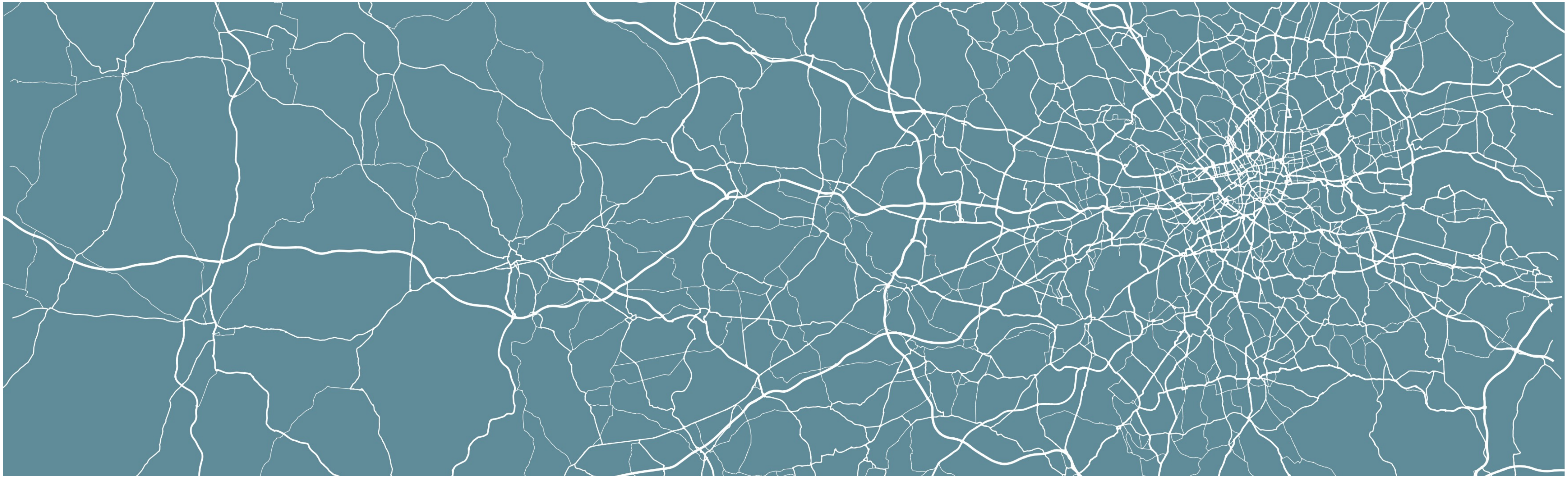
Geocomputation

Cartography and Visualisation

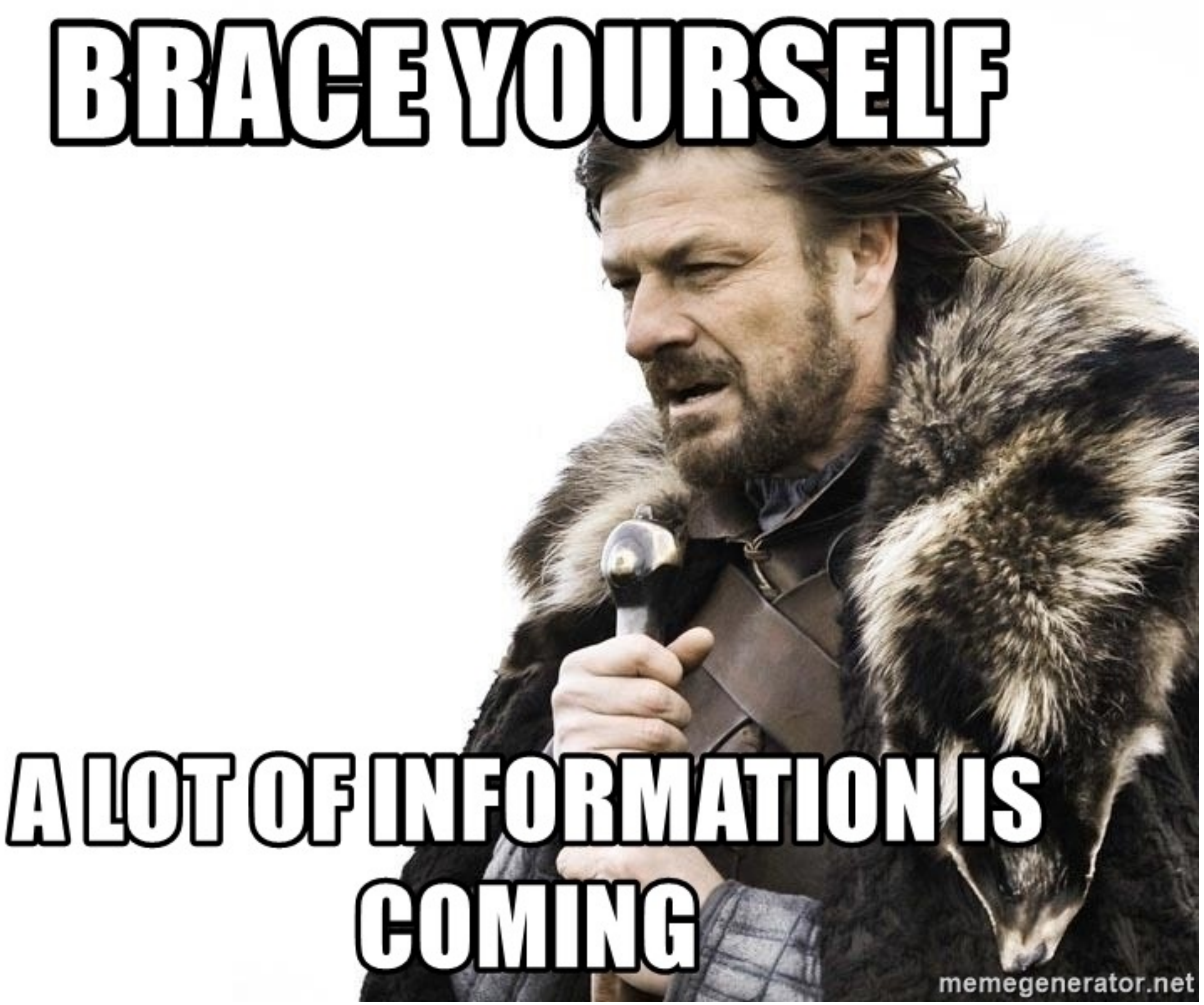


Where are we at?

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


BRACE YOURSELF
A LOT OF INFORMATION IS COMING
memegenerator.net

This week

Part I: Making maps

- Types of maps, specifically the choropleth map
- A note on ecological fallacy and MAUP
- Map conventions

Part II: Positioning the map

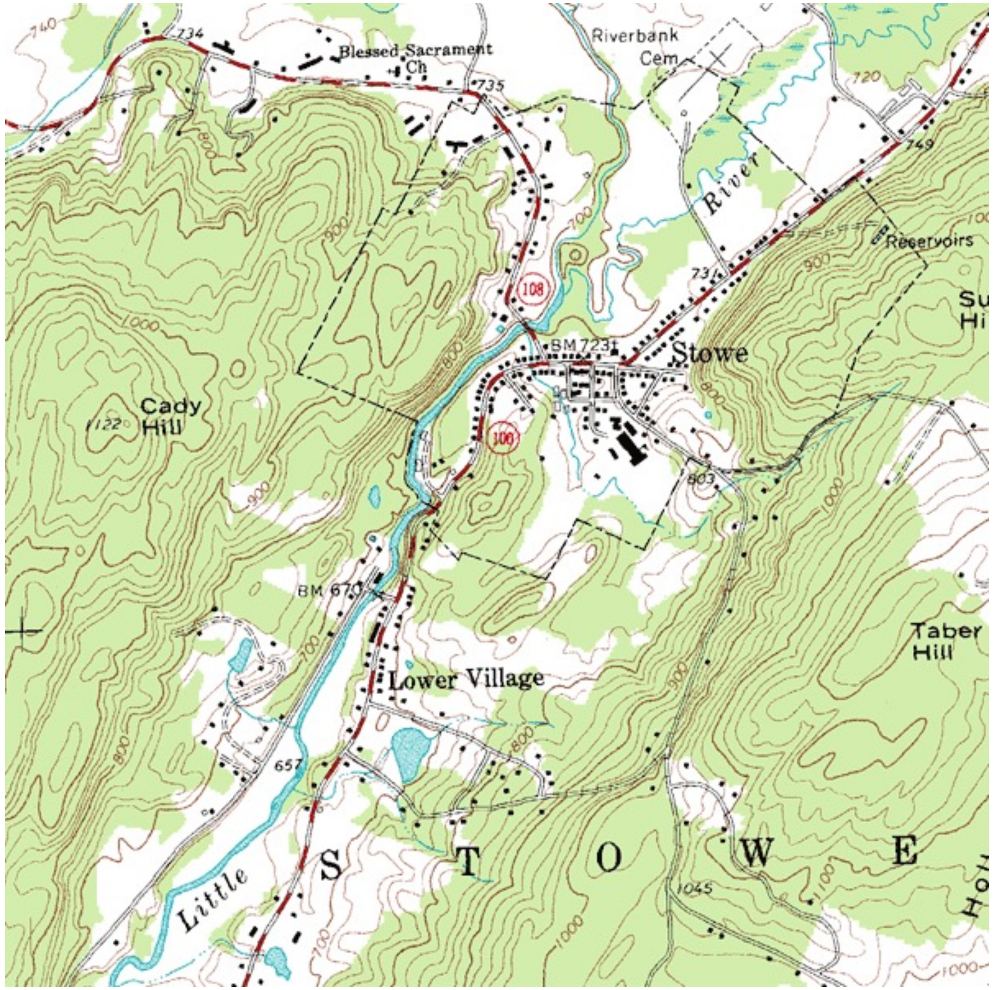
- Coordinate Reference Systems and Projections

Part I: Making maps

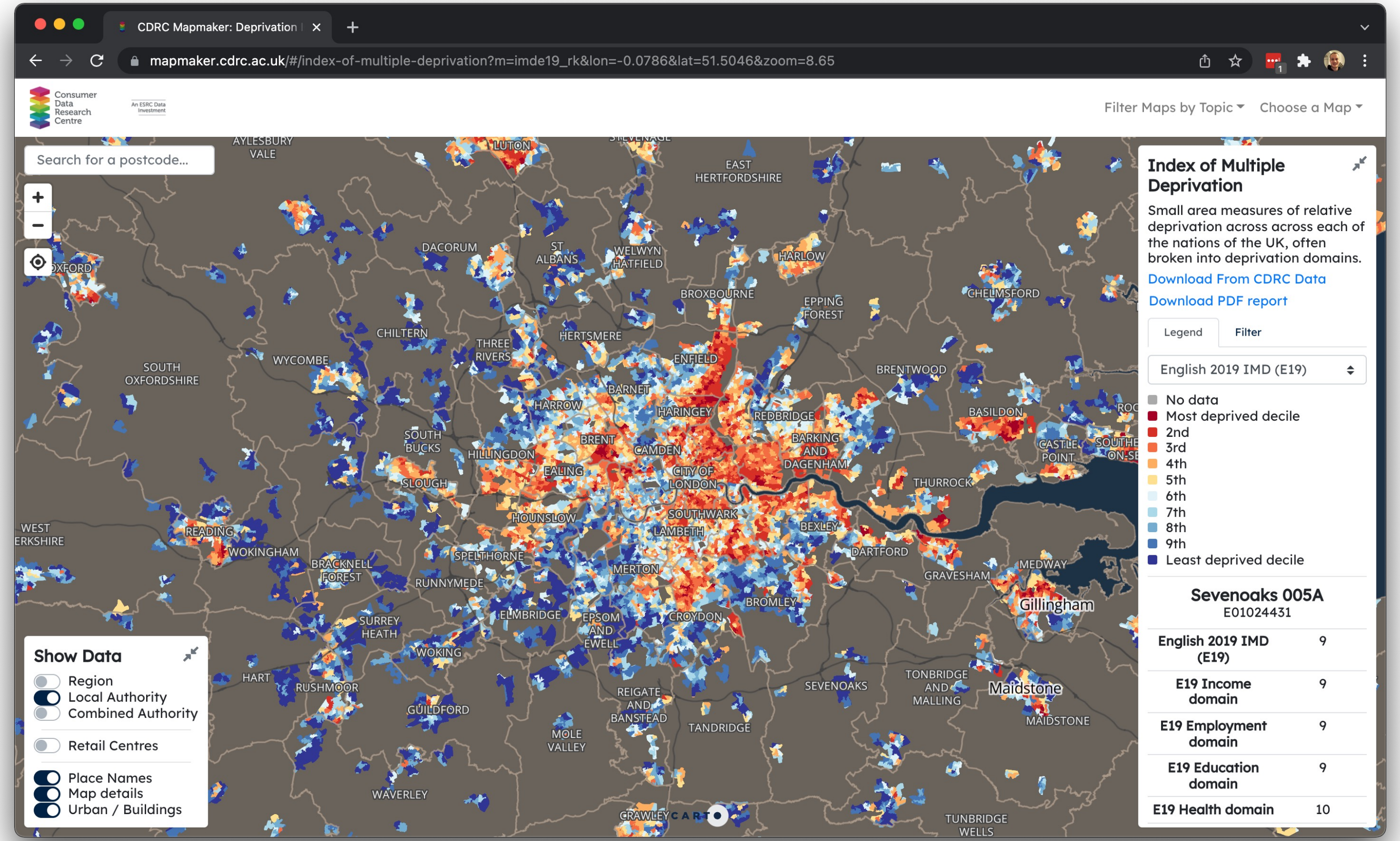
Types of maps

- Topographic maps: indicate features of the land’s surface (e.g. mountains, hills, valleys, rivers).
- Thematic maps: graphical outputs that typically show geographic patterns of a particular theme in a geographic area.

