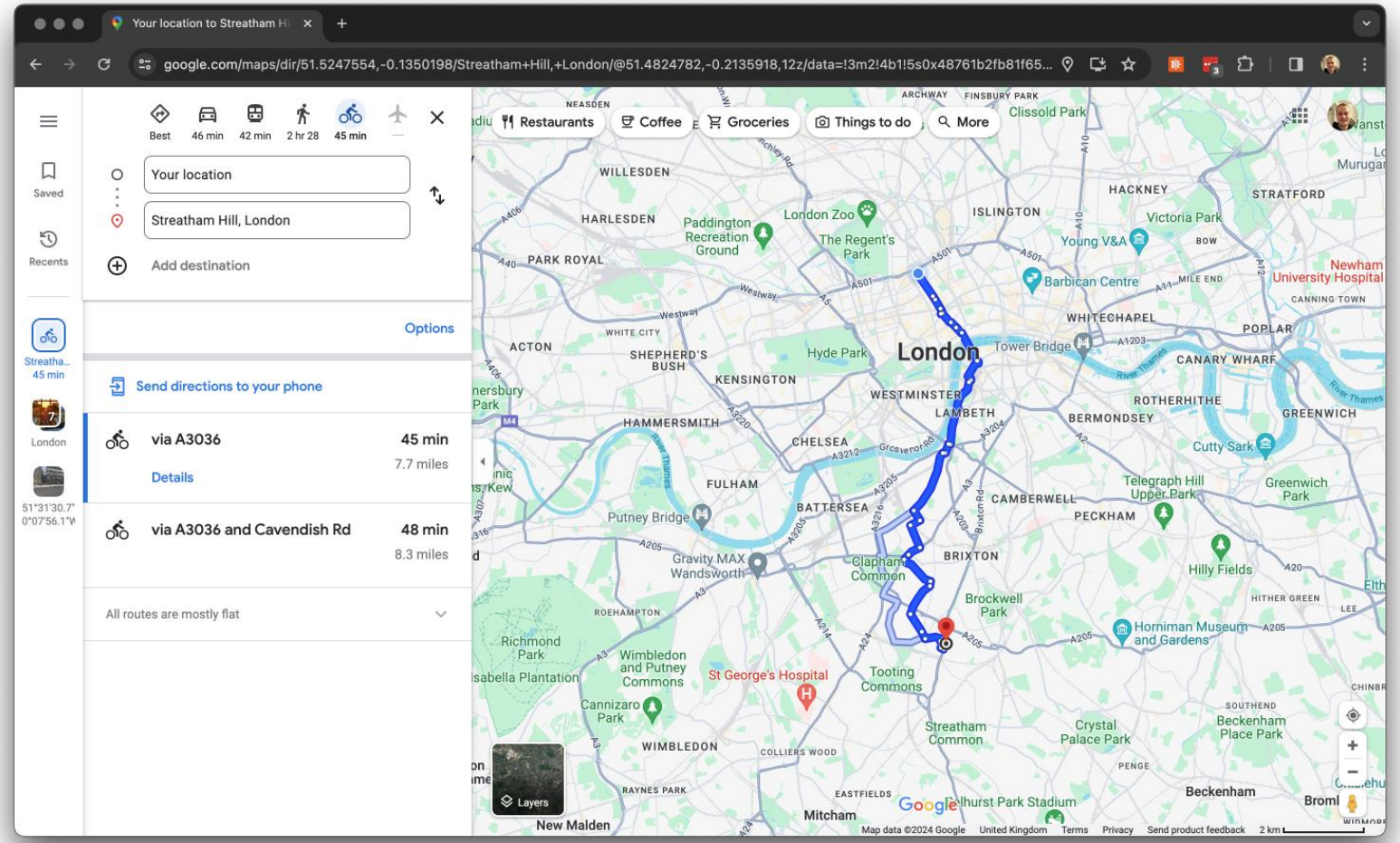
Spatial analysis



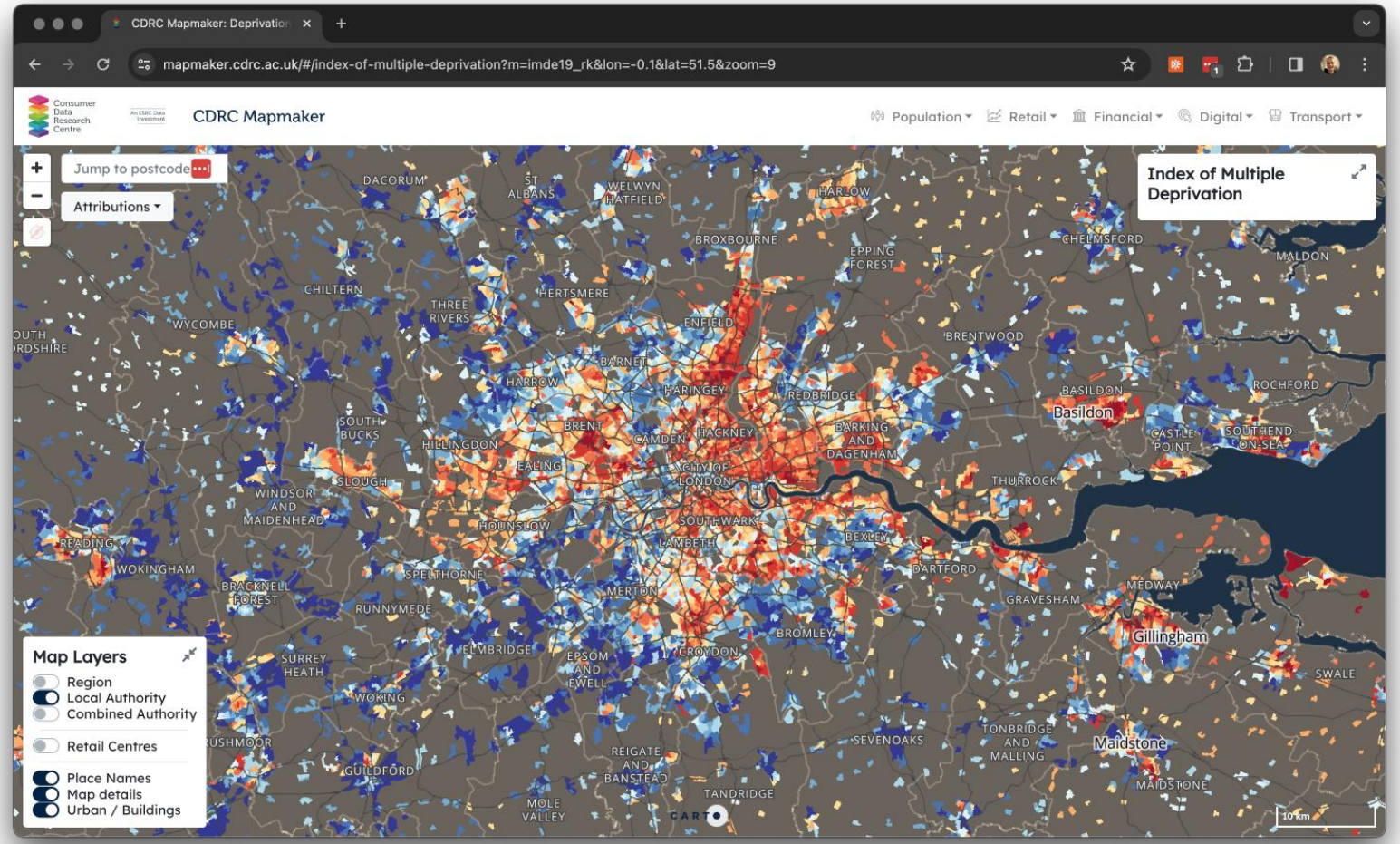
Spatial analysis



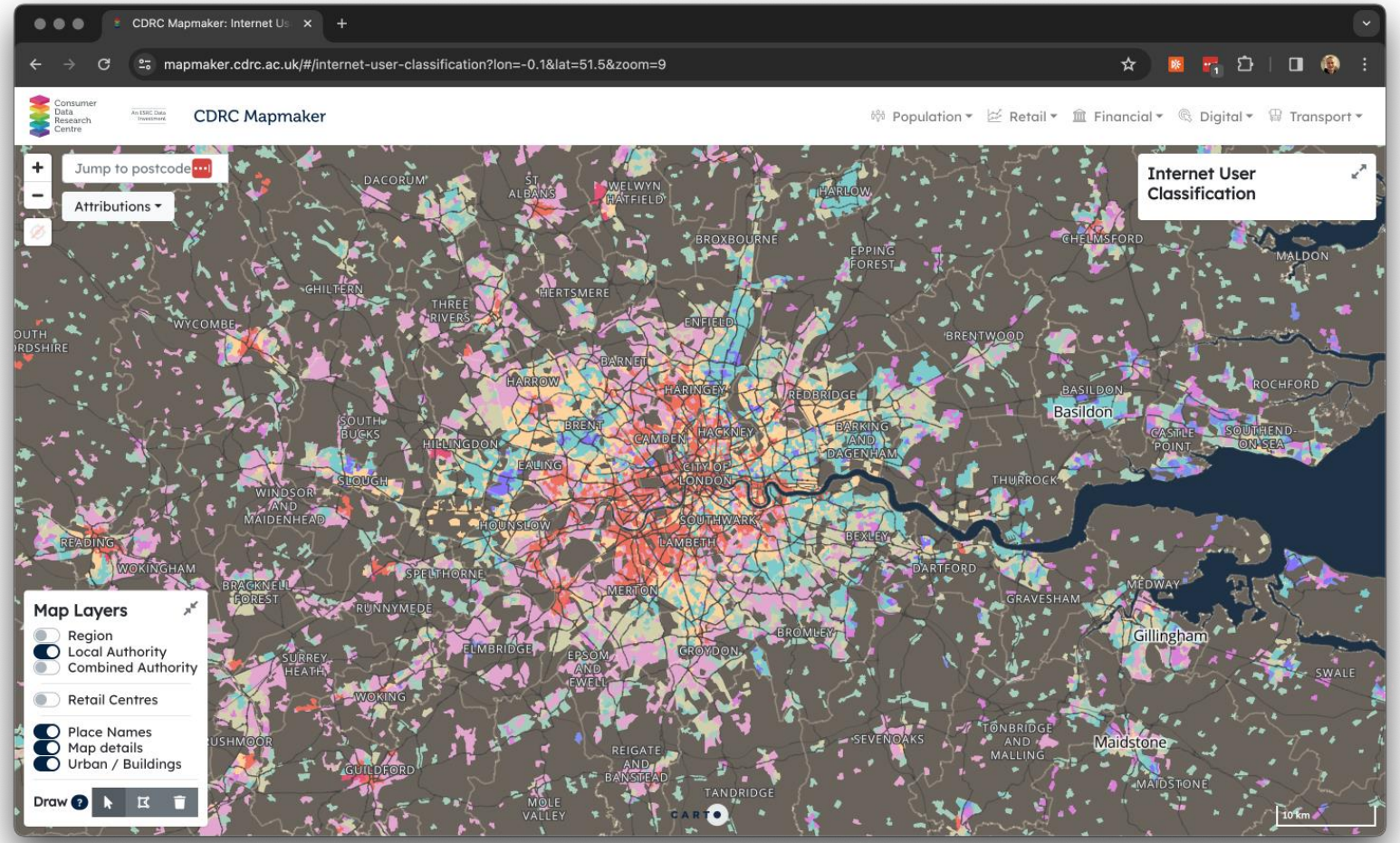
Spatial analysis



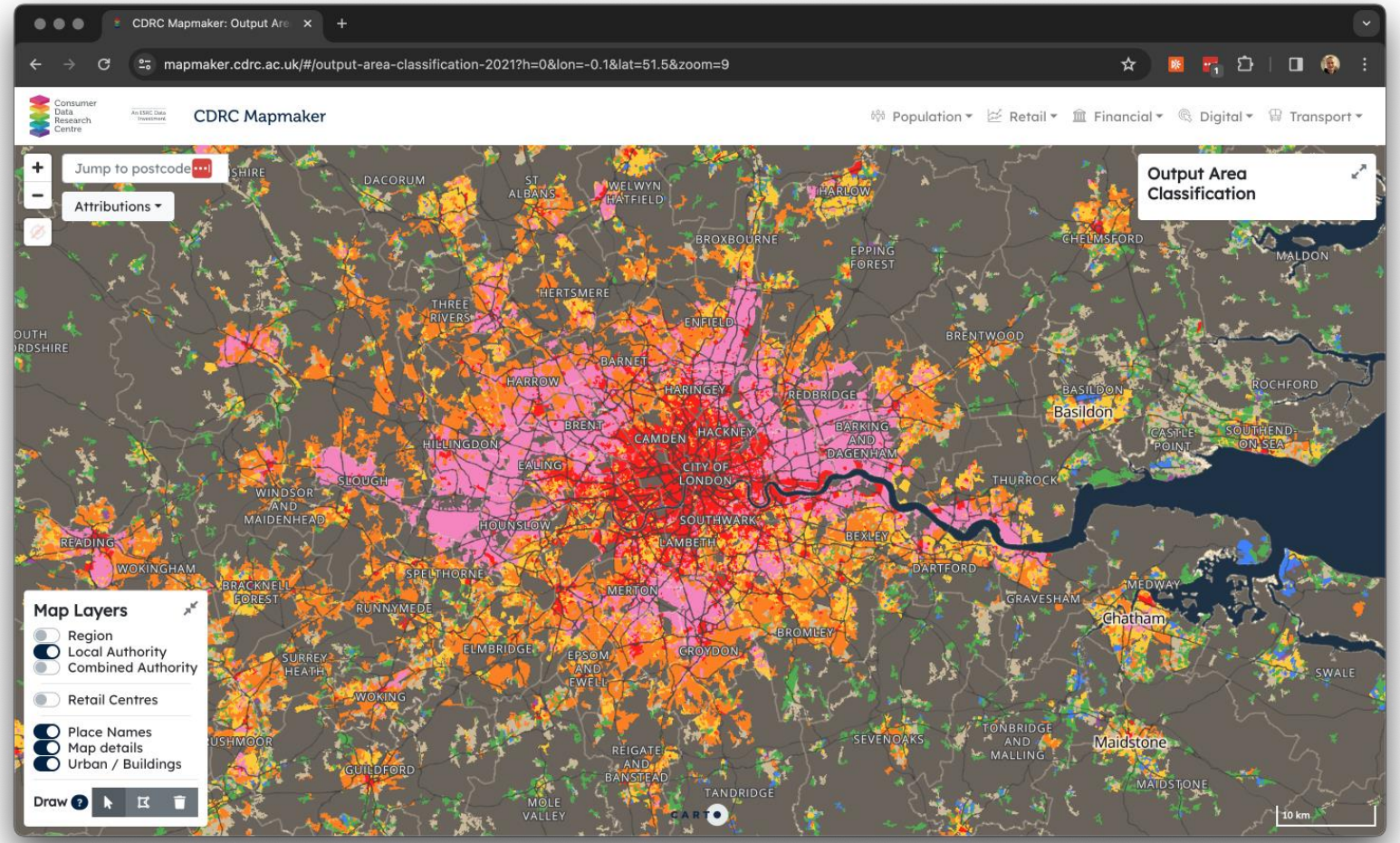
Spatial analysis



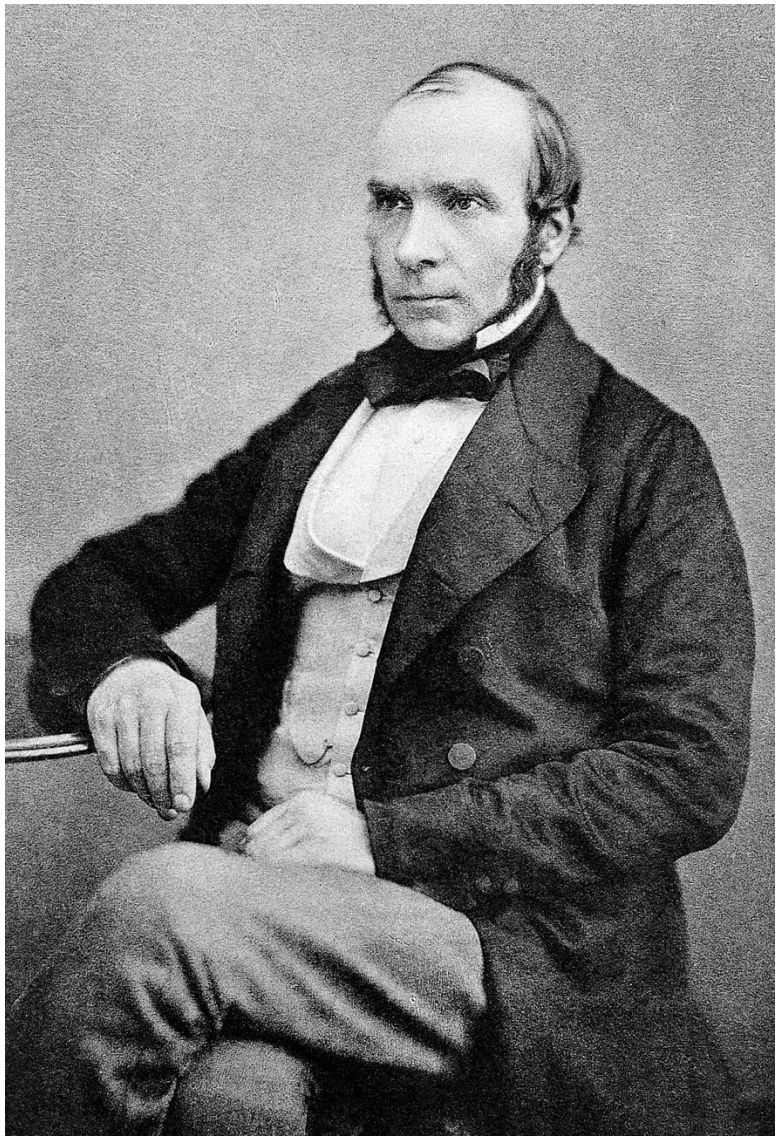
Spatial analysis



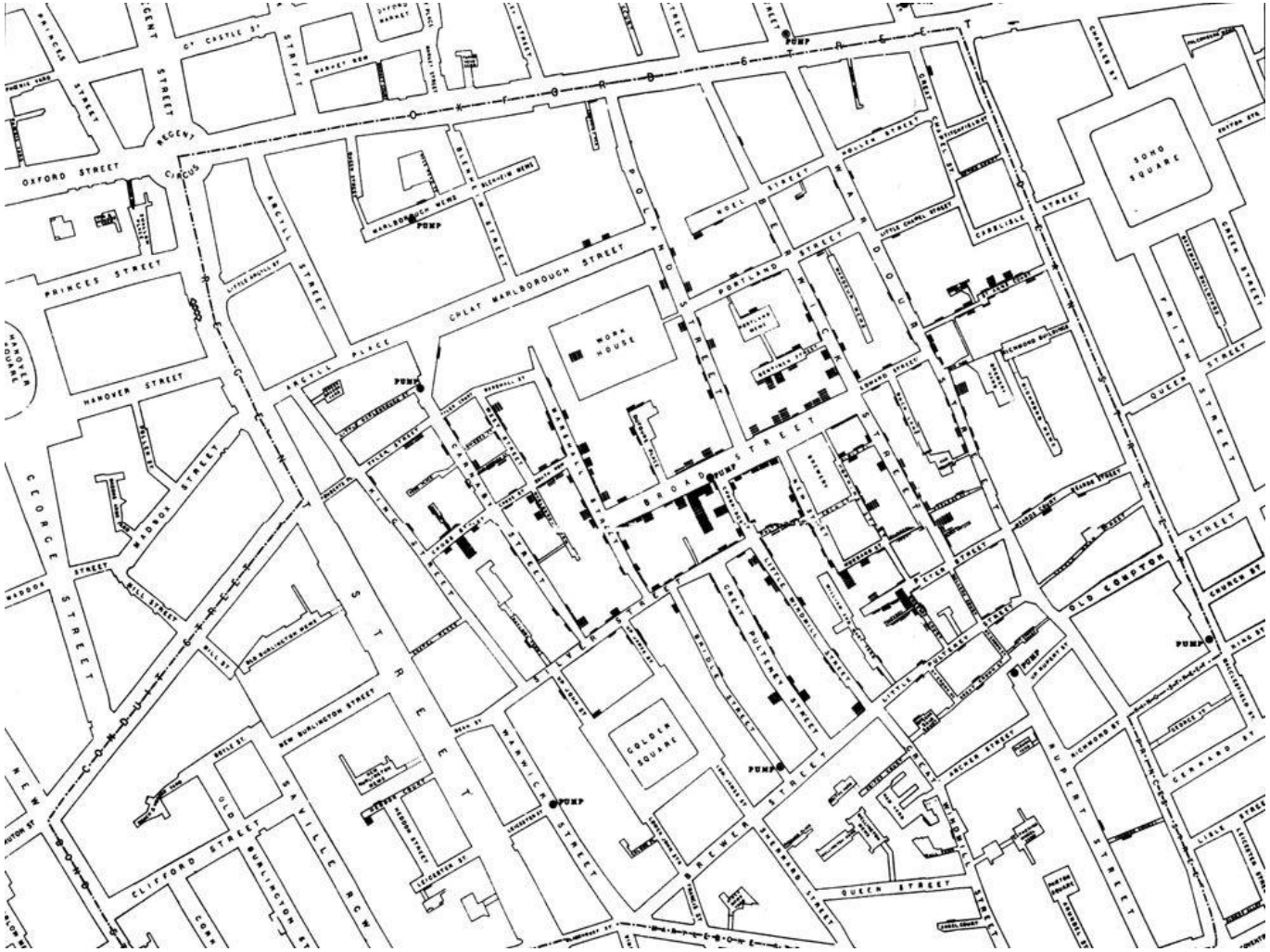
Spatial analysis



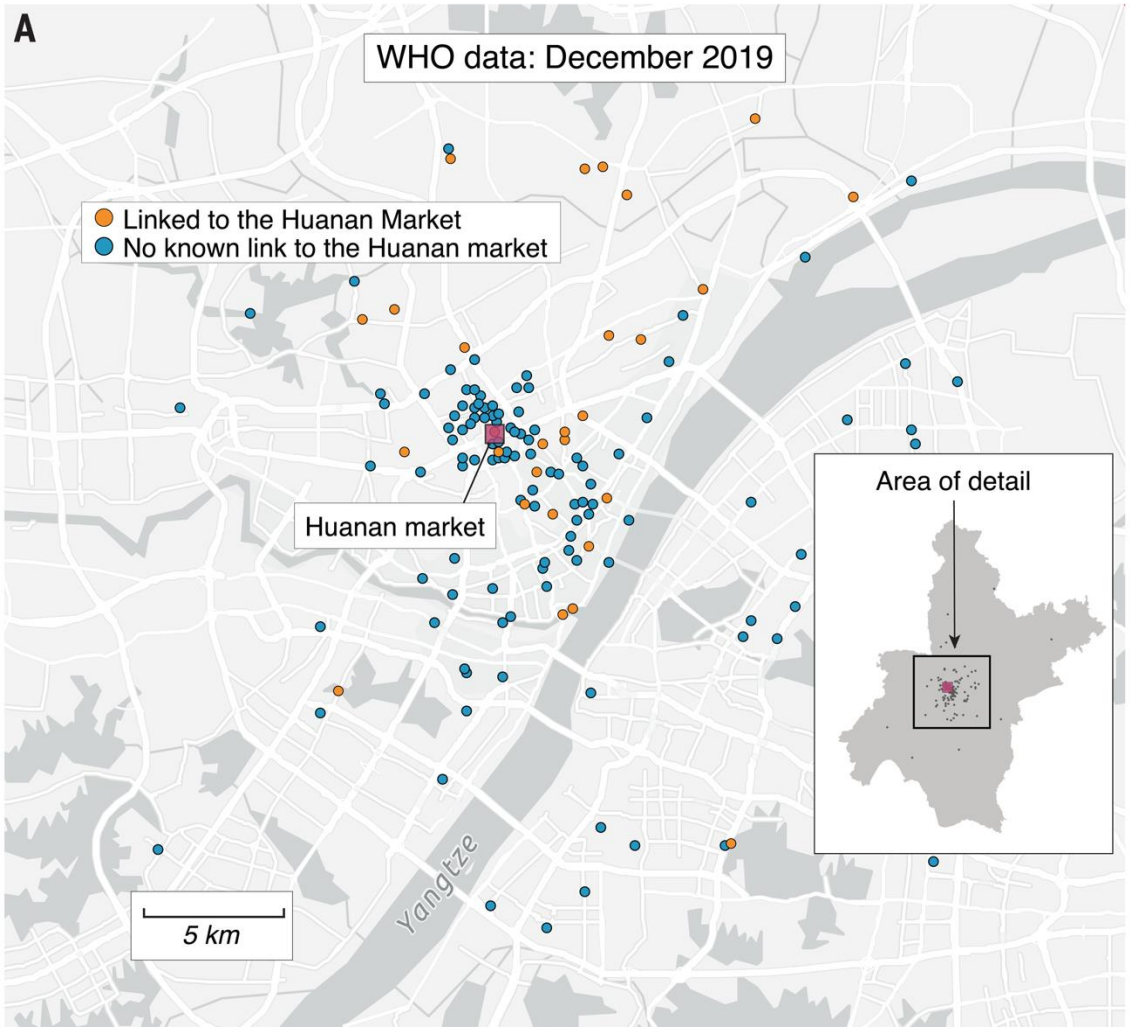