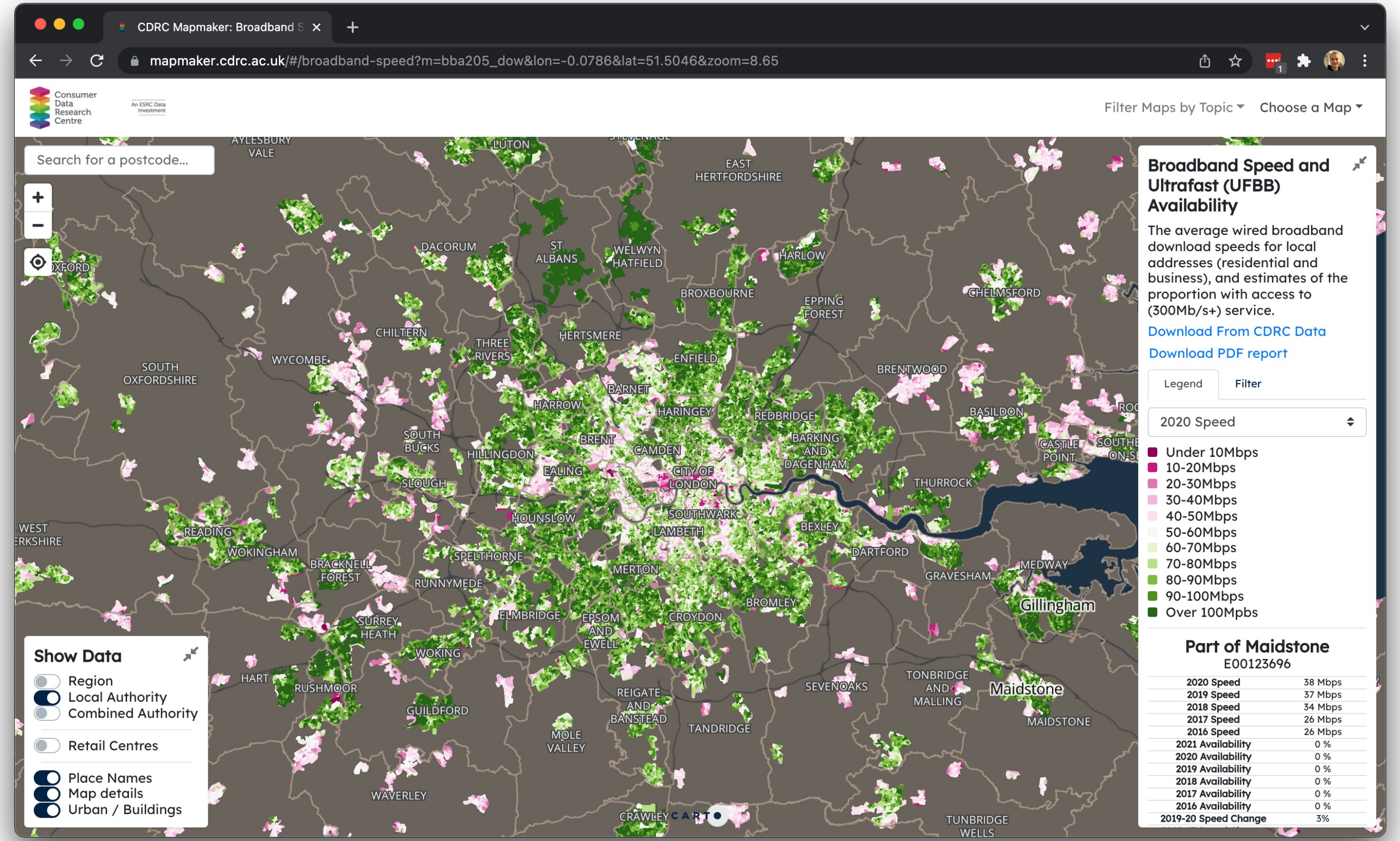
Topographic maps



Thematic maps



Thematic maps



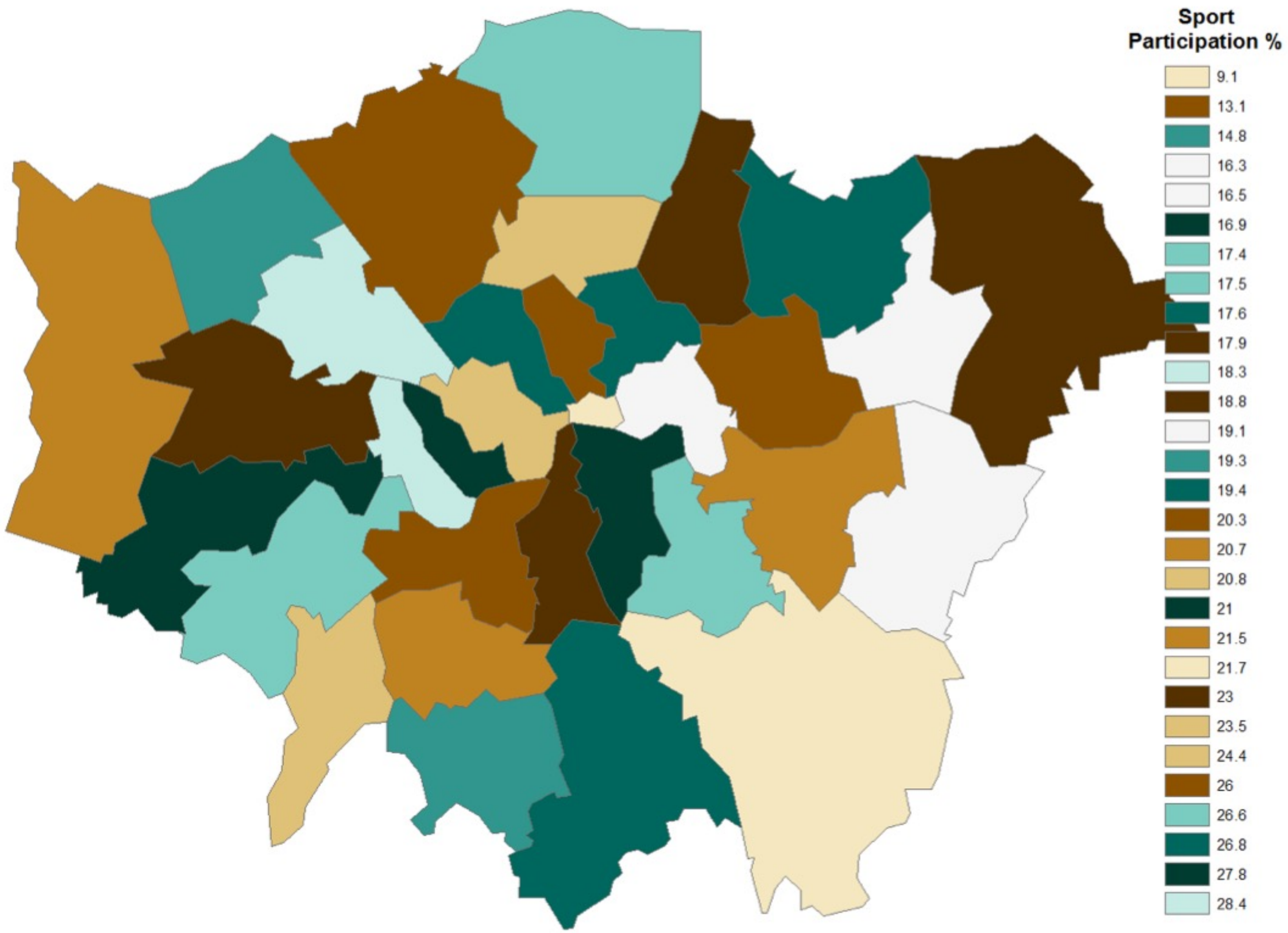
Choropleth maps

- A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map.
- Pretty much the most common style of map you will encounter or produce.
- Shows two types of data: categorical or continuous

Mapping categorical data

- ‘Individual’ colours for unique values in the data, no gradient!
- Descriptive legends to help interpret colours.
- Colours can be ordered to infer rank for ordinal data.

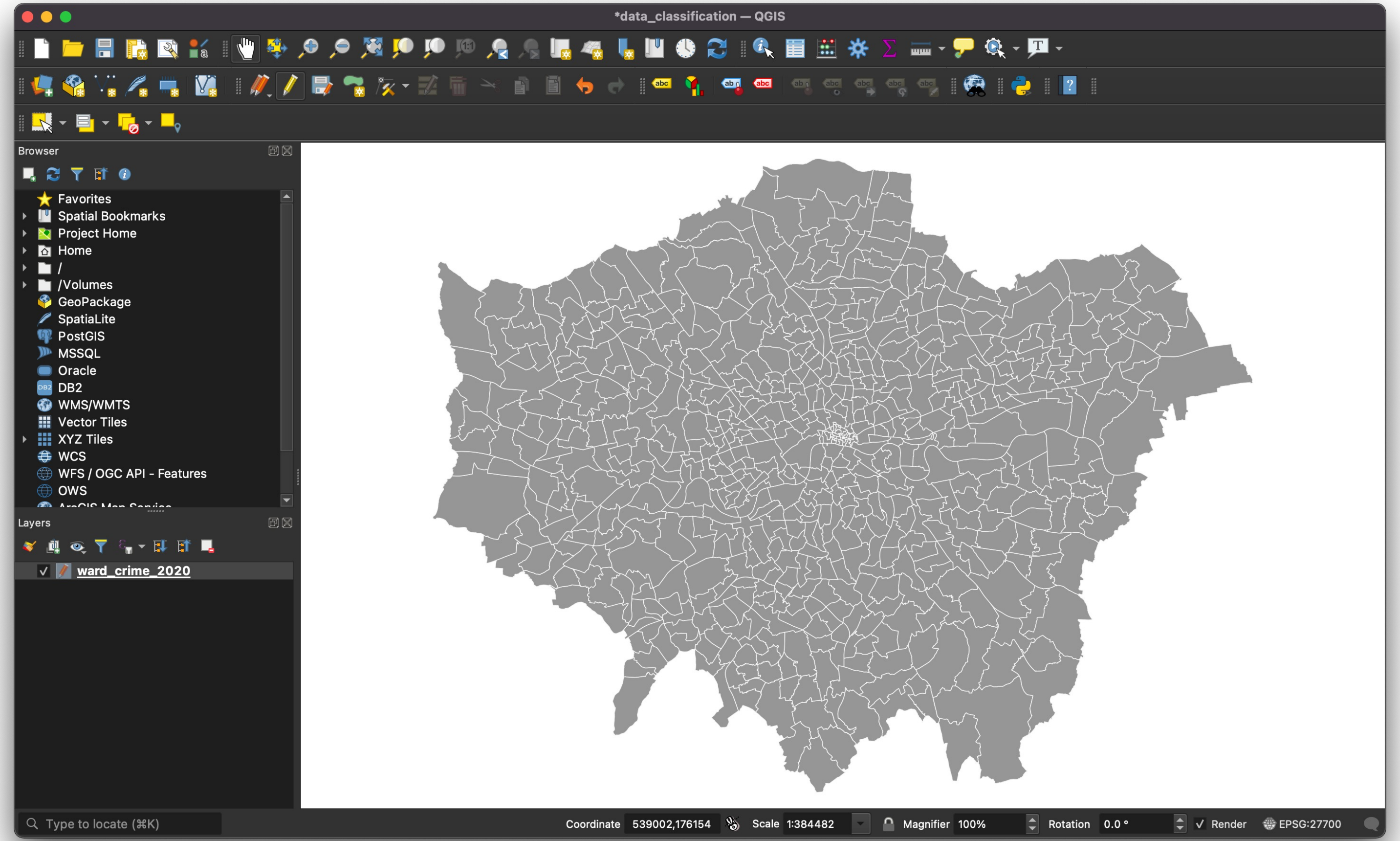
Mapping categorical data



Mapping continuous data

- Continuous values divided into classes – spatial equivalent of a histogram.
- Number of classes dependent on data (number of bins).
- Utilise data distribution (via histogram) to determine best classification approach (how to bin) and breaks.
- Approach to classification will be determined by distribution of data (how to bin?).
- Always a trade-off between information loss and readability / simplicity.

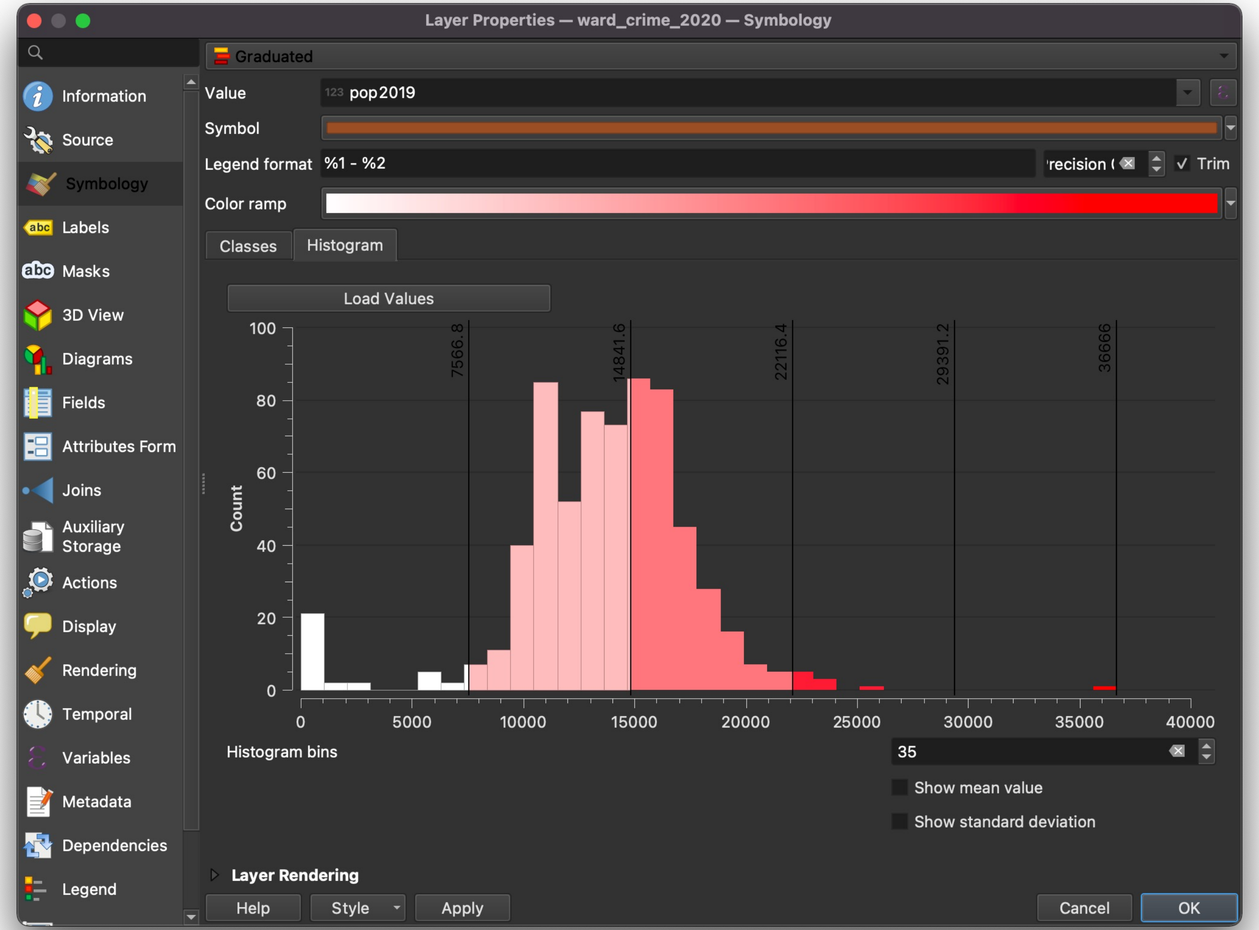
Mapping continuous data



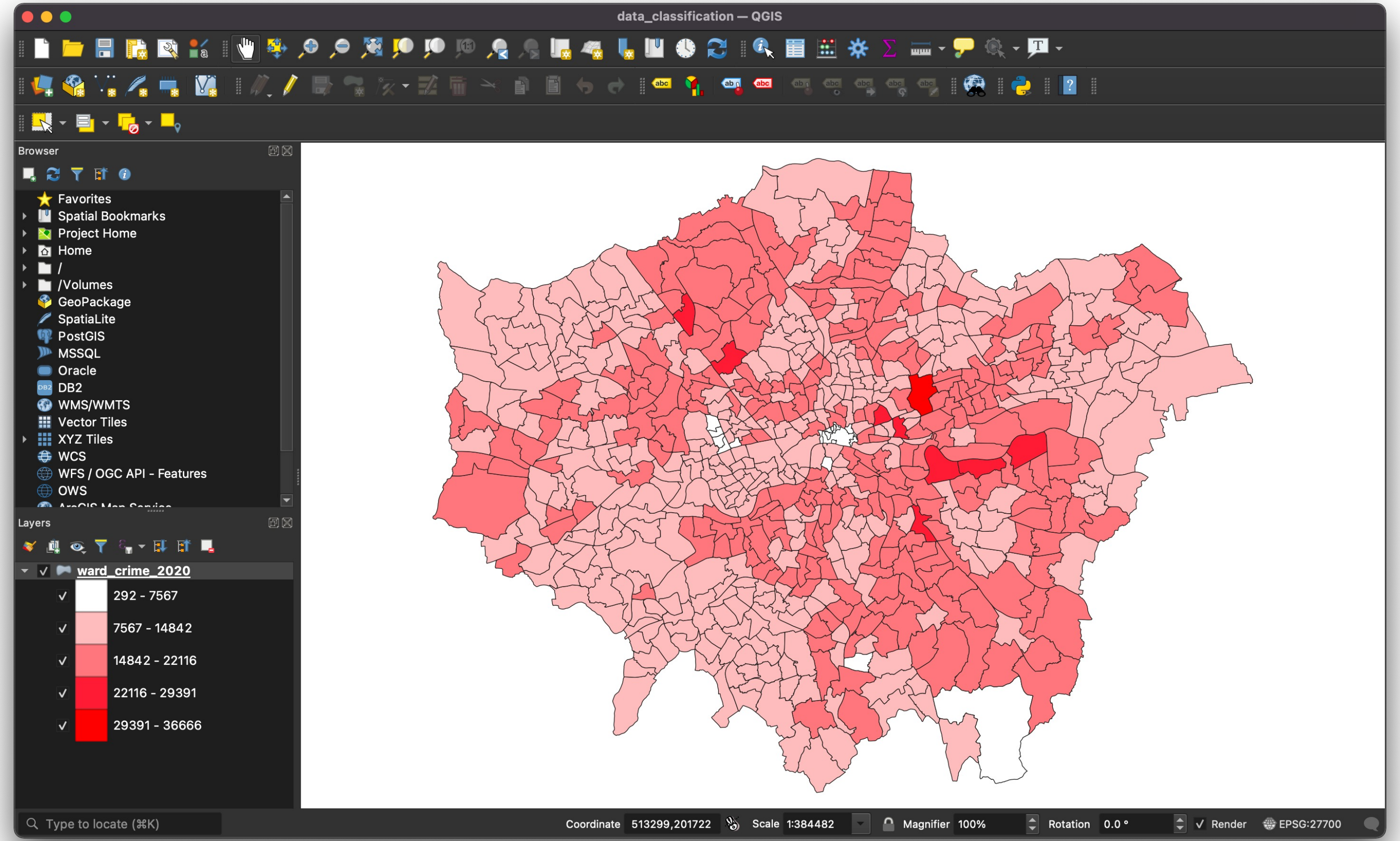
Equal interval

- Take the range of the data values to represent and split it into equal intervals.
- Splits are based on the numerical value.
- If distribution is skewed: more weight to outliers.

Equal interval



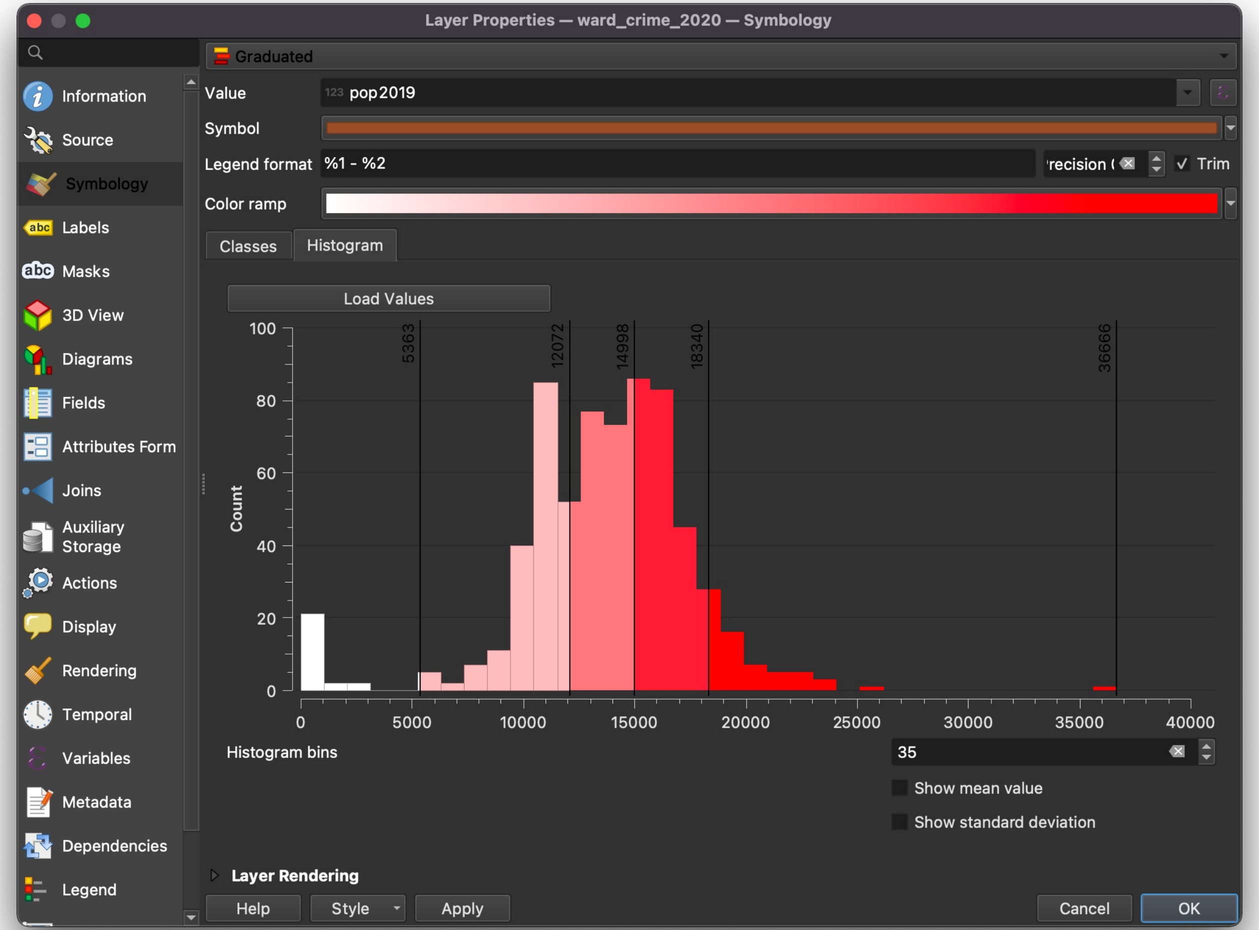
Equal interval



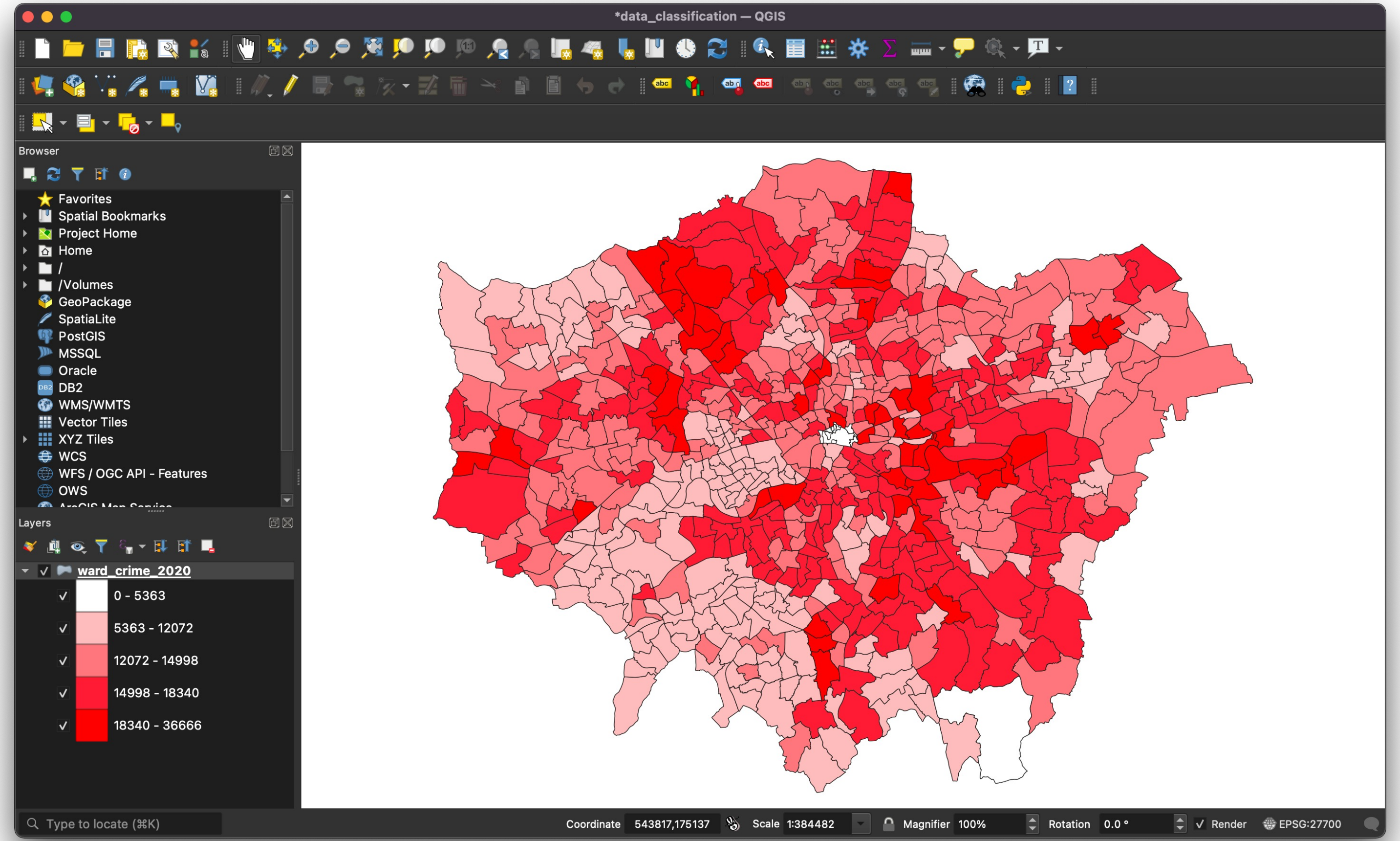
Natural breaks

- Based on minimisation of each class’s average deviation from the class mean and maximisation of each class's deviation from the means of the other classes.
- Splits are based on data clustering.
- If distribution is skewed: outliers in own class (although dependent on number of classes).