Spatial analysis



Spatial analysis



Spatial analysis

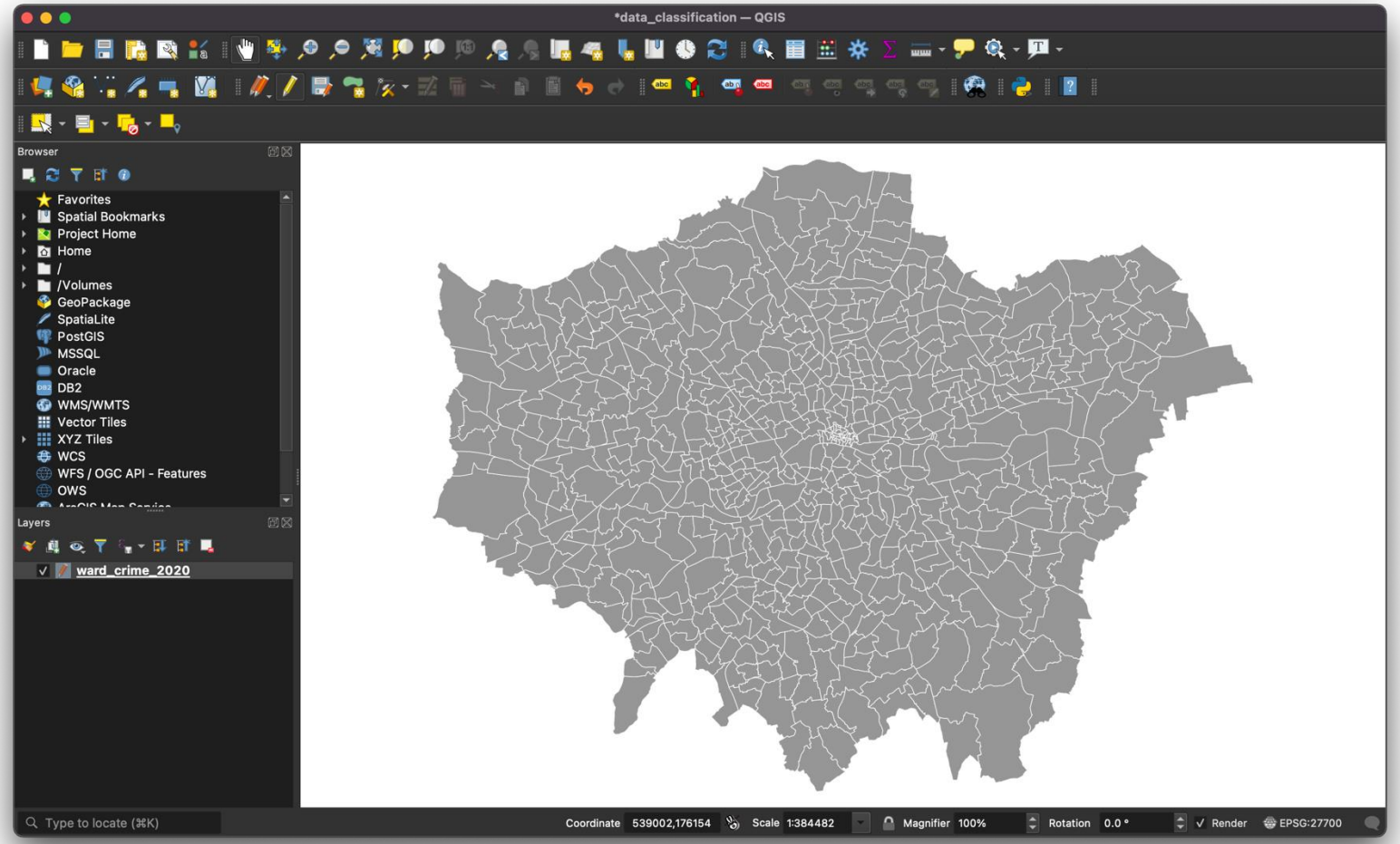


Worobey *et al*. 2022

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.