Natural breaks



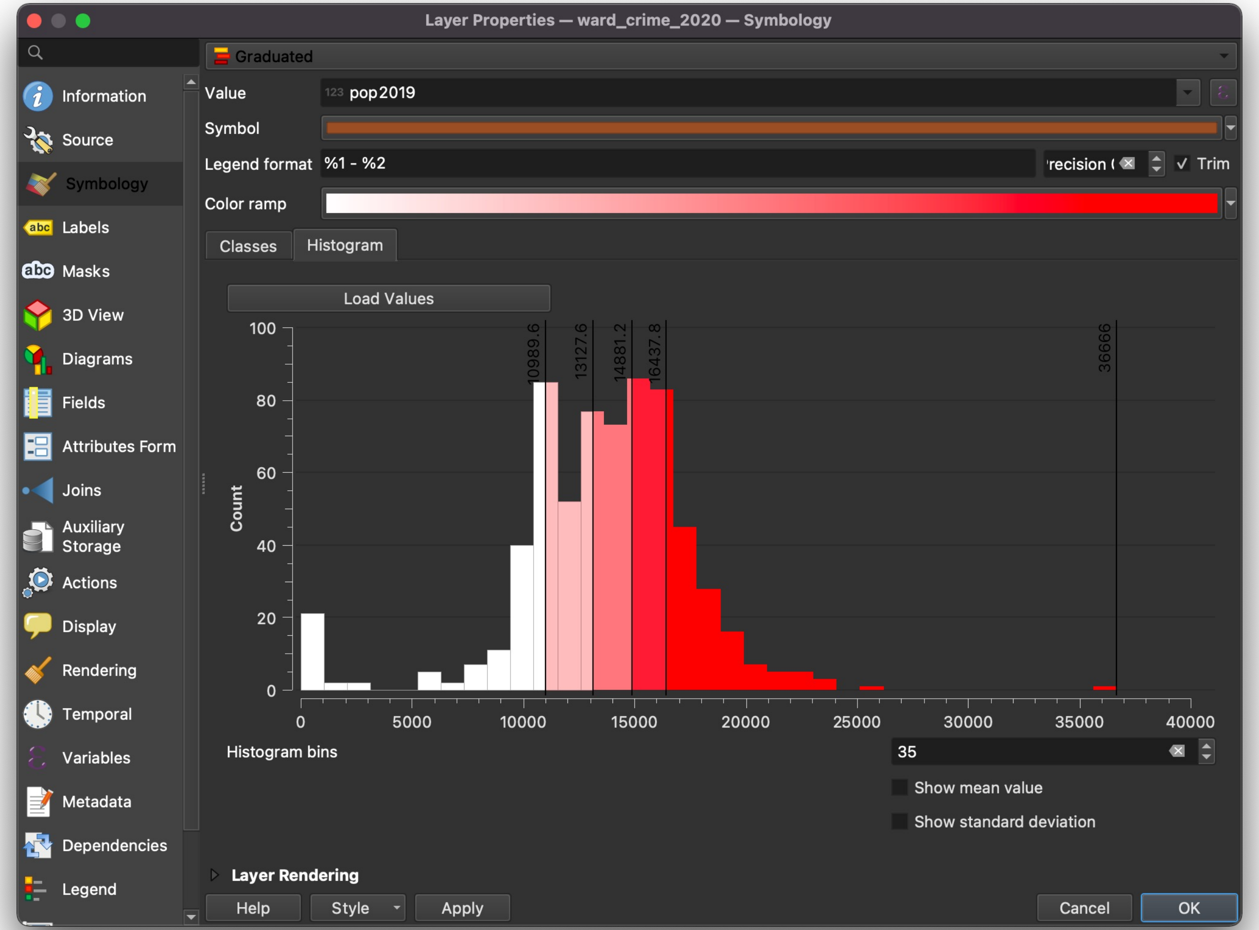
Natural breaks



Quantiles

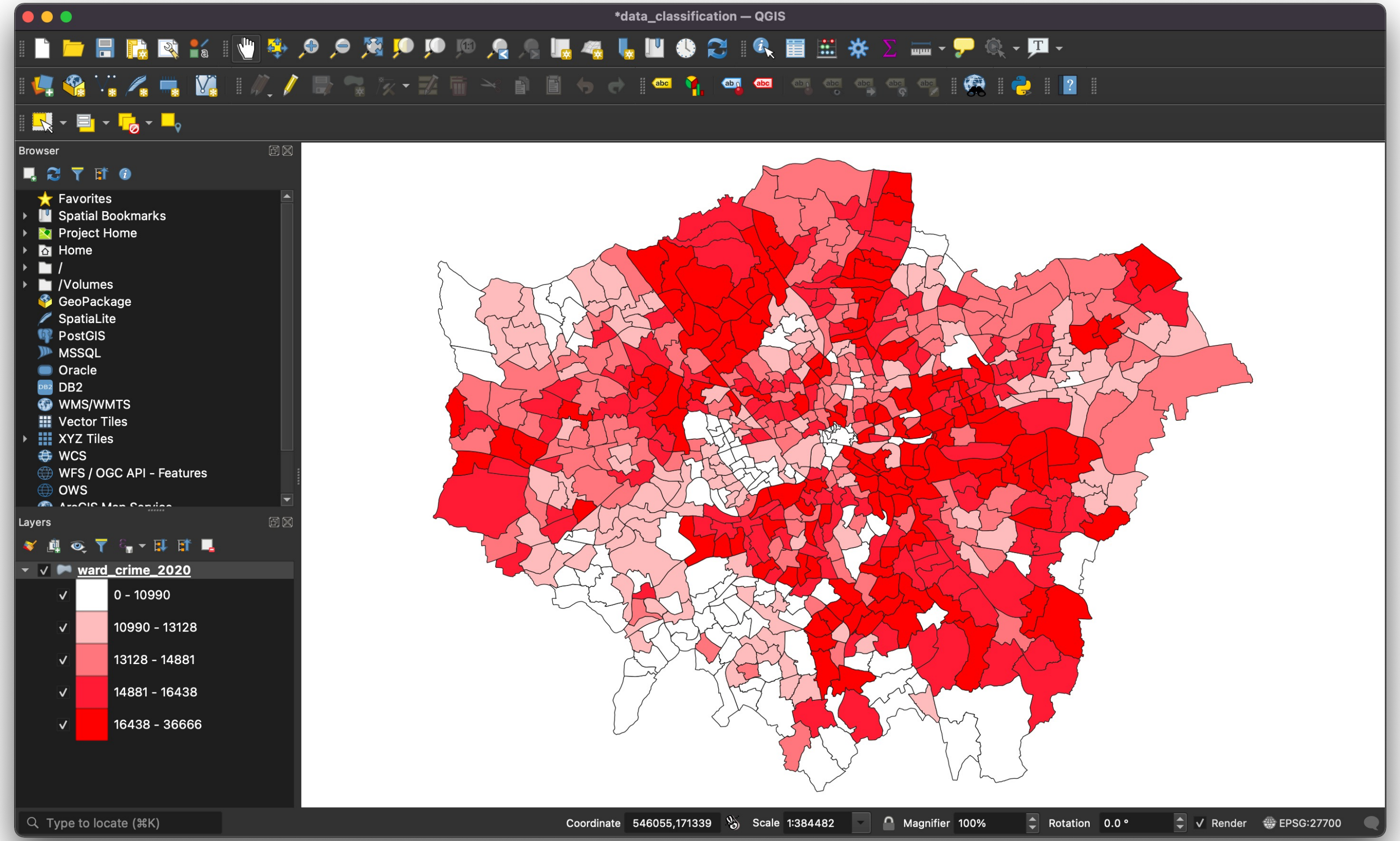
- Split the distribution keeping the same number of values in each bin.
- Splits are based on the rank of the value.
- If distribution is skewed: very different values can be binned together.

Quantiles



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
10989.6
13127.6
14881.2
16437.8
36666
Count
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

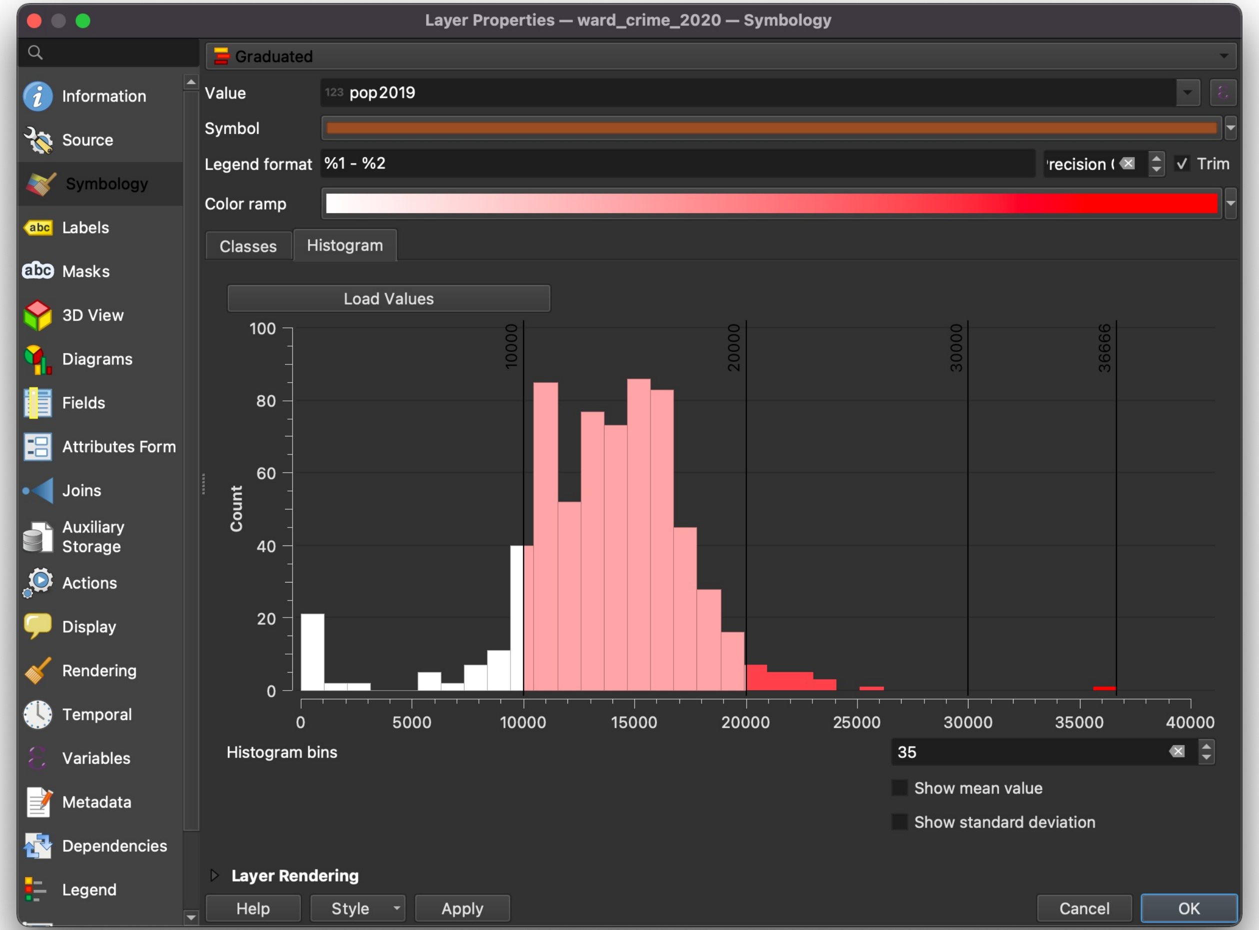
Quantiles


*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

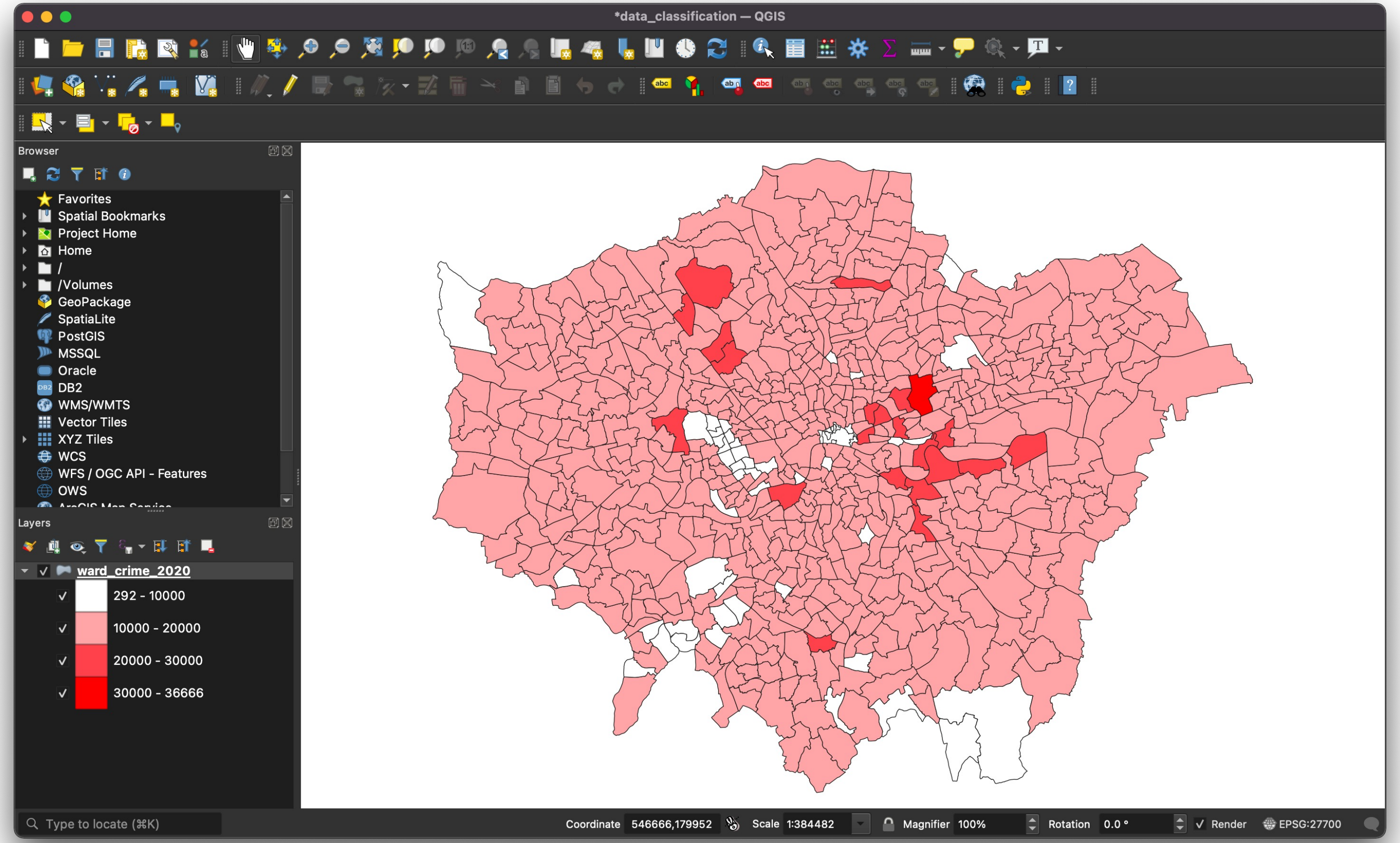
Pretty

- Split the distribution using a series of rounded values.
- Splits are not based on data or distribution.
- If distribution is skewed: bins with a low number of values.

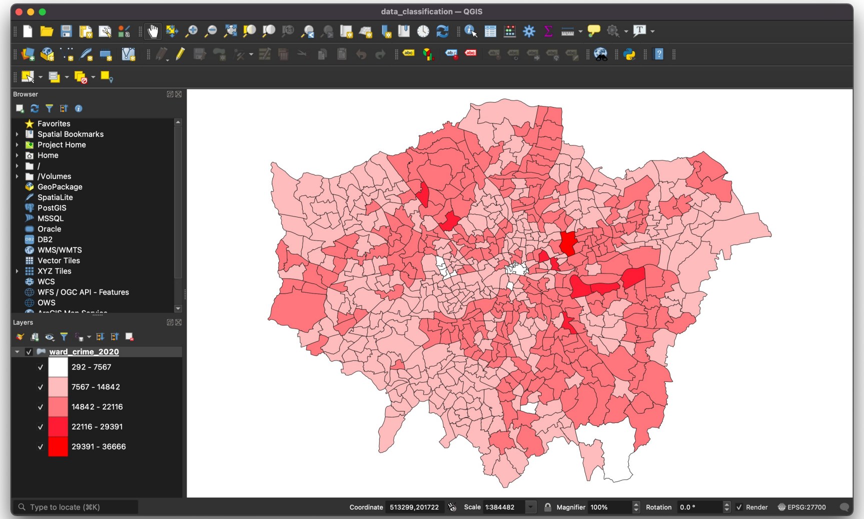
Pretty

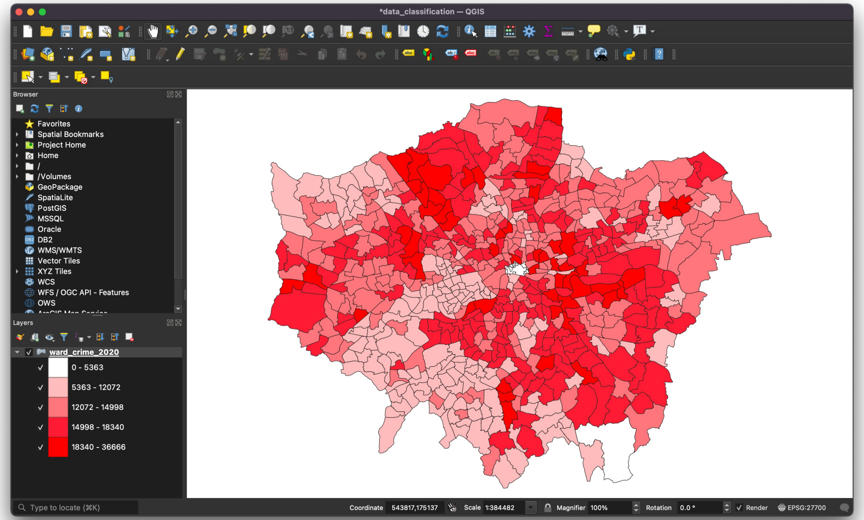

Layer Properties — ward_crime_2020 — Symbology
Information
Source
Symbology
Labels
Masks
3D View
Diagrams
Fields
Attributes Form
Joins
Auxiliary Storage
Actions
Display
Rendering
Temporal
Variables
Metadata
Dependencies
Legend
Graduated
Value
pop2019
Symbol
Legend format
%1 - %2
Trim
Color ramp
Classes
Histogram
Load Values
Count
100
80
60
40
20
0
10000
20000
30000
36666
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

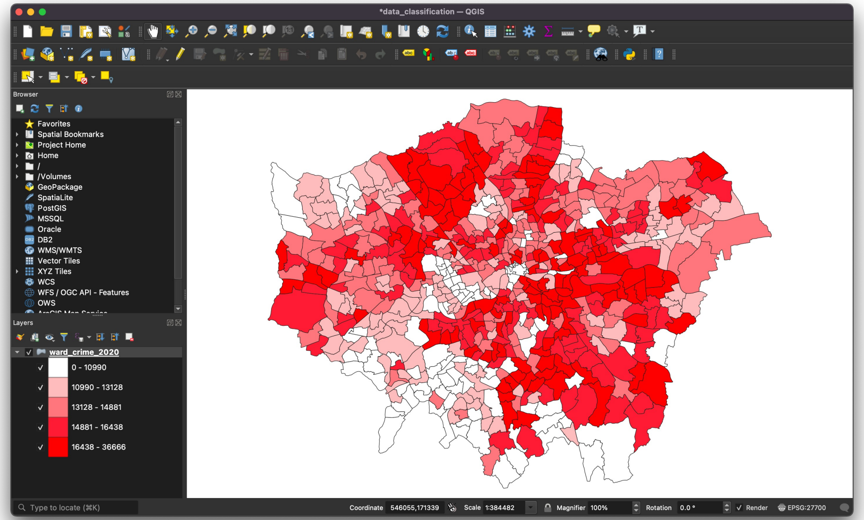
Pretty

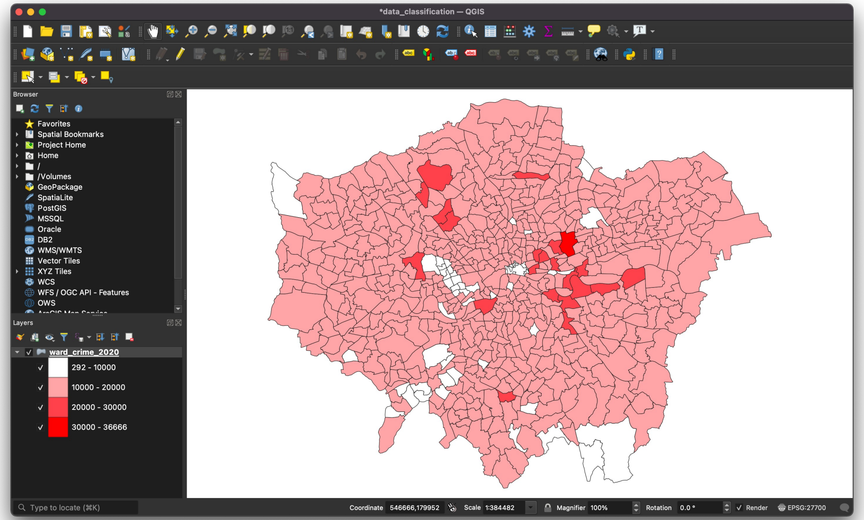

*data_classification — QGIS
Browser
Favorites
Spatial Bookmarks
Project Home
Home
/
/Volumes
GeoPackage
SpatiaLite
PostGIS
MSSQL
Oracle
DB2
WMS/WMTS
Vector Tiles
XYZ Tiles
WCS
WFS / OGC API - Features
OWS
Layers
ward_crime_2020
292 - 10000
10000 - 20000
20000 - 30000
30000 - 36666
Type to locate (⌘K)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

Different classifications, different maps









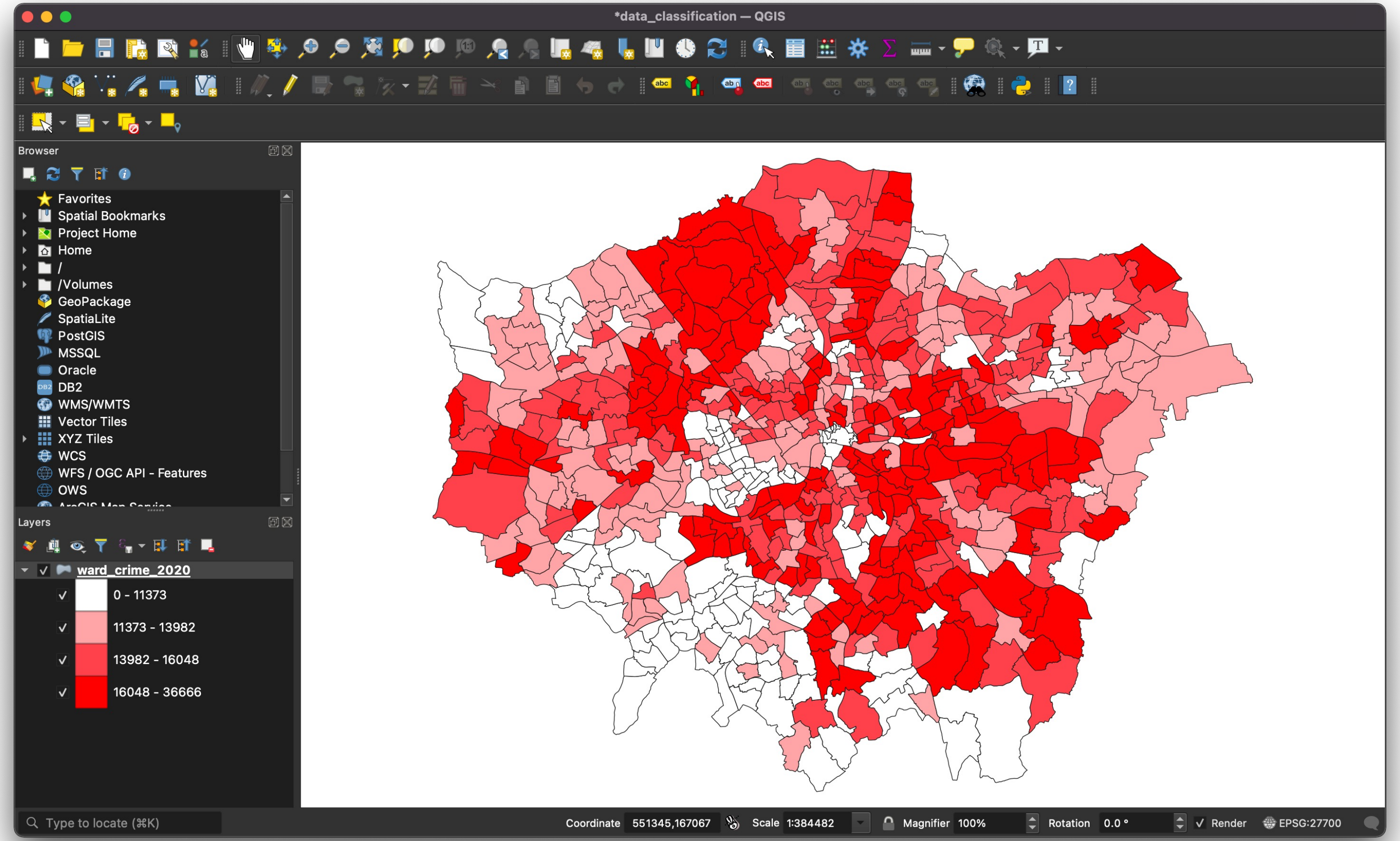
Some other classification options

- Standard deviation: shows you how much a feature's attribute value varies from the mean.
- Manual: good to highlight specific features in the data.

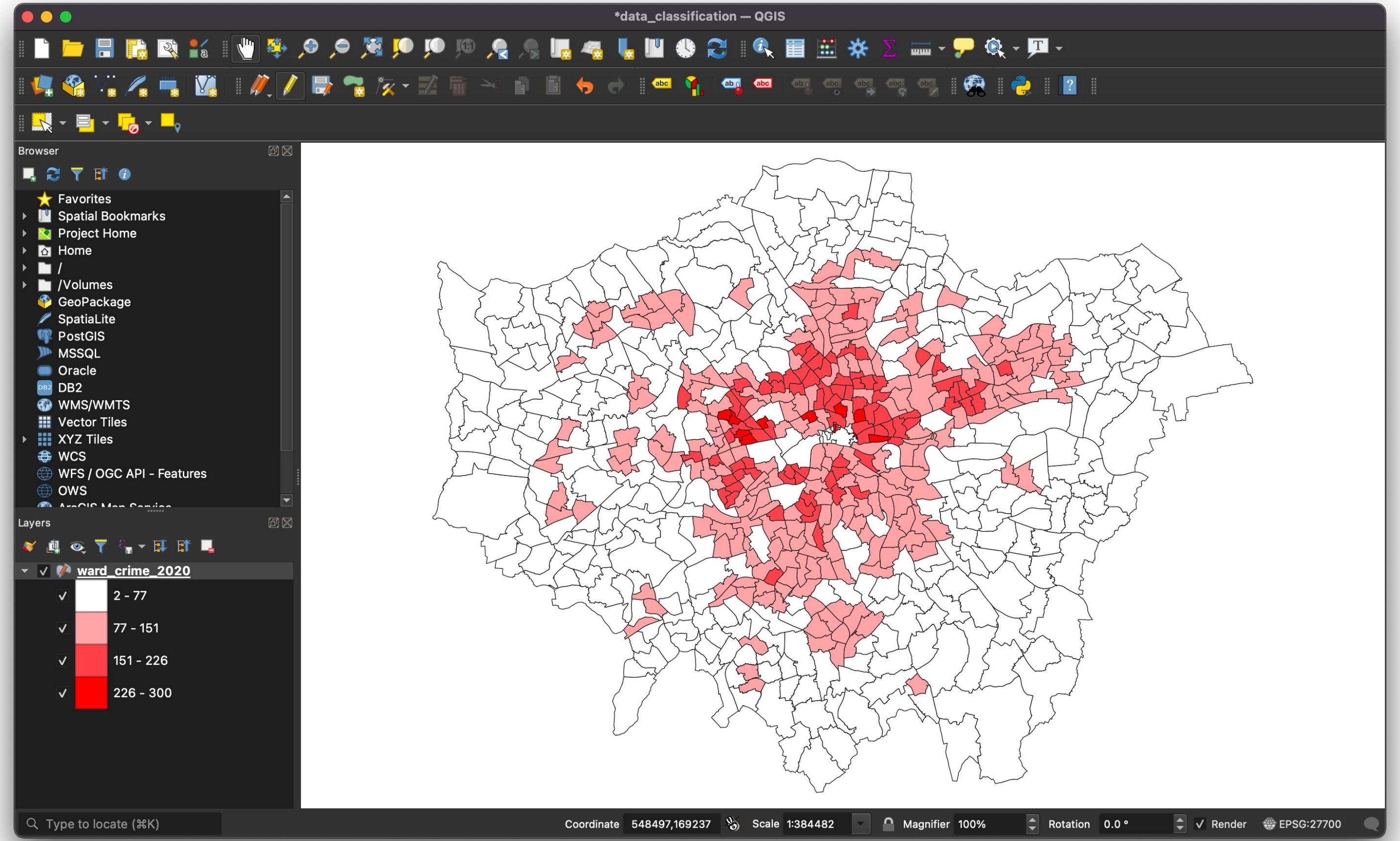
Standardisation

- Plotting raw numbers in choropleth maps should be avoided (almost all of the time).
- Standardisation can be achieved by dividing the variable of interest (numerator) by some standardising variable (denominator) – e.g. total population, area etc.

Standardisation



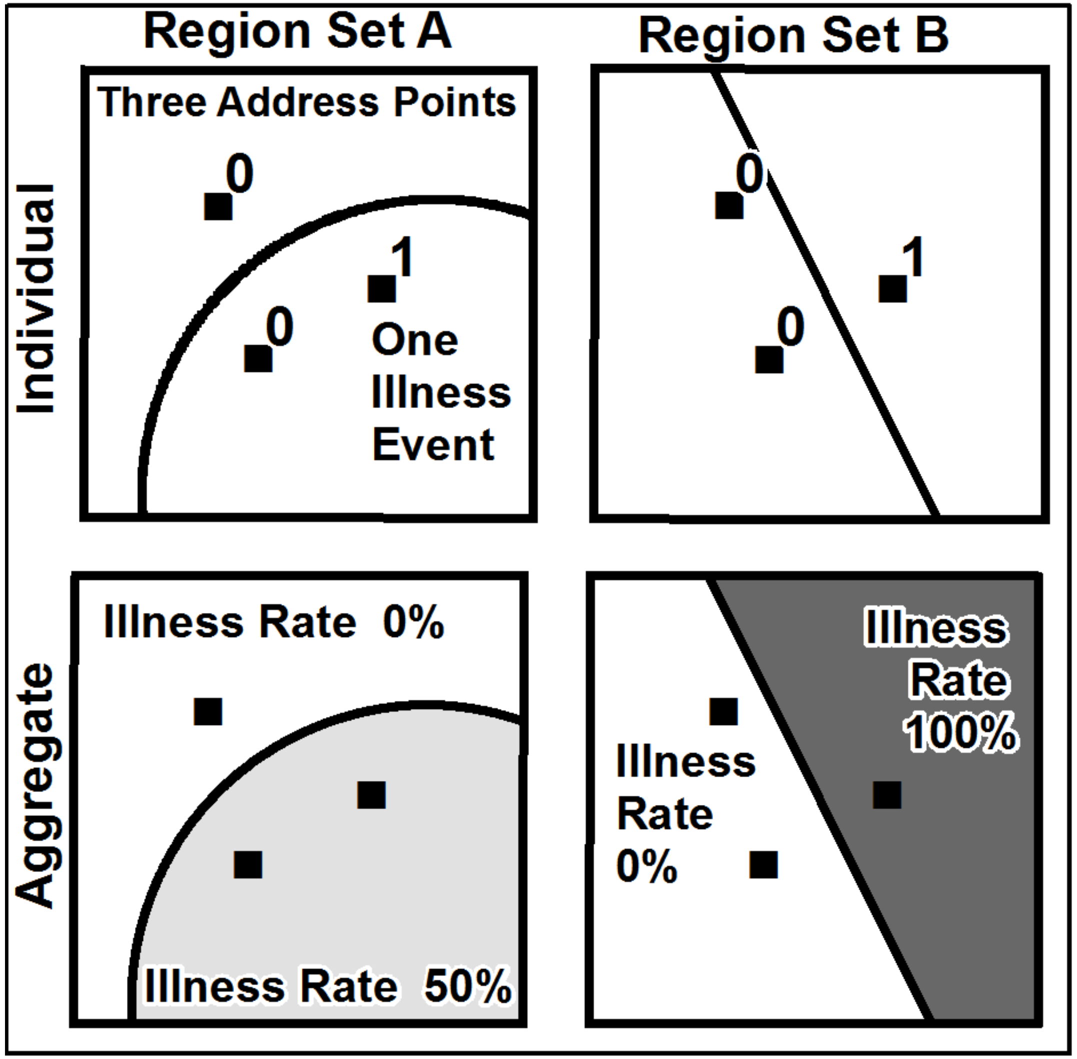
Standardisation



Be aware when mapping

- Ecological fallacy
- Modifiable Areal Unit Problem (MAUP)