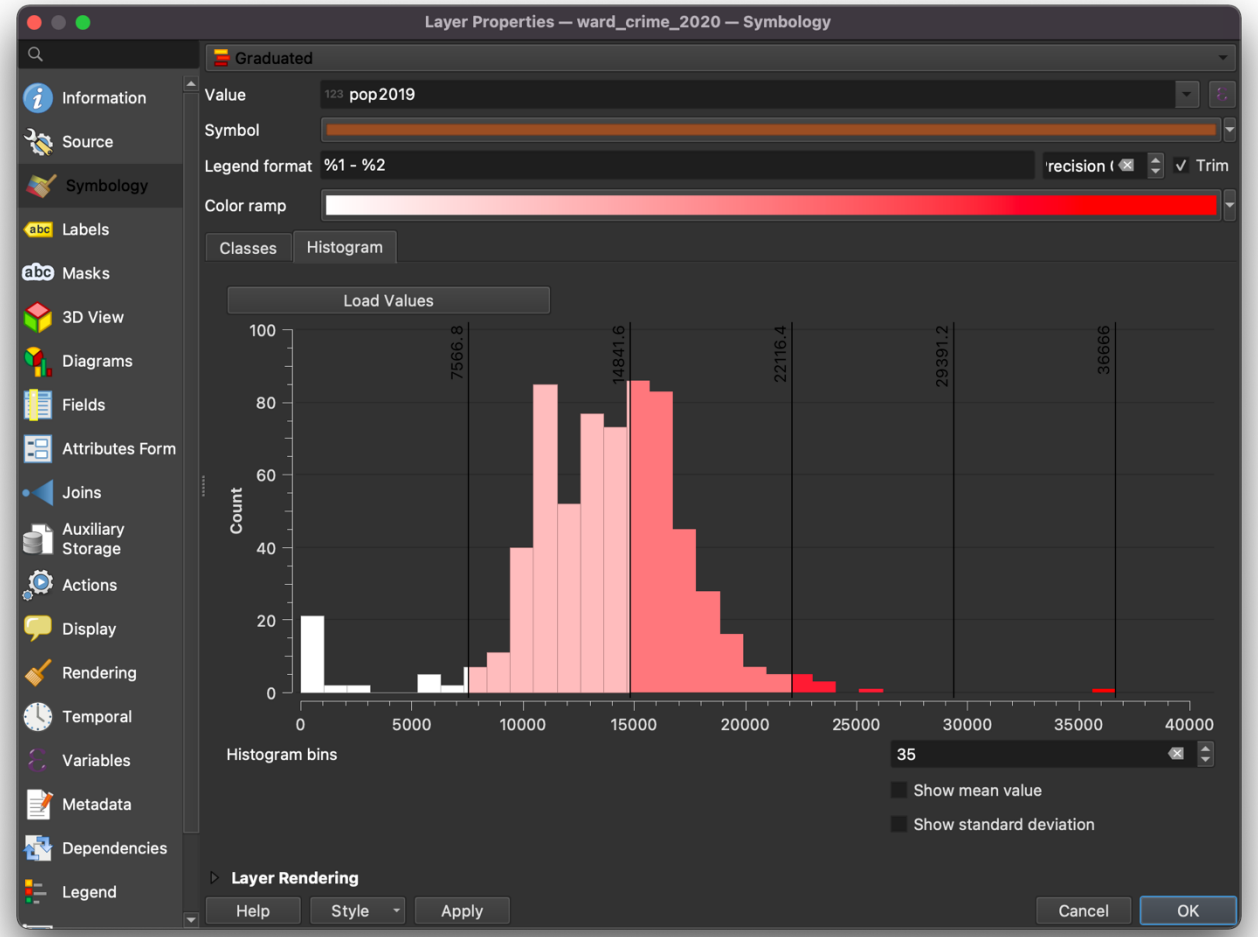
Mapping 101


*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
Type to locate (⌘K)
Coordinate 539002,176154
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

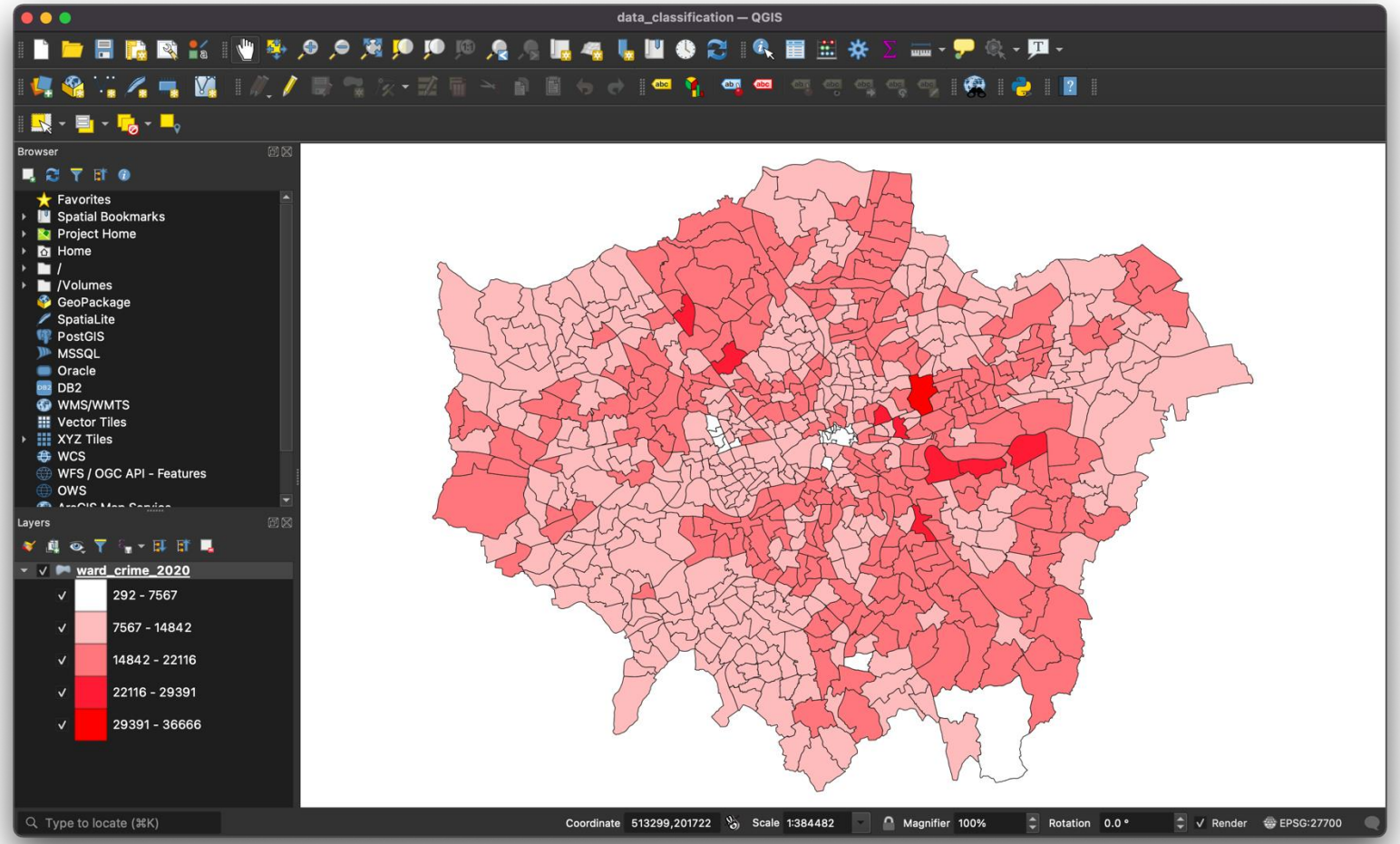
Mapping 101

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


Layer Properties — ward_crime_2020 — Symbology
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Metadata
Dependencies
Legend
Graduated
Value
pop2019
Symbol
Legend format
%1 - %2
Trim
Color ramp
Classes
Histogram
Load Values
7566.8
14841.6
22116.4
29391.2
36666
Count
Histogram bins
35
Show mean value
Show standard deviation
Layer Rendering
Help
Style
Apply
Cancel
OK

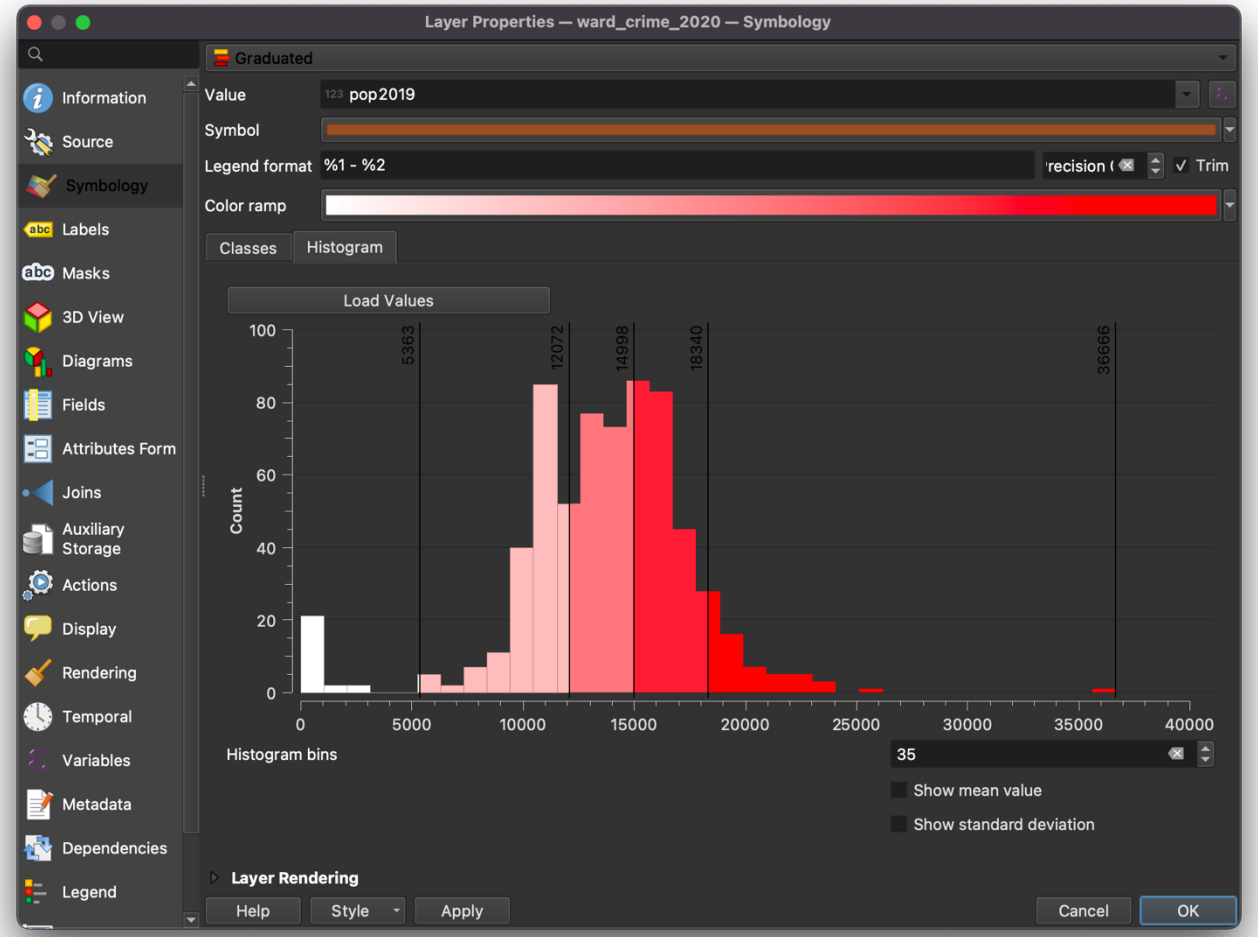
Mapping 101


data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
292 - 7567
7567 - 14842
14842 - 22116
22116 - 29391
29391 - 36666
Type to locate (⌘K)
Coordinate 513299,201722
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

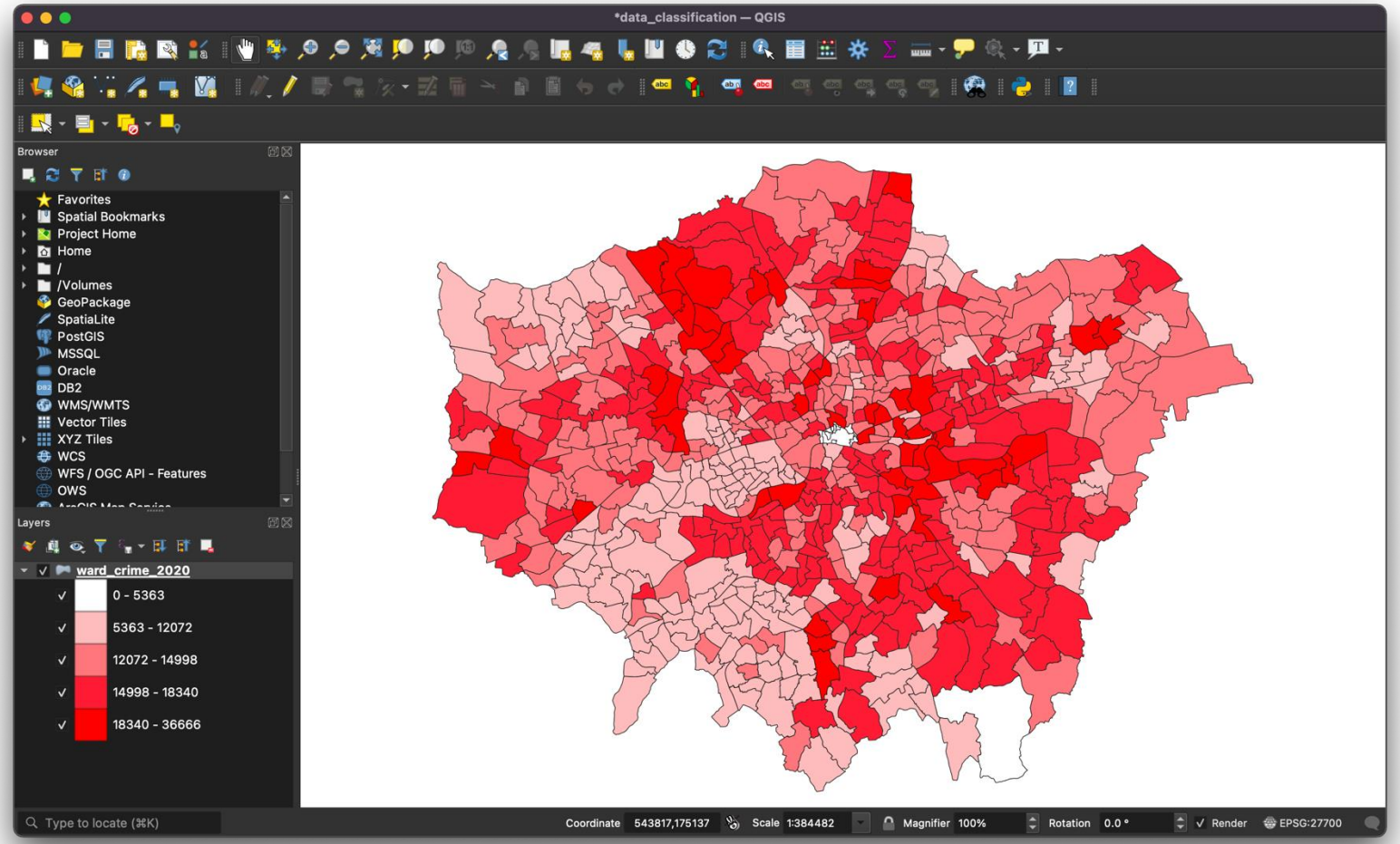
Mapping 101

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


Layer Properties — ward_crime_2020 — Symbology
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Metadata
Dependencies
Legend
Graduated
Value
pop2019
Symbol
Legend format
%1 - %2
Trim
Color ramp
Classes
Histogram
Load Values
5363
12072
14998
18340
36666
Count
Histogram bins
35
Show mean value
Show standard deviation
Layer Rendering
Help
Style
Apply
Cancel
OK