Modifiable Areal Unit Problem



Map conventions

All maps should have a:

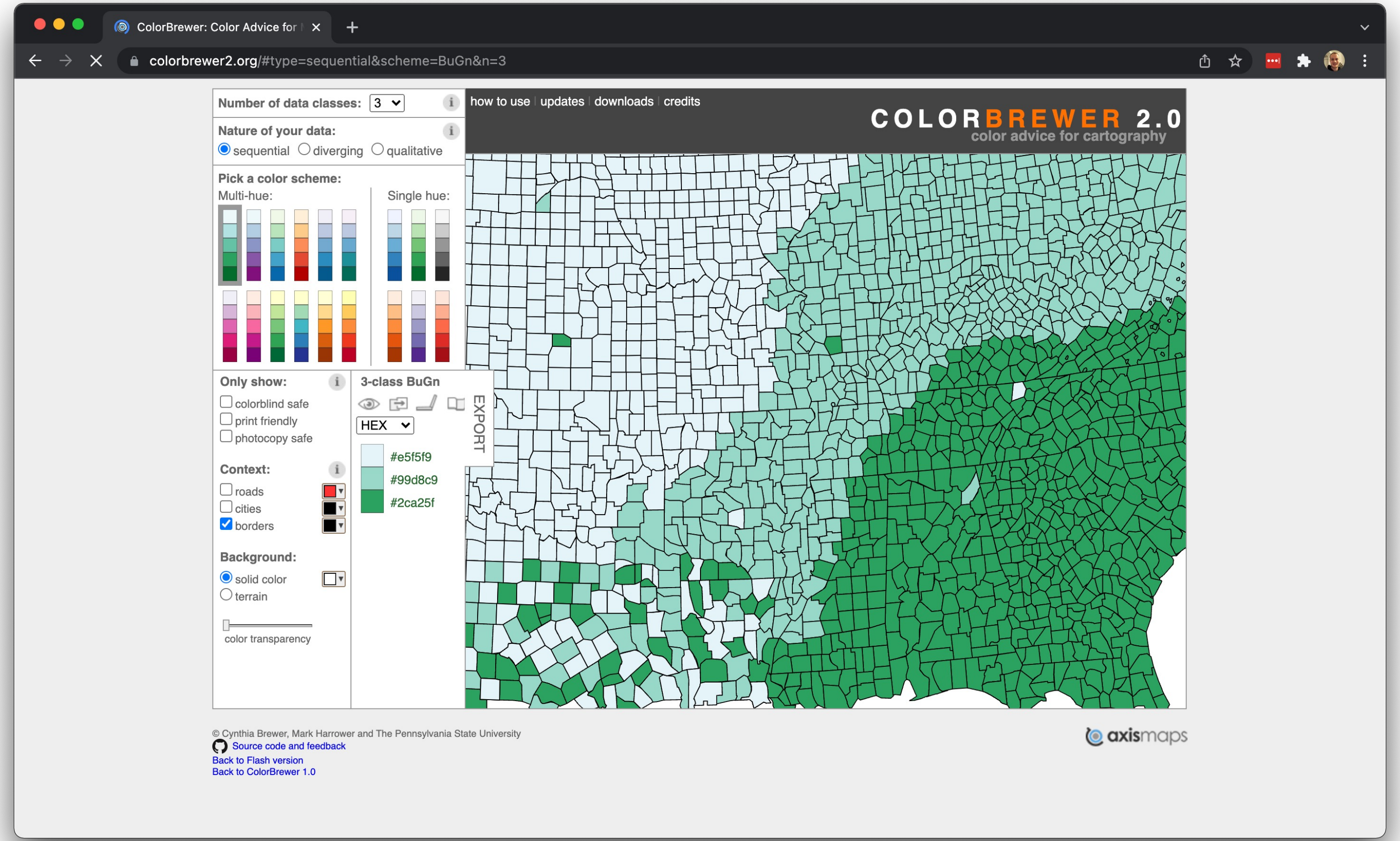
- Title
- Legend
- Scale bar / scale text
- North arrow
- And: attribution and data sources

Optional:

- Text and labels – sometimes important for context, sometimes distracting!

Map conventions

Colour choices

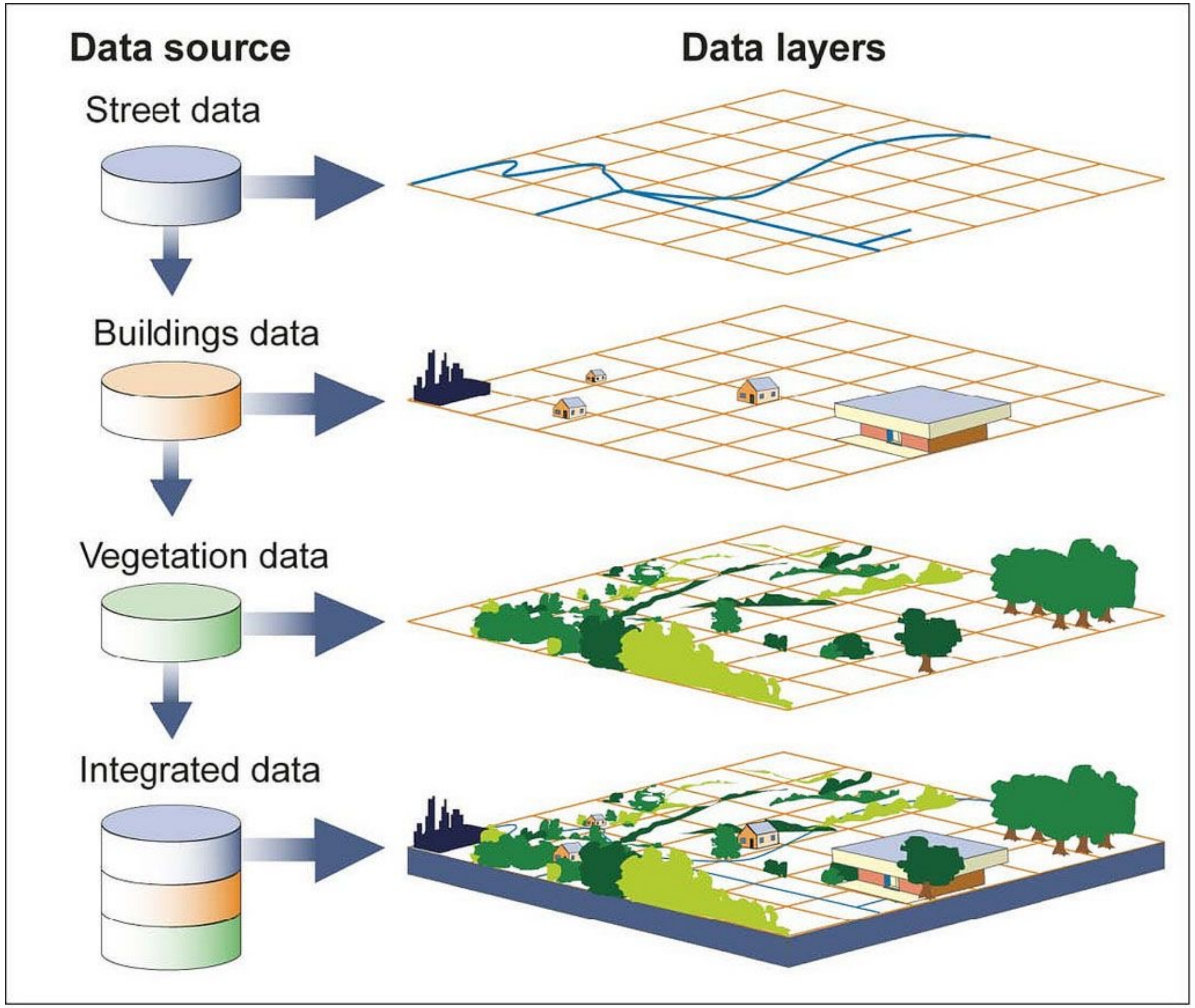


Effective data visualization is hard

- Requires the combination of good analysis and good visualisation. GIGO.
- What is the map trying to accomplish?
- Is the map suited to the audience?
- Have you included sufficient attribution information for data sources etc.?
- What are the likely impressions of the map?
- Are the data appropriate for the map's purpose?
- Does the symbolisation reflect the character of the phenomenon?
- Is the level of generalisation appropriate?

Part II: Positioning the map

Positioning the map



Source: GAO.

Positioning the map

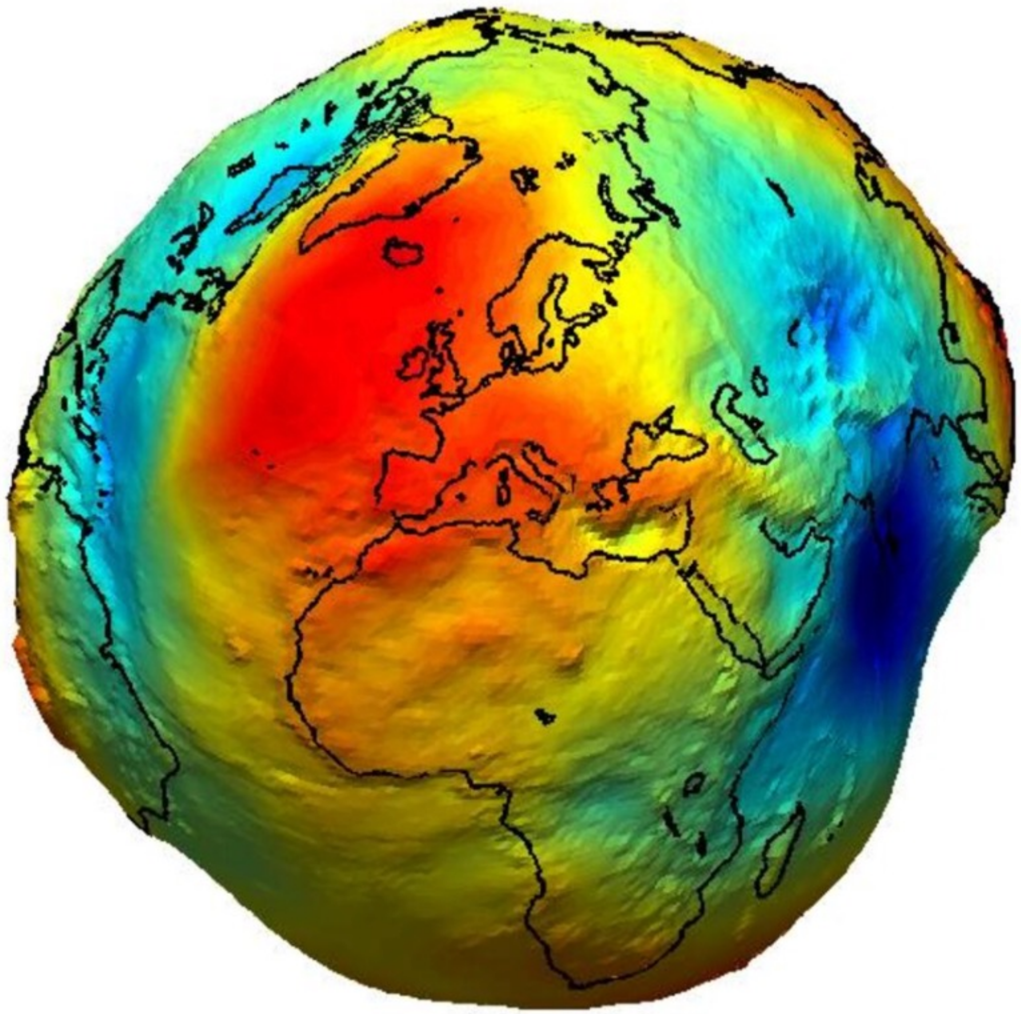
- Spatial data includes numerical information that allows you to position it on earth.
- These numbers are part of a coordinate system that provides a frame of reference for your data, to locate features on the surface of the earth, to align your data relative to other data, to perform spatially accurate analysis, and to create maps.
- How does this work?