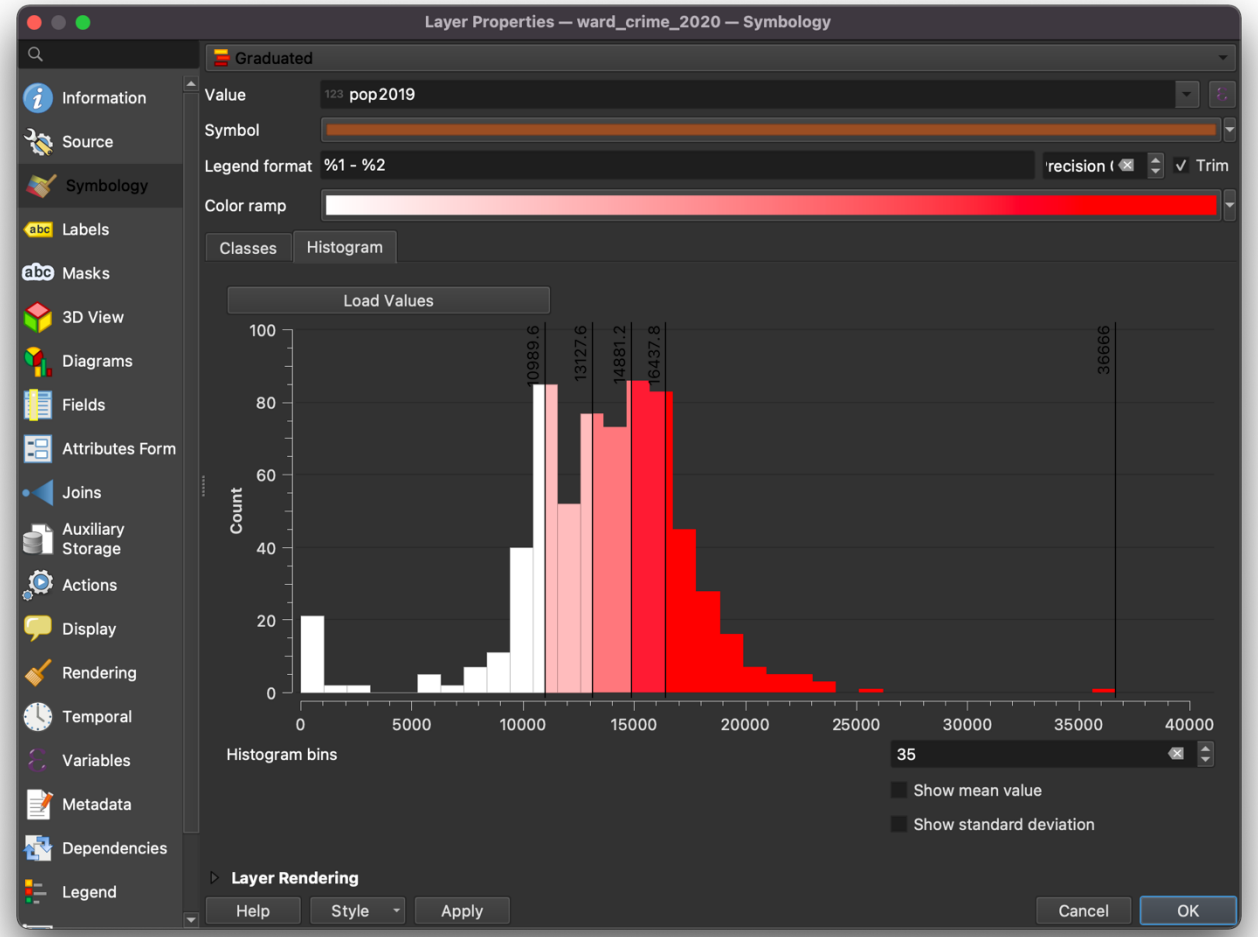
Mapping 101


*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
0 - 5363
5363 - 12072
12072 - 14998
14998 - 18340
18340 - 36666
Type to locate (⌘K)
Coordinate 543817,175137
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

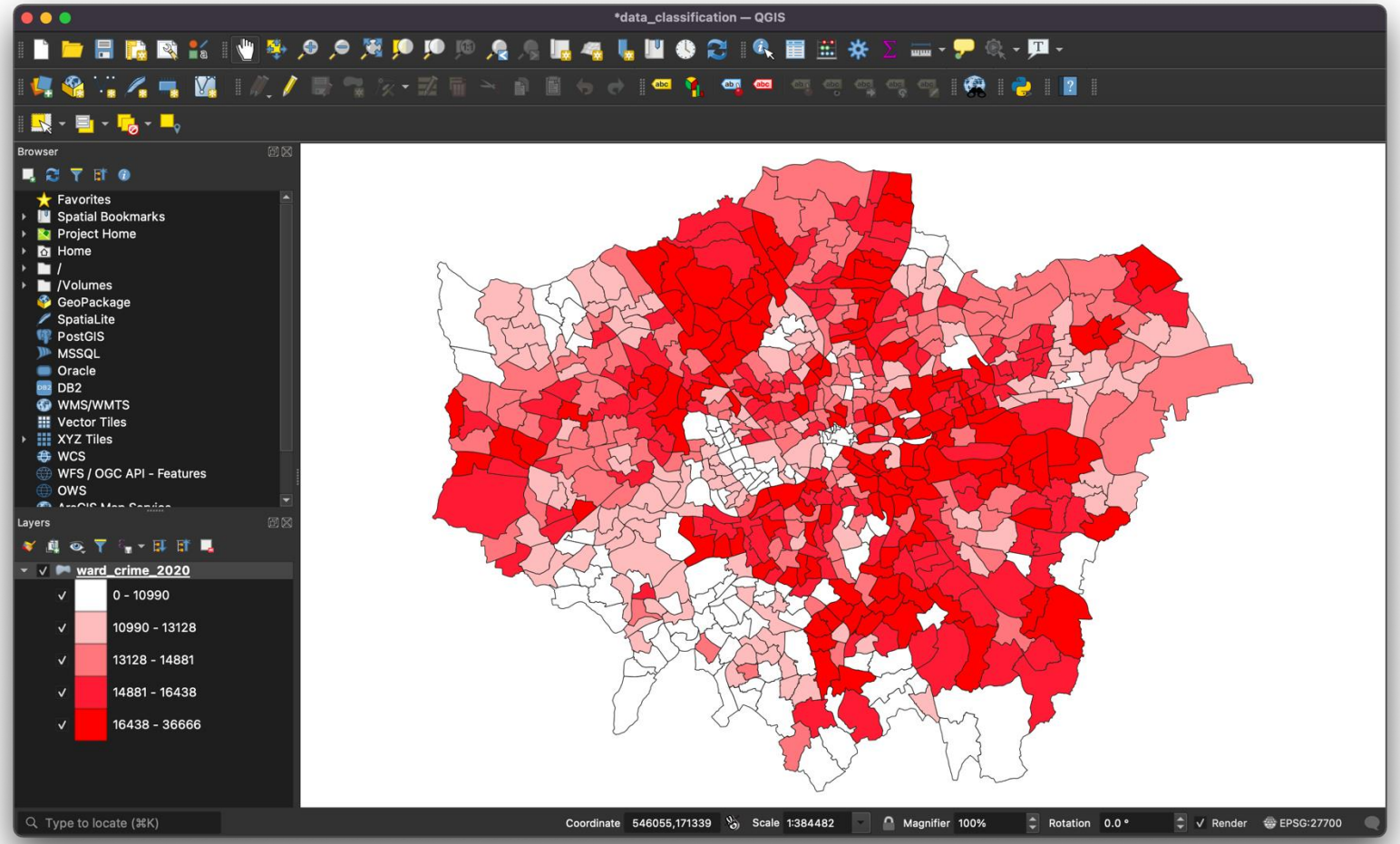
Mapping 101

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


Layer Properties — ward_crime_2020 — Symbology
Graduated
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Metadata
Dependencies
Legend
Value
pop2019
Symbol
Legend format
%1 - %2
Trim
Color ramp
Classes
Histogram
Load Values
Count
100
80
60
40
20
0
0
5000
10000
15000
20000
25000
30000
35000
40000
Histogram bins
35
Show mean value
Show standard deviation
Layer Rendering
Help
Style
Apply
Cancel
OK

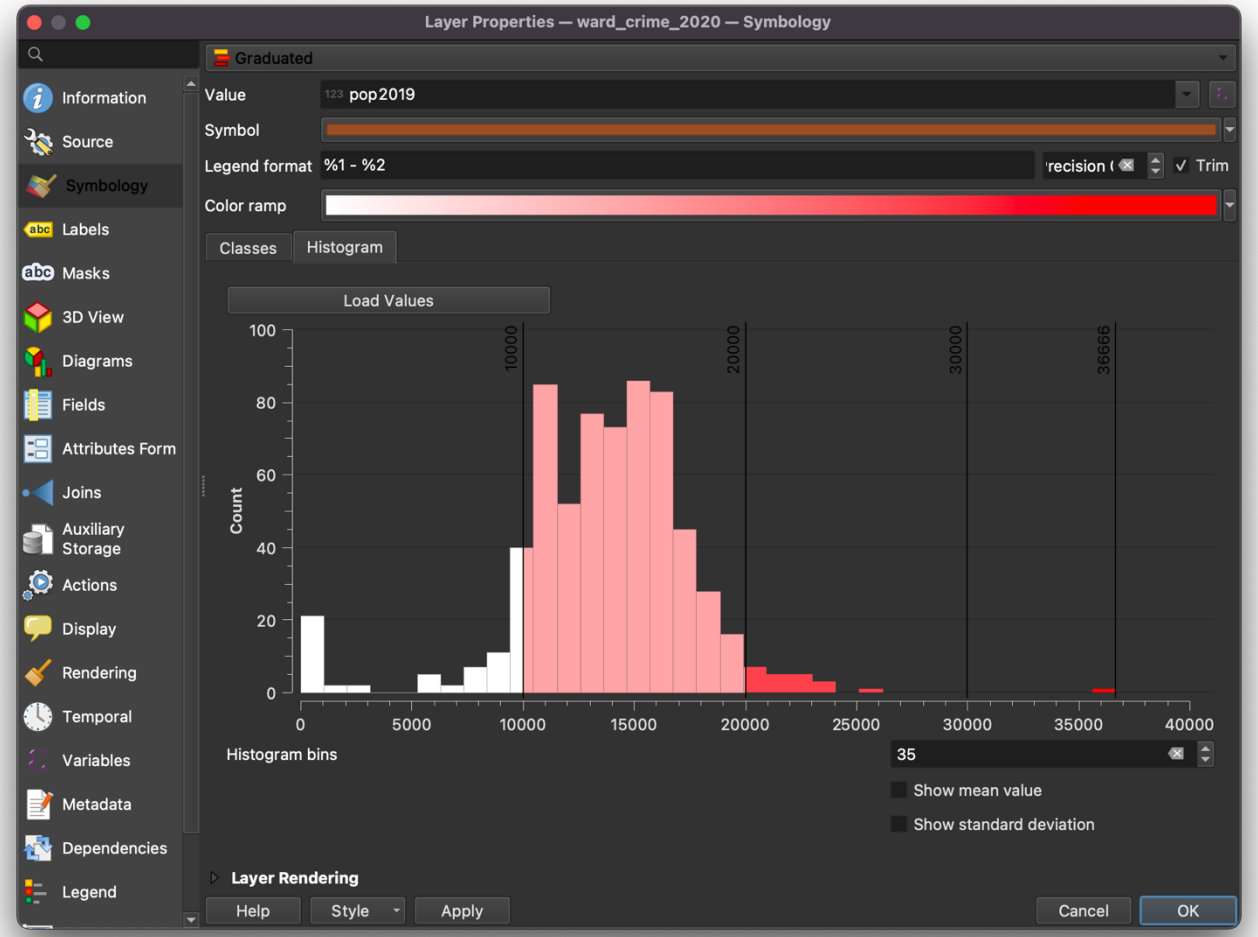
Mapping 101


*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
0 - 10990
10990 - 13128
13128 - 14881
14881 - 16438
16438 - 36666
Type to locate (⌘K)
Coordinate 546055,171339
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

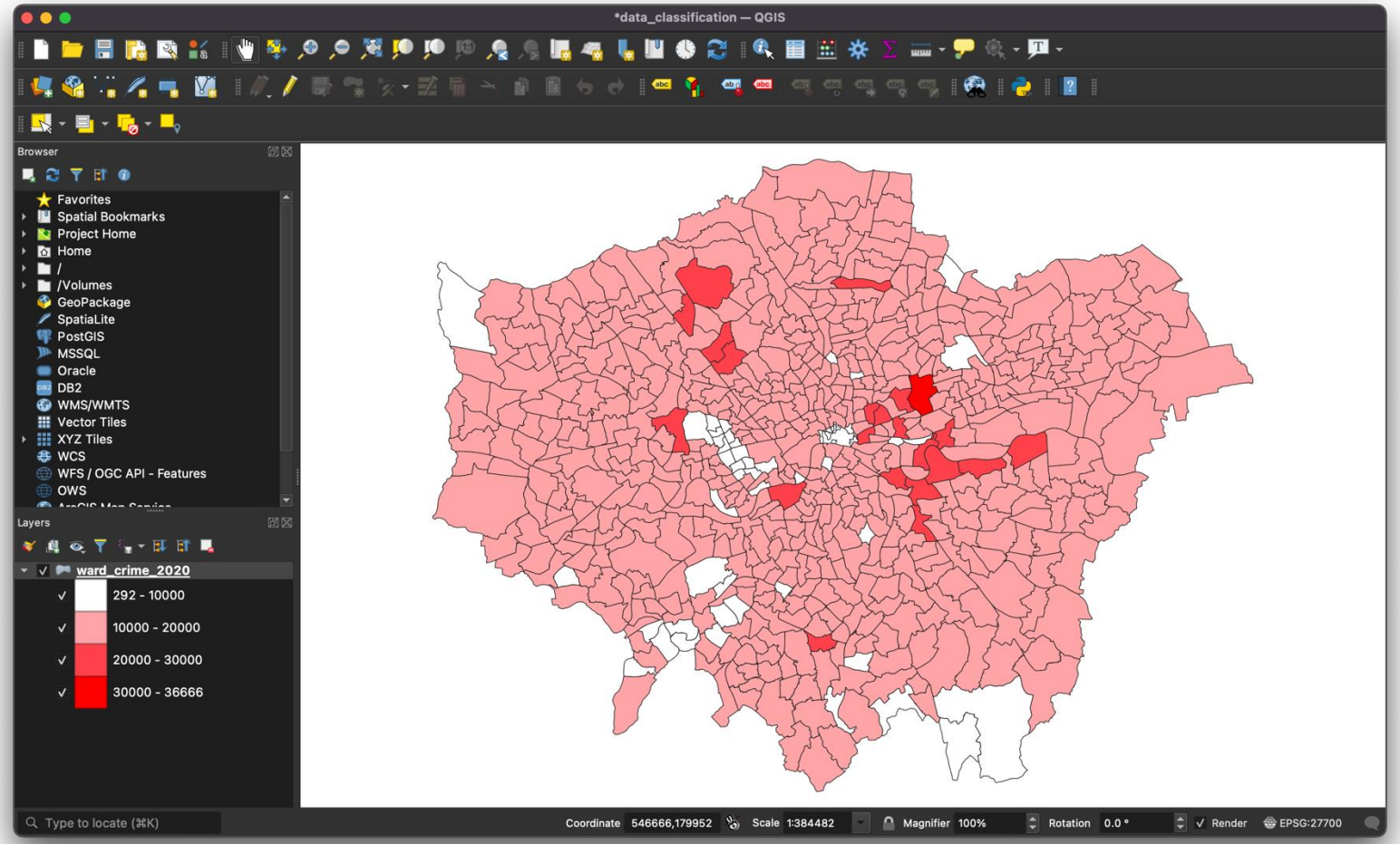
Mapping 101

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.

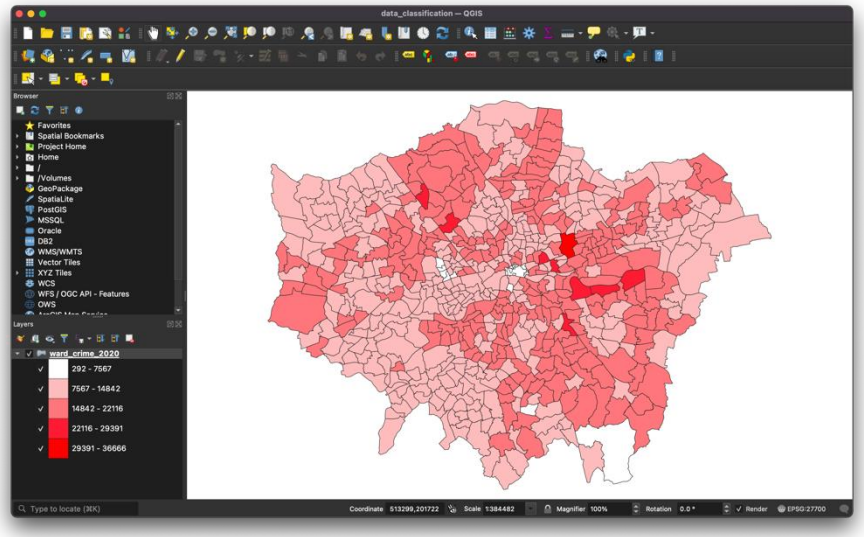
Mapping 101

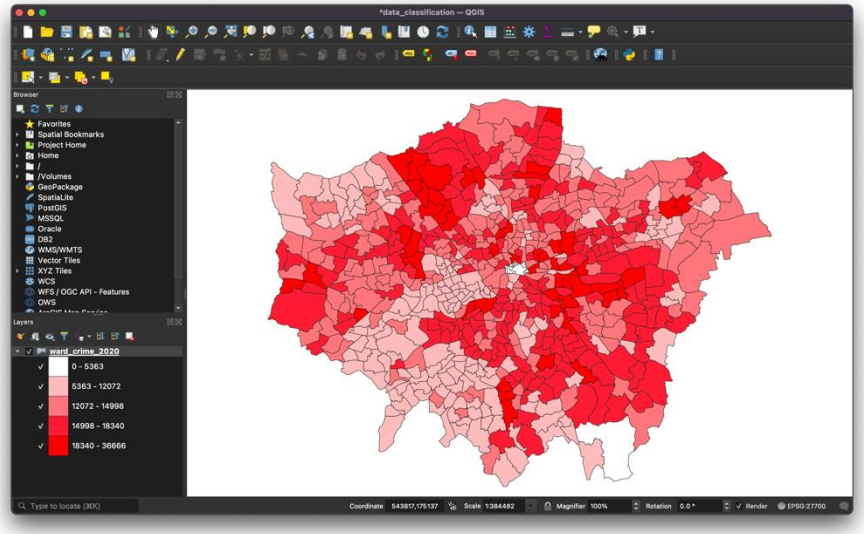

Layer Properties — ward_crime_2020 — Symbology
Graduated
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Metadata
Dependencies
Legend
Value
pop2019
Symbol
Legend format
%1 - %2
Trim
Color ramp
Classes
Histogram
Load Values
Count
Histogram bins
35
Show mean value
Show standard deviation
Layer Rendering
Help
Style
Apply
Cancel
OK