Positioning the map

To be able to locate, integrate and visualise spatial data accurately within a GIS system or digital map, spatial data needs to have two things:

- Coordinate Reference System (CRS) / Geographic Coordinate System (GCS)
- Map Projection / Projected Coordinate System (PCS)

Representing the globe



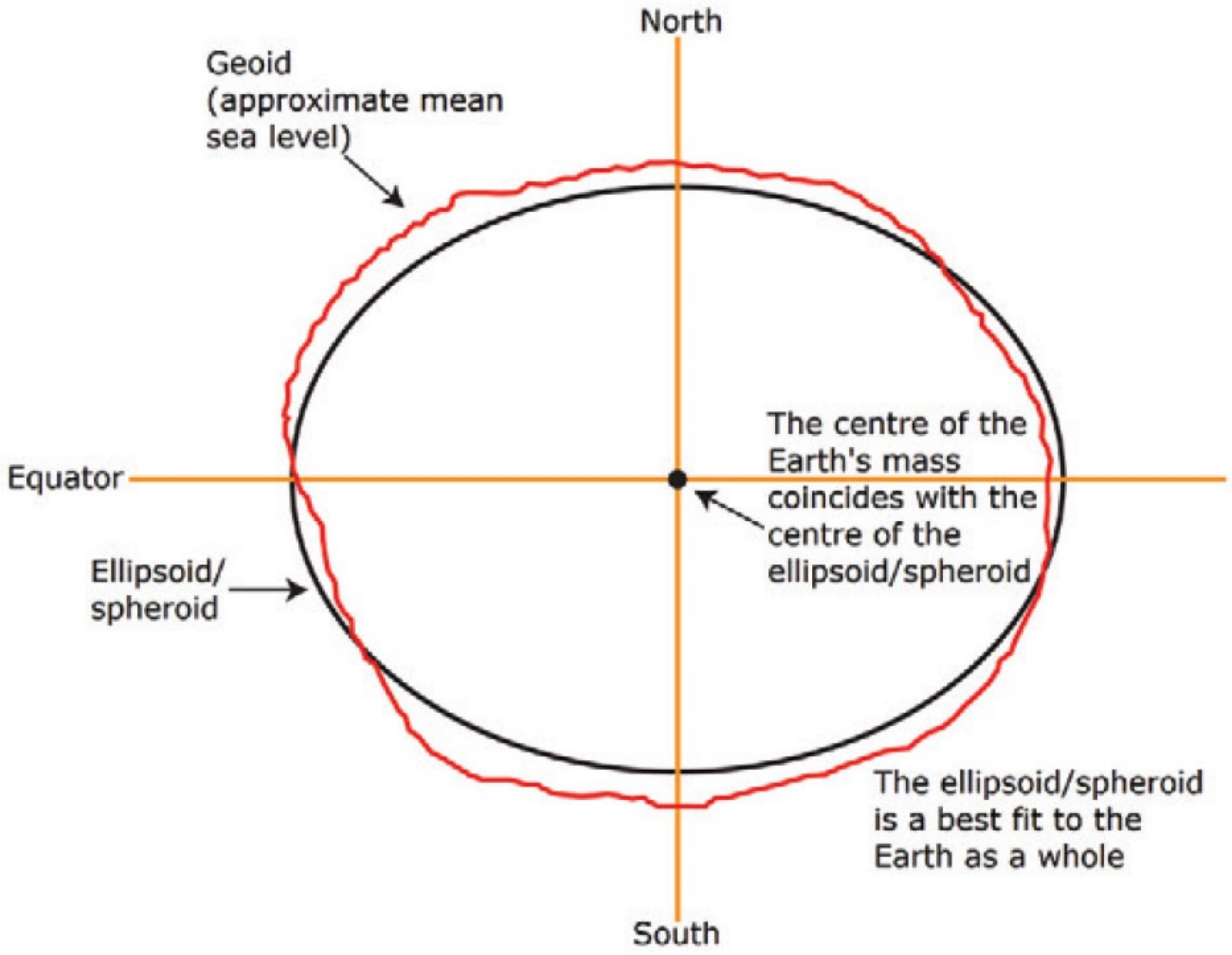
Representing the globe

- Representing the earth’s true shape, the geoid, as a mathematical model is crucial for a GIS environment.
- The earth is not a perfect sphere (although this makes calculations much easier), but more closely resembles an ellipsoid.
- In GIS we use a reference ellipsoid as an approximation of the surface of the earth. A reference ellipsoid is a mathematically-defined surface that roughly matches the model of the earth when its topographical features are removed.

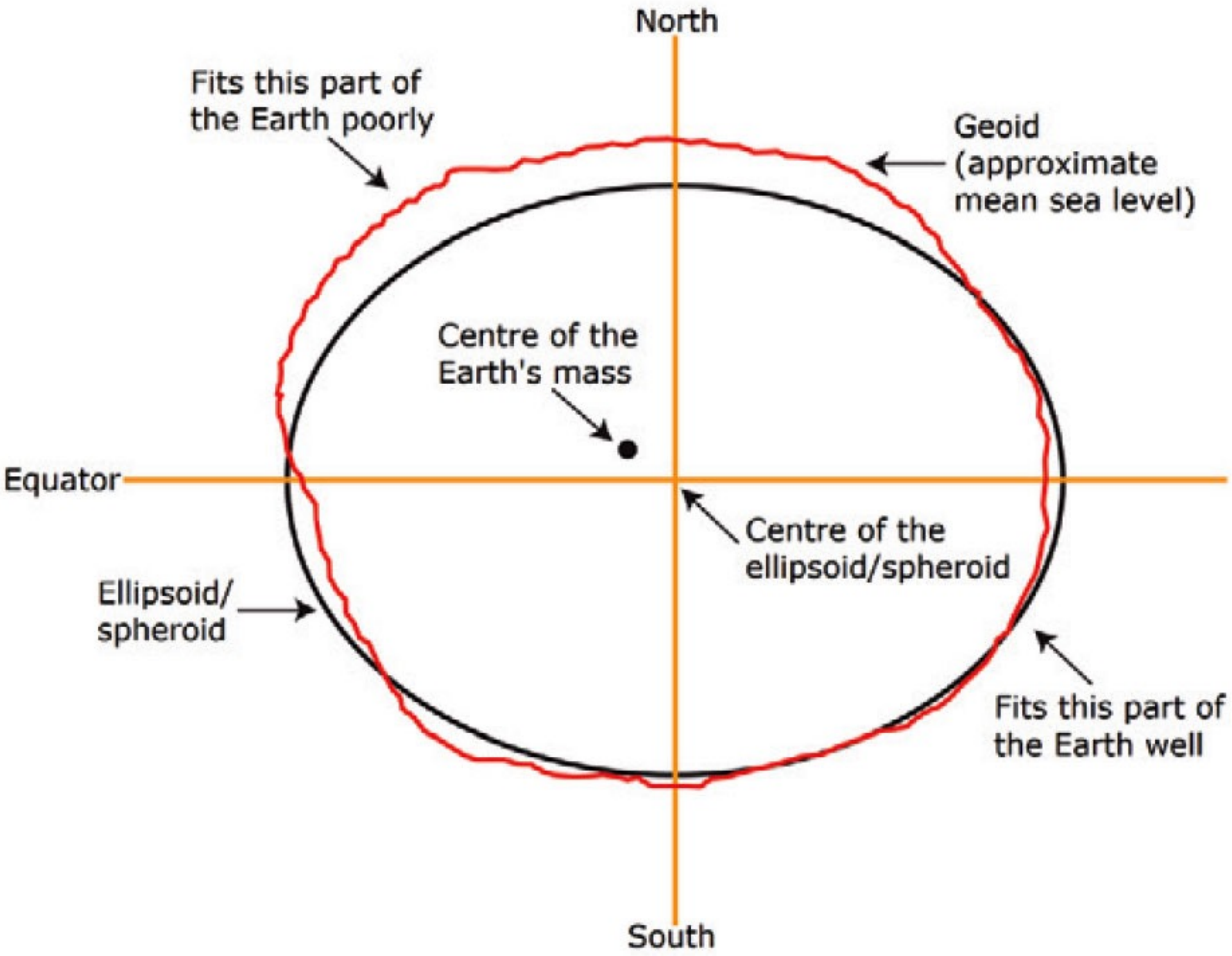
Representing the globe

- Once we have the reference ellipsoid, we can try to align the geoid with the reference ellipsoid representation of the earth and to map the earth’s surface features onto the reference ellipsoid.
- Because a reference ellipsoid is an approximation, there are locations where an ellipsoid exactly matches the geoid and there are locations where an ellipsoid is way way off.

Aligning the geoid with the reference ellipsoid



Aligning the geoid with the reference ellipsoid



Geographic Coordinate Systems

- The alignment can be local where the ellipsoid surface is closely fit to the geoid at a particular location on the earth’s surface or geocentric where the ellipsoid is aligned with the center of the earth.
- How one chooses to align the ellipsoid to the geoid defines a datum.
- The combination of a geoid, reference ellipsoid and the way these two are aligned (the datum) is called a Geographic Coordinate System (GCS).

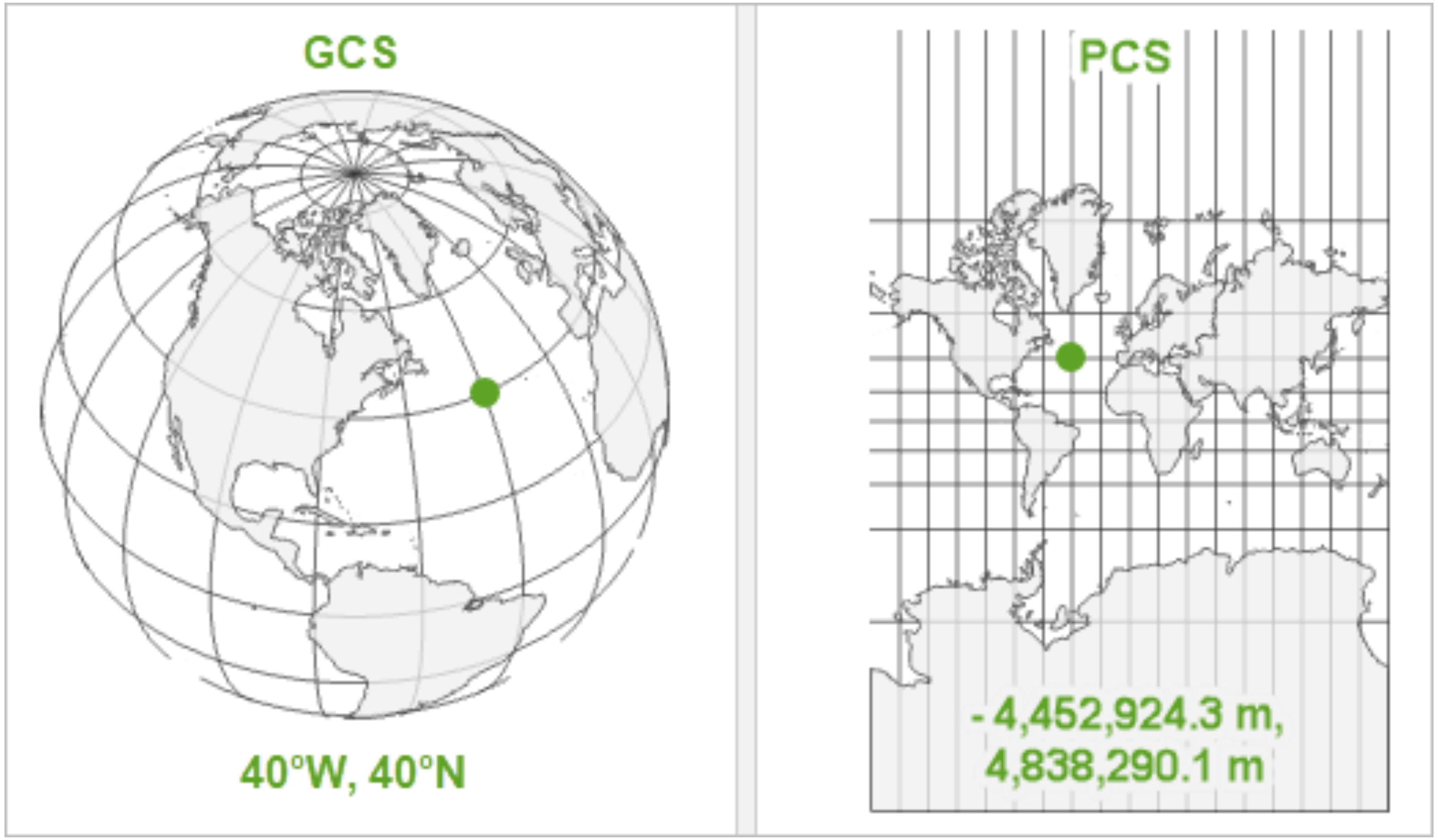
Geographic Coordinate Systems

- A Geographic Coordinate System (GCS) defines where the data is located on the earth's surface.
- In a GCS locations are measured in angular units from the center of the earth relative to two planes: a plane defined by the equator and a plane defined by the prime meridian.
- In a GCS a location is thus defined by two values: a latitudinal value (in reference to the equator plane) and a longitudinal values (in reference to the plane as defined by the prime meridian).

Projected Coordinate Systems

- Where a Geographic Coordinate System (GCS) defines where the data is located on the earth's surface, a Projected Coordinate System (PCS) defines how to draw the data on a flat surface.
- A GCS is round, and so records locations in angular units (usually degrees). A PCS is flat, so it records locations in linear units (usually meters) on a cartesian plane.

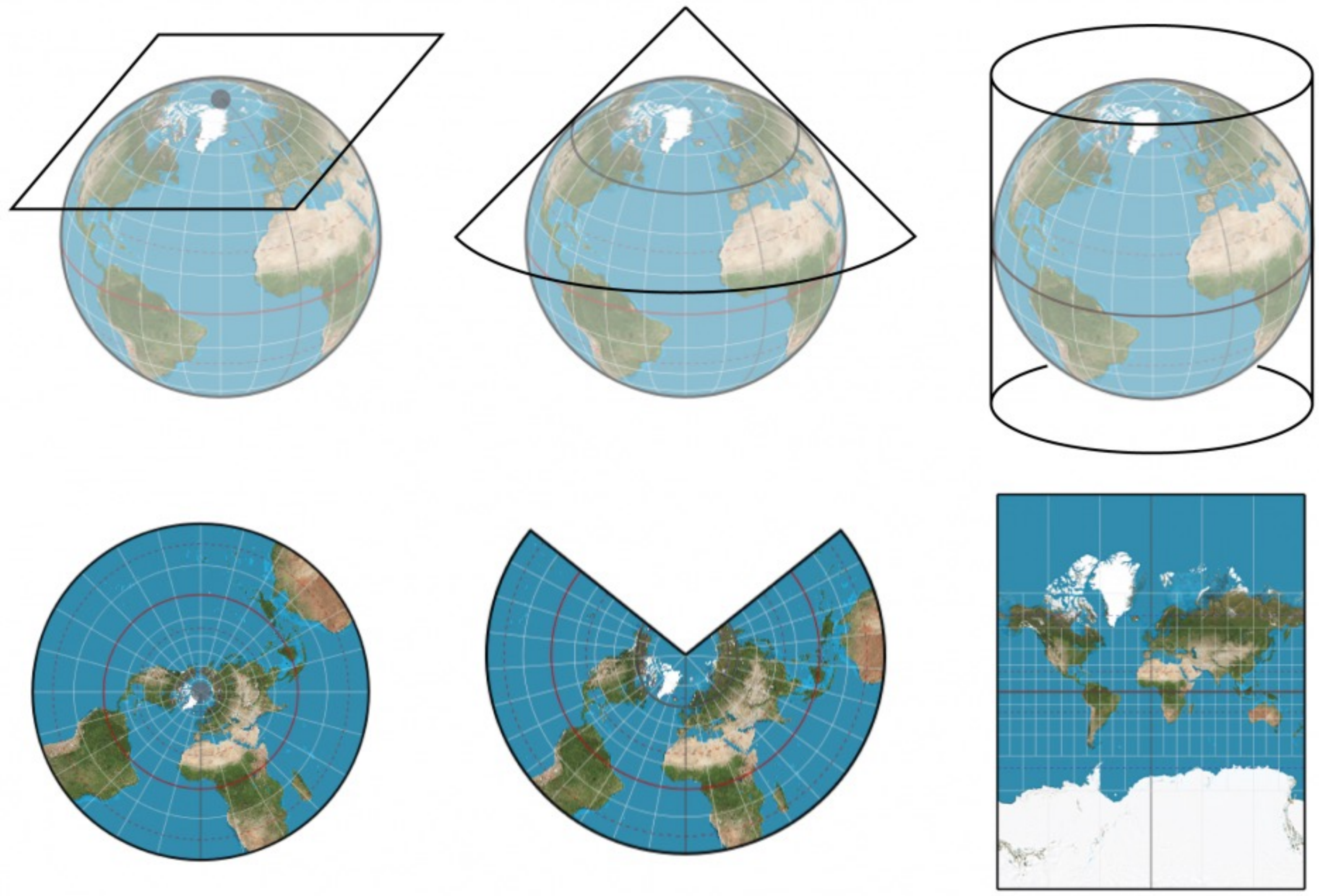
Projected Coordinate Systems



Projected Coordinate Systems

- A PCS is a reference system for identifying locations and measuring features on a flat (map) surface.
- Going from a GCS to a PCS requires mathematical transformations.
- Many ways of doing this, but there are three groups: planar, conical, and cylindrical.