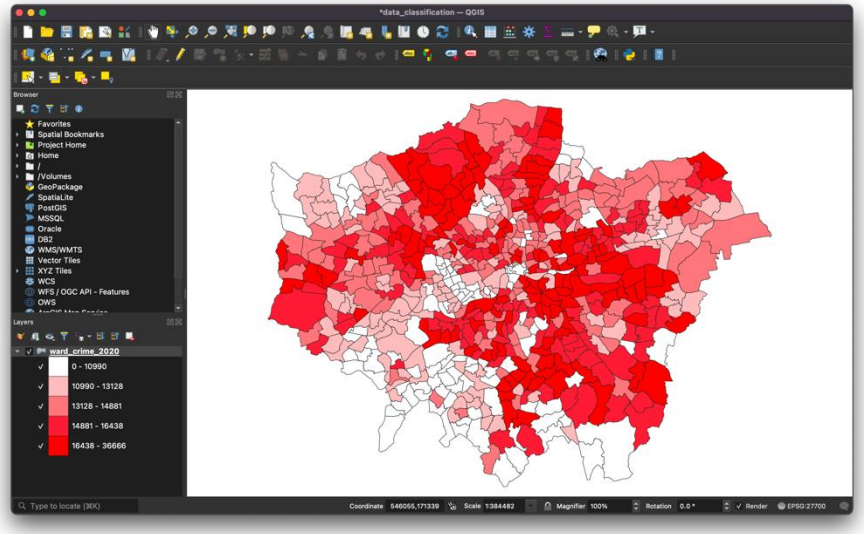
Mapping 101

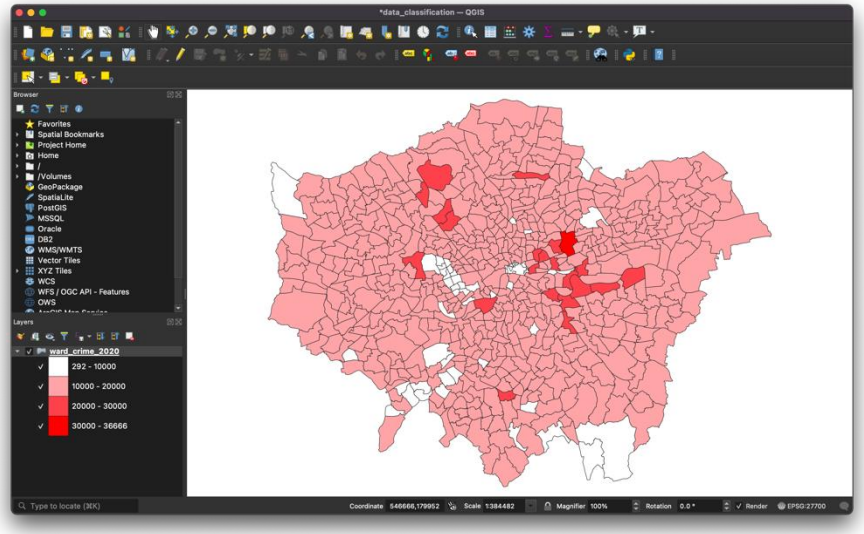

*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
292 - 10000
10000 - 20000
20000 - 30000
30000 - 36666
Type to locate (⌘K)
Coordinate 546666,179952
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

Different classifications, different maps





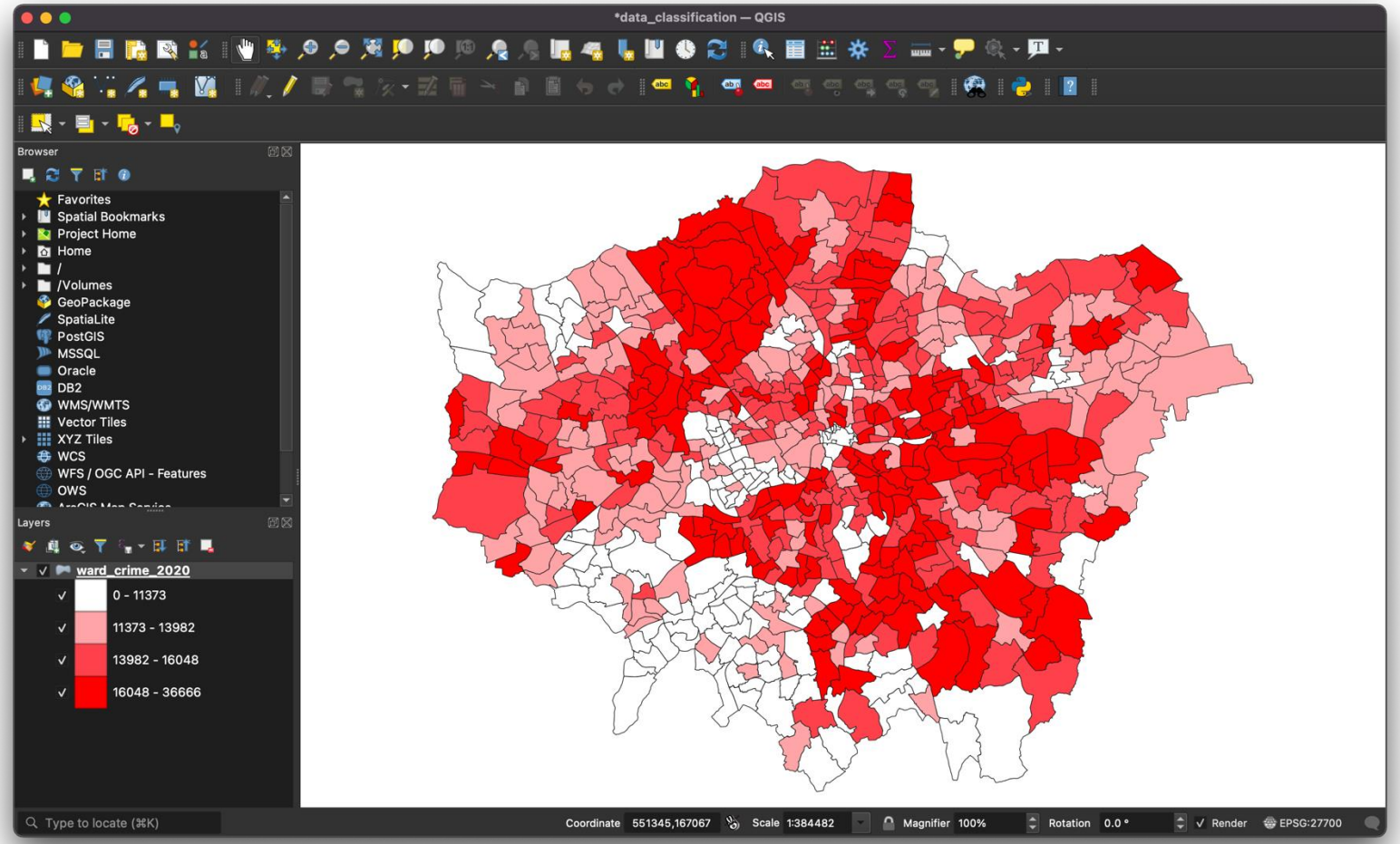




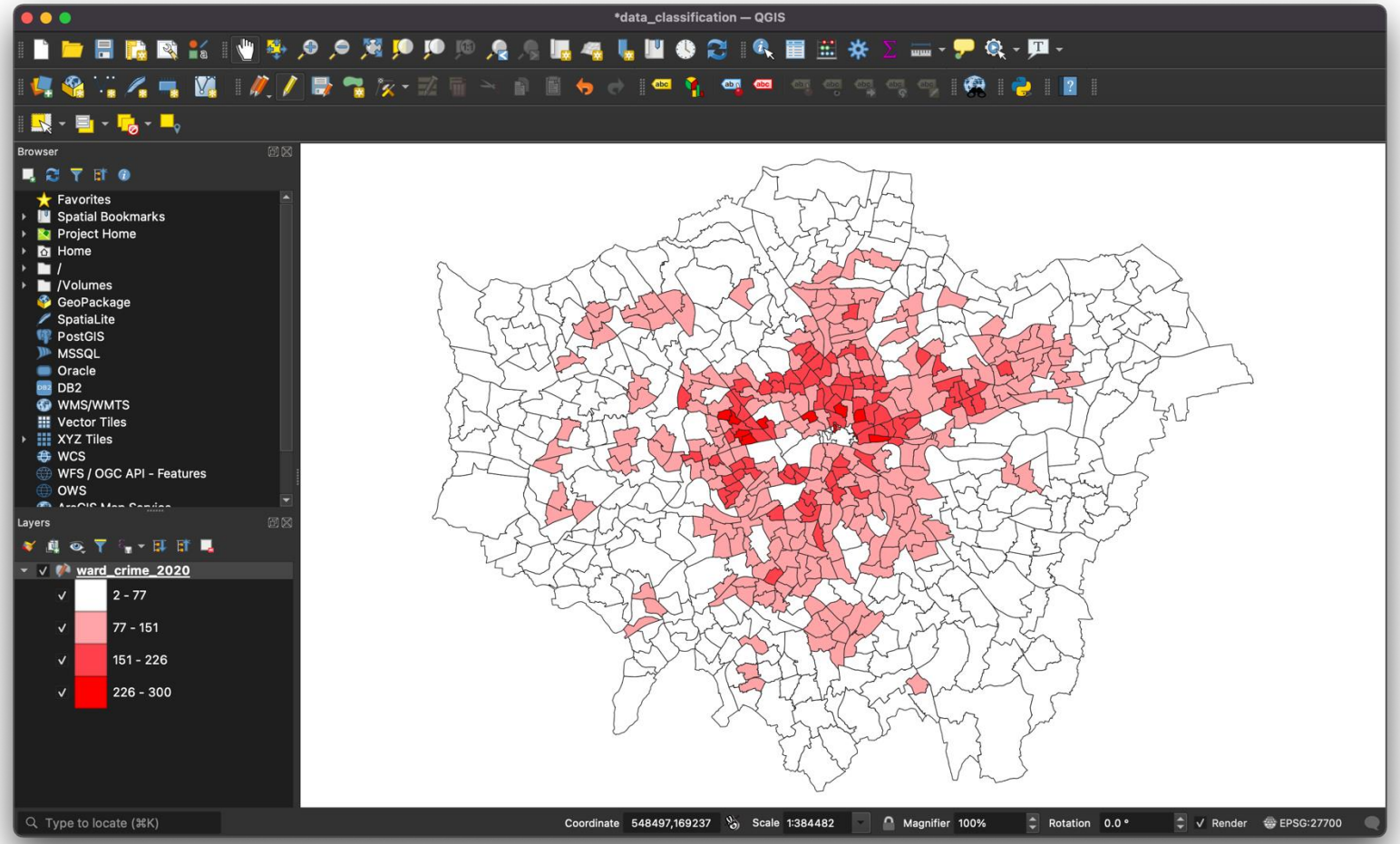
Mapping 101

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


*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
0 - 11373
11373 - 13982
13982 - 16048
16048 - 36666
Type to locate (⌘K)
Coordinate 551345,167067
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

Mapping 101


*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
2 - 77
77 - 151
151 - 226
226 - 300
Type to locate (⌘K)
Coordinate 548497,169237
Scale 1:384482
Magnifier 100%
Rotation 0.0 °
Render
EPSG:27700

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