Projected Coordinate Systems



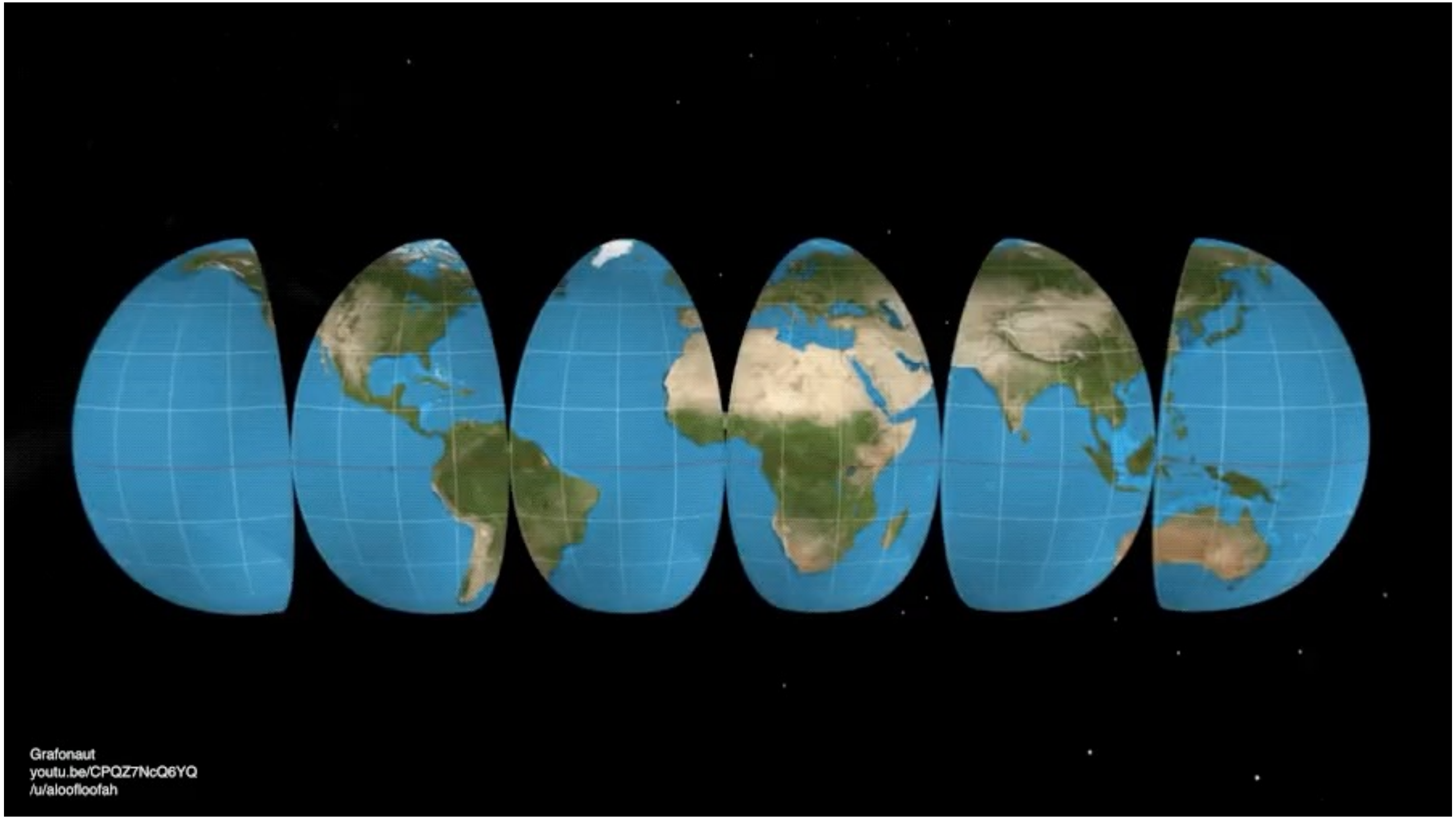
Projected Coordinate Systems

- A planar projection maps the earth surface features to a flat surface that touches the earth's surface at a point. Often used at the poles.
- A conical map projection maps the earth surface onto a map rolled into a cone.
- A cylindrical map projection maps the earth surface onto a map rolled into a cylinder (which can then be flattened into a plane).
- All projections distort real-world geographic features in terms of shape, area, distance, or direction.
- Each map projection is good at preserving only one or two of the four spatial properties.

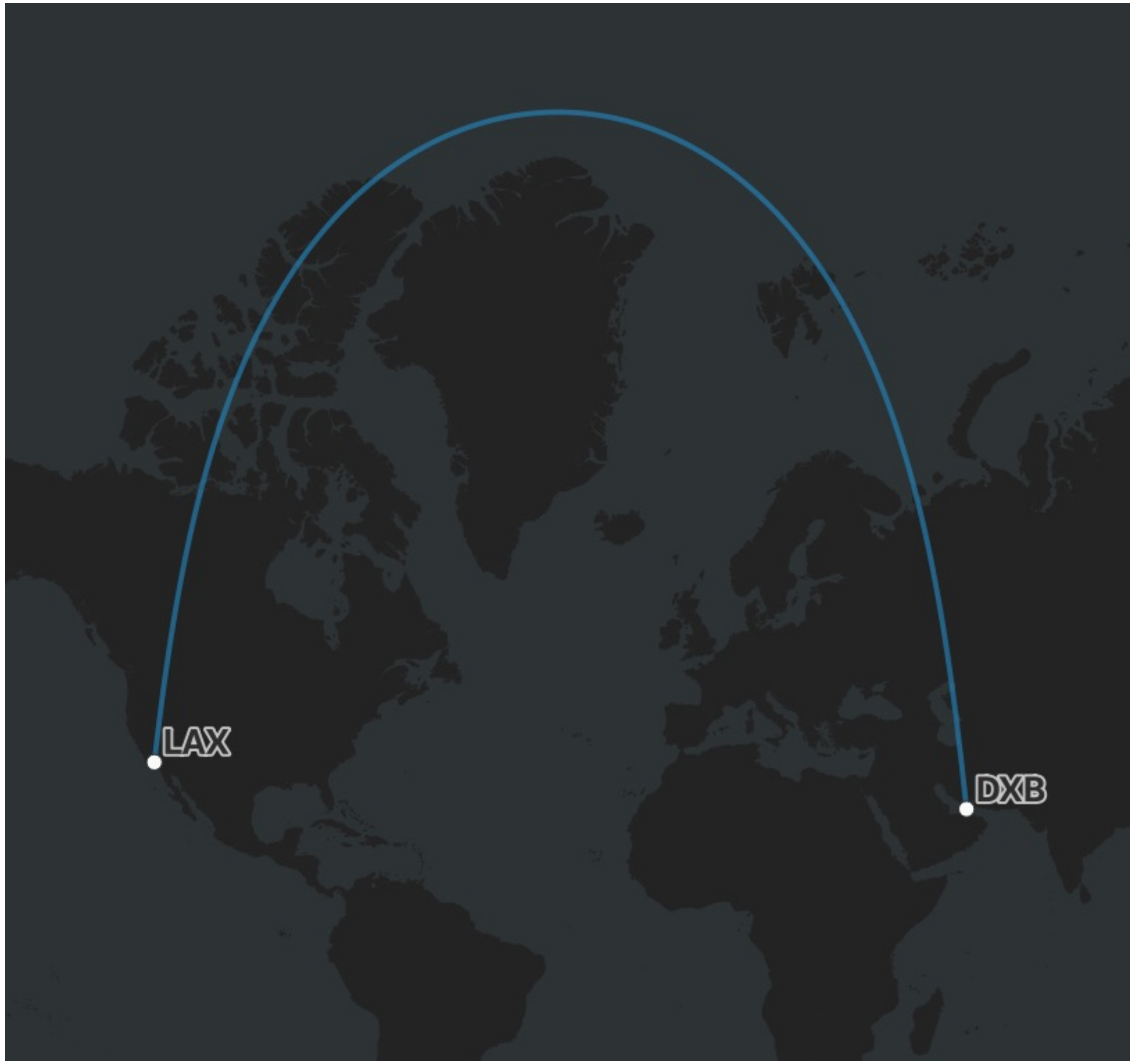
Projected Coordinate Systems

- Probably the most in(famous) projection to represent the world on a flat surface is the Mercator projection. Often used on the web.
- However: it exaggerates the size of countries near the poles, while downplaying the size of those near the equator.

Projected Coordinate Systems

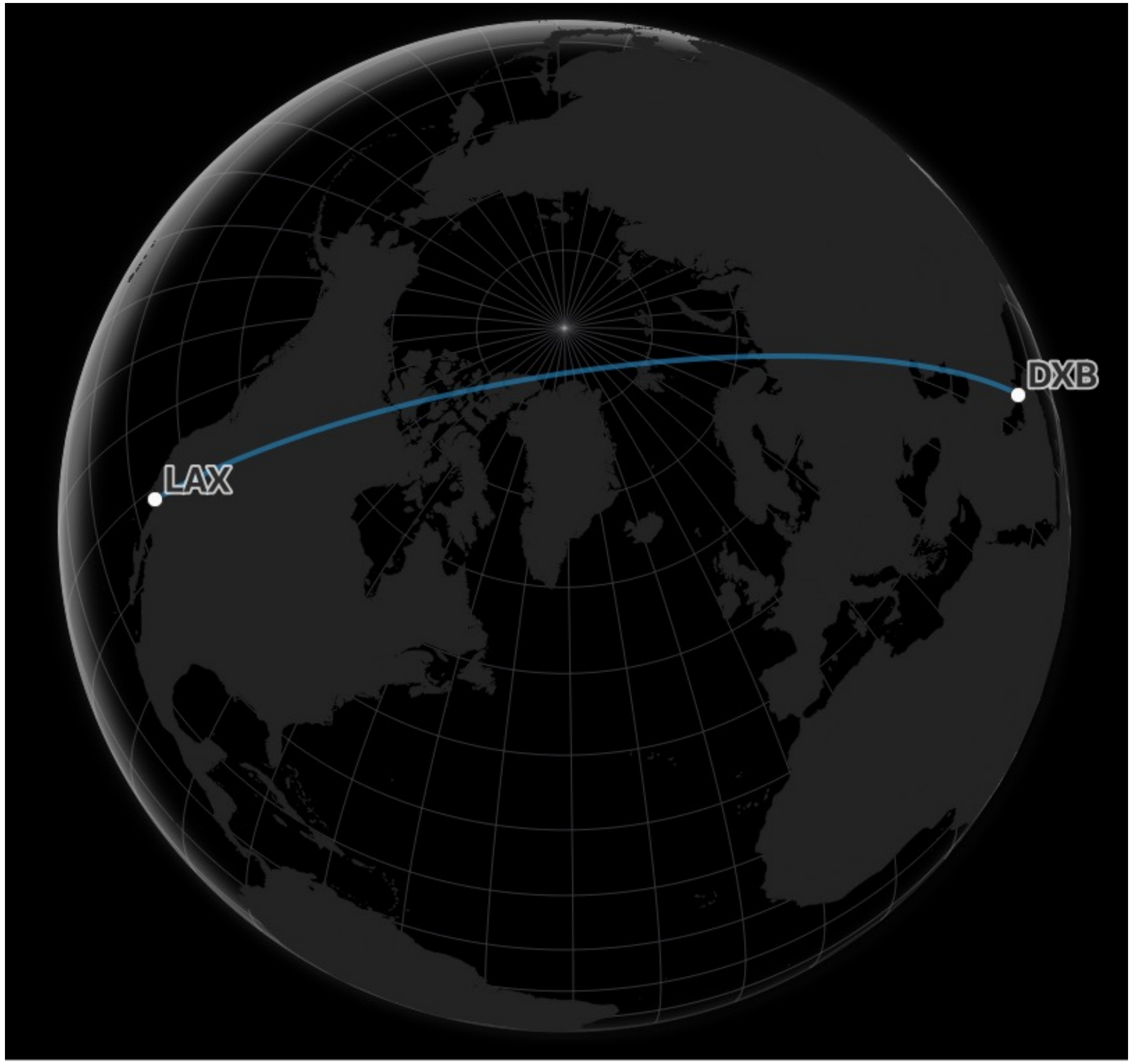


Projected Coordinate Systems



Alasdair Rae. *https://automaticknowledge.co.uk/*

Projected Coordinate Systems



Alasdair Rae. *https://automaticknowledge.co.uk/*

Projected Coordinate Systems

Effective visualisation of the effects of the Mercator projection on individual countries:

https://thetruesize.com/

Working with GCS and PCS

- World Geodetic System 1984 (WGS 1984) is designed as a one-size-fits-all GCS, good for mapping global data. But may not be the best option locally. Uses latitude and longitude.
- British National Grid PCS is based on UTM (Universal Transverse Mercator); one of the advantages of the U.K. national grid over the global UTM coordinate system is that it eliminates the boundary between different UTM zones.
- For large countries or areas, different regions may need different projections.
- Existing data sources will typically already have a GCS and projection assigned; make sure that when combining different sources, the projections are the same.

Working with GCS and PCS

- Details of projection are often contained in an EPSG Code. The EPSG registry is a public registry of geodetic datums, spatial reference systems, Earth ellipsoids, coordinate transformations and related units of measurement.
- EPSG: 4326 WGS84 (often used for GPS coordinates)
- EPSG: 27000 British National Grid
- Spatial datasets can be transformed from one GCS or PCS to another: “reprojected”

Conclusion

Part I

- To design a good map, lots of choices have to be made: type of map, type of classification, standardisation, MAUP, colours.
- Good maps adhere to the map conventions: elements that should all be present on a map such as title, legend, scale bar.

Part II

- GCS and PCS are essential for GIS as they determine how spatial layers are linked to the earth.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk