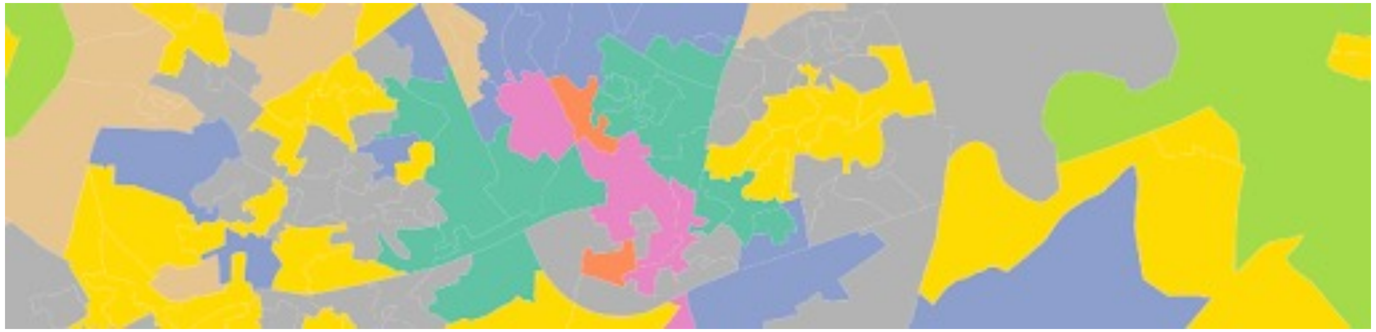
Principles of Spatial Analysis

WEEK 03: SPATIAL OPERATIONS



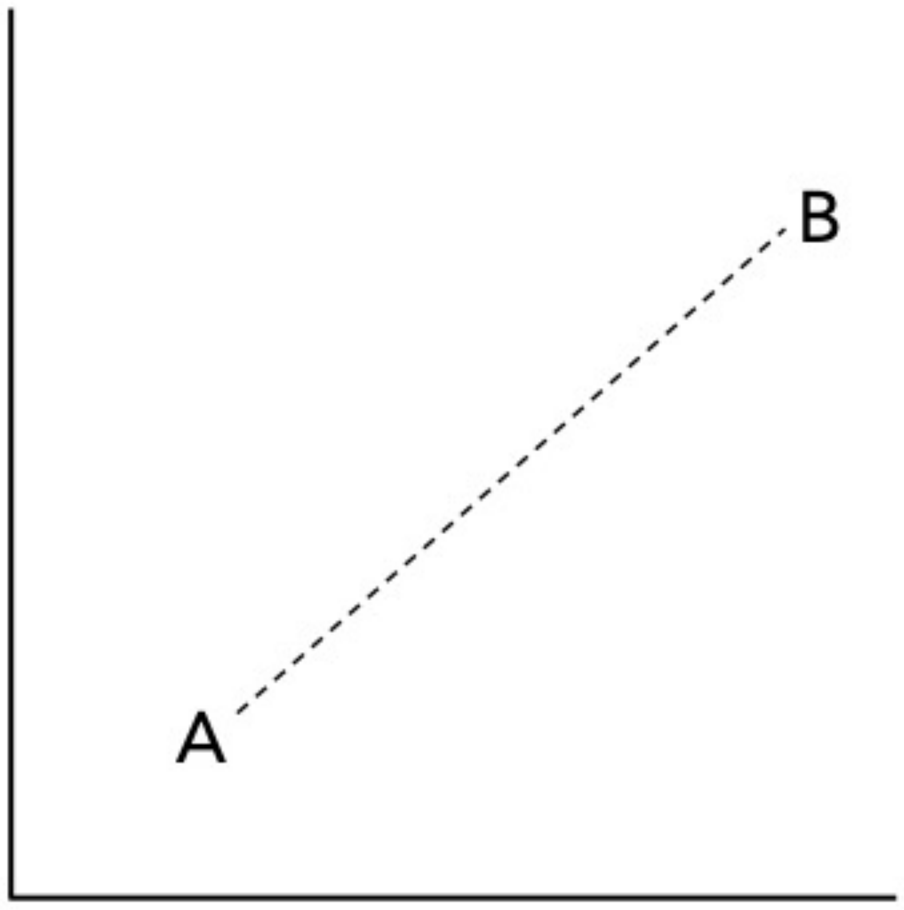
This week

- Spatial properties
- Geometric operations
- Spatial relationships
- Coordinate systems and projections

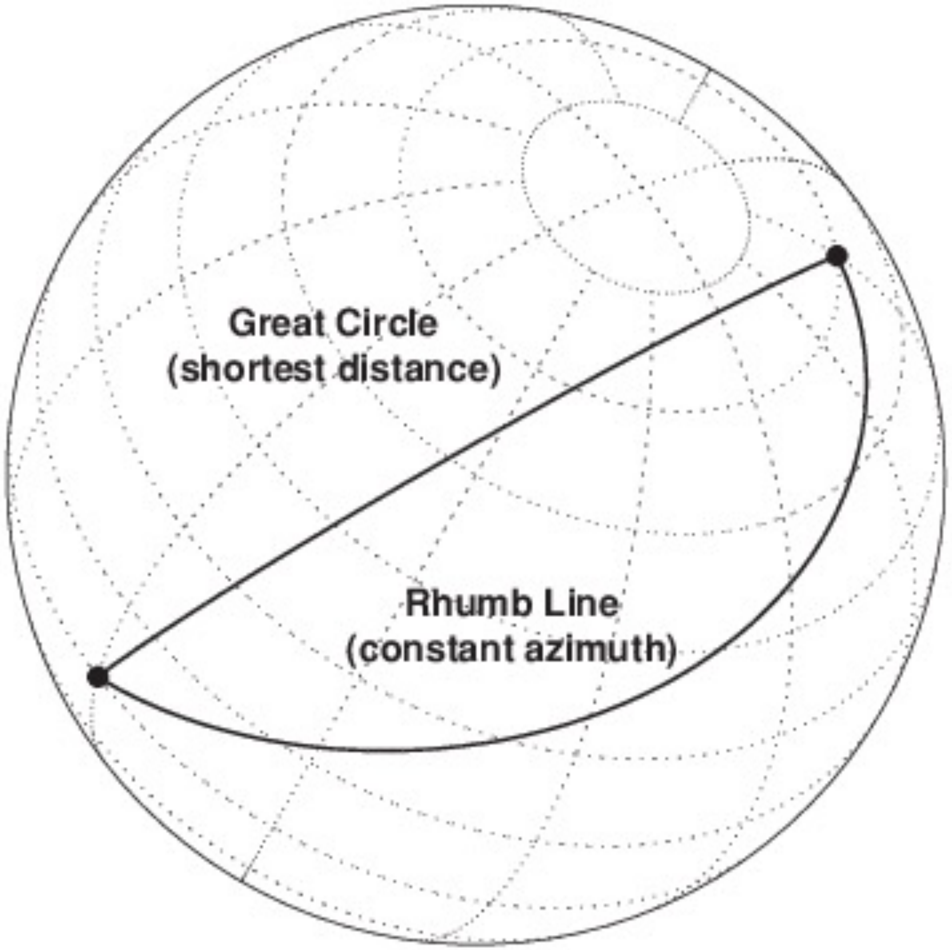
Spatial properties

- Much of spatial data processing is ‘spatial maths’ on several spatial properties.
- We typically work with ‘things’ like distance, area, and shape.
- Different ways to think about these properties and how to conceptualise them.

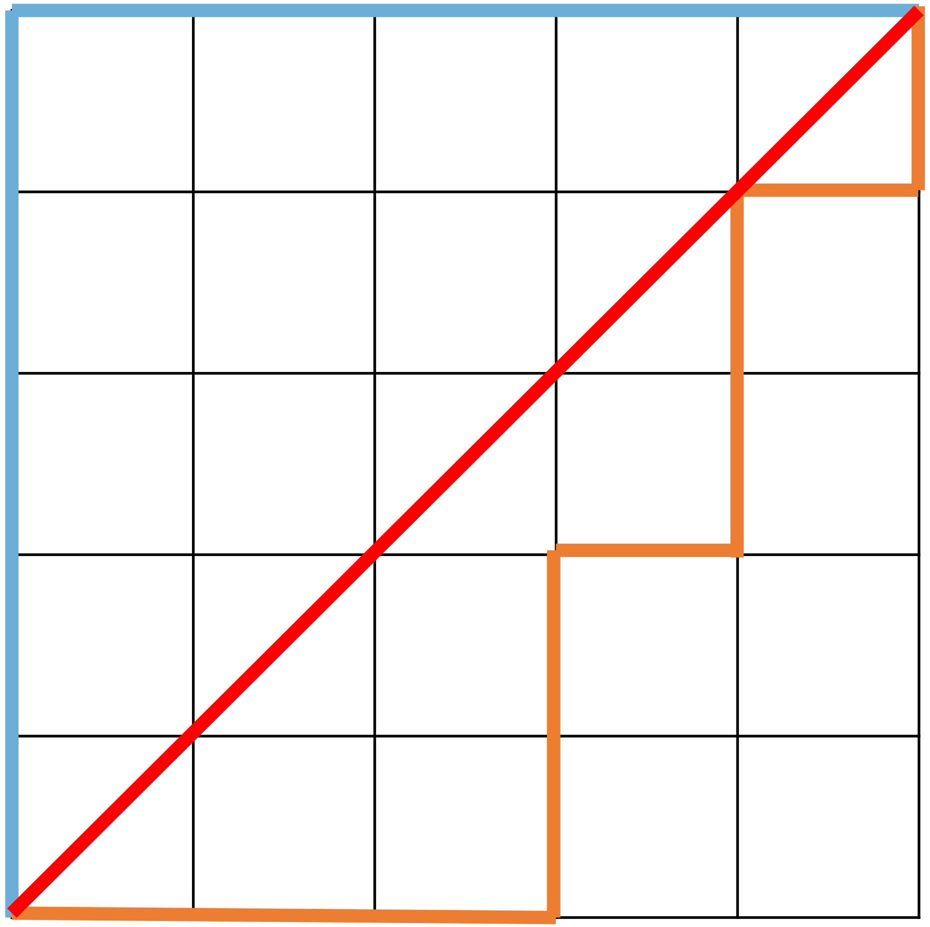
Distance I



Distance II


Great Circle
(shortest distance)
Rhumb Line
(constant azimuth)

Distance III

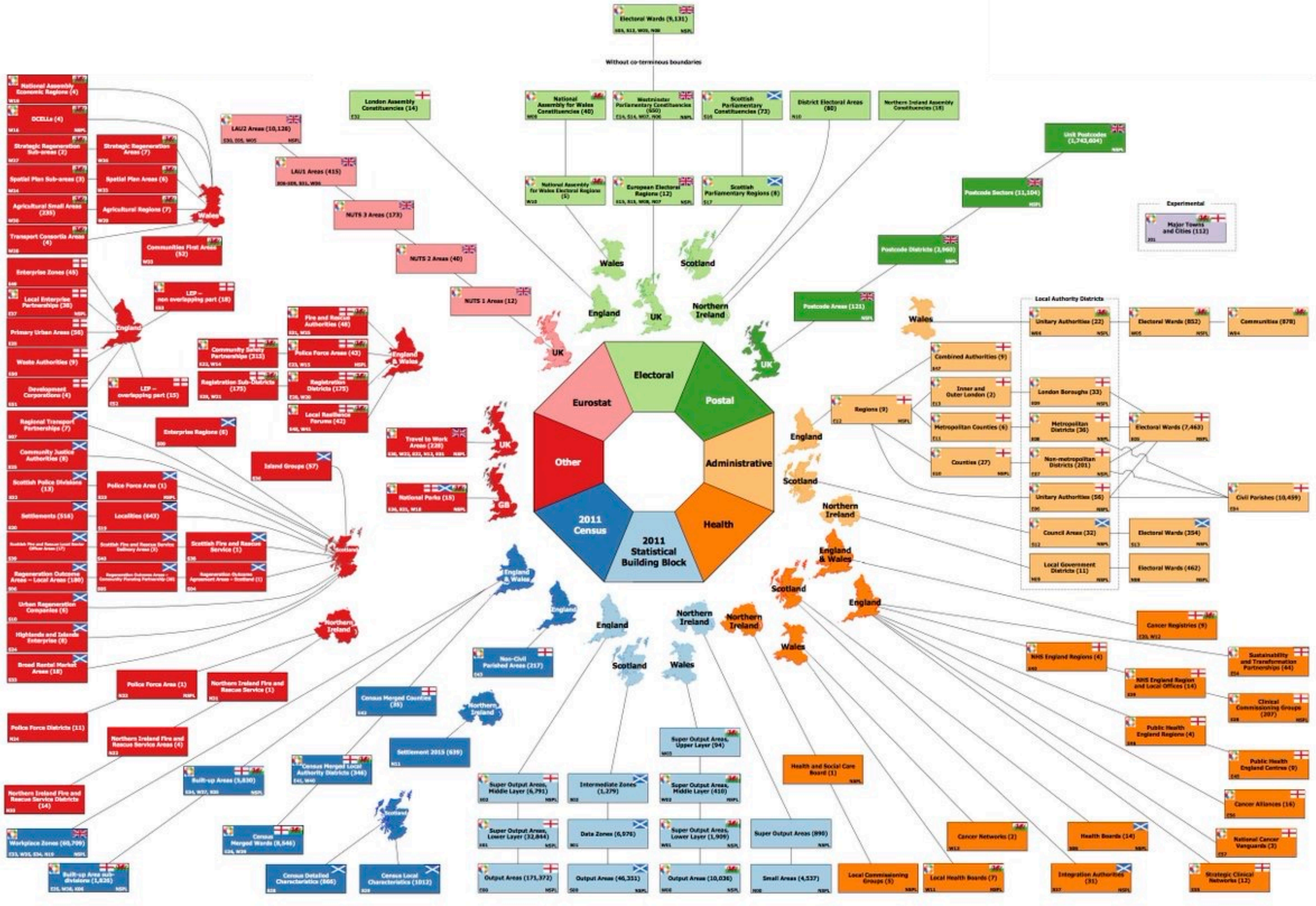


Area I

Area II

- Precision of the data source will affect calculations (e.g. simplified topology).
- Decision of which geography to use is crucial and depends on what you want to investigate – and keep in mind MAUP.

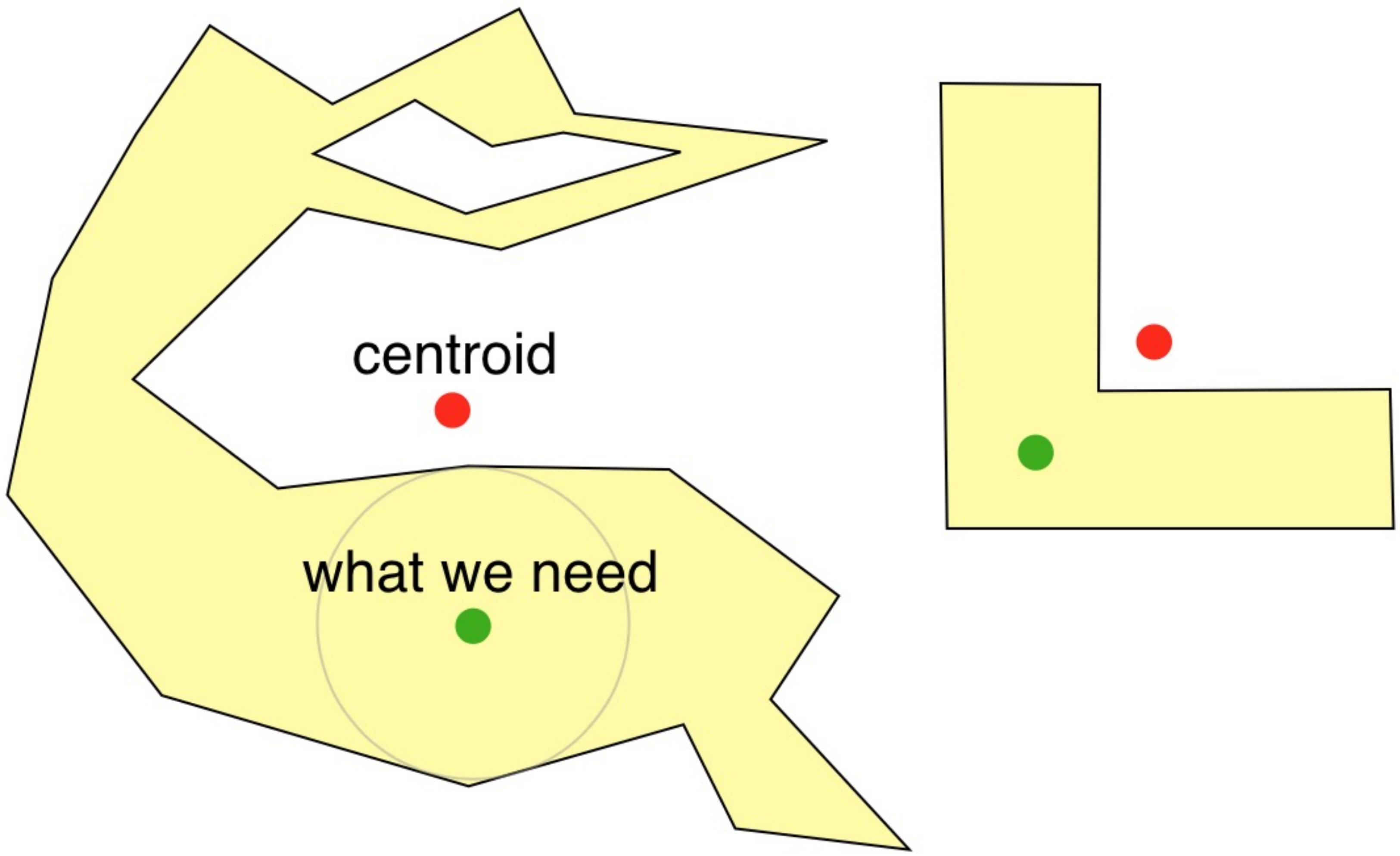
Area III


Electoral
Postal
Administrative
Health
2011 Statistical Building Block
2011 Census
Other
Eurostat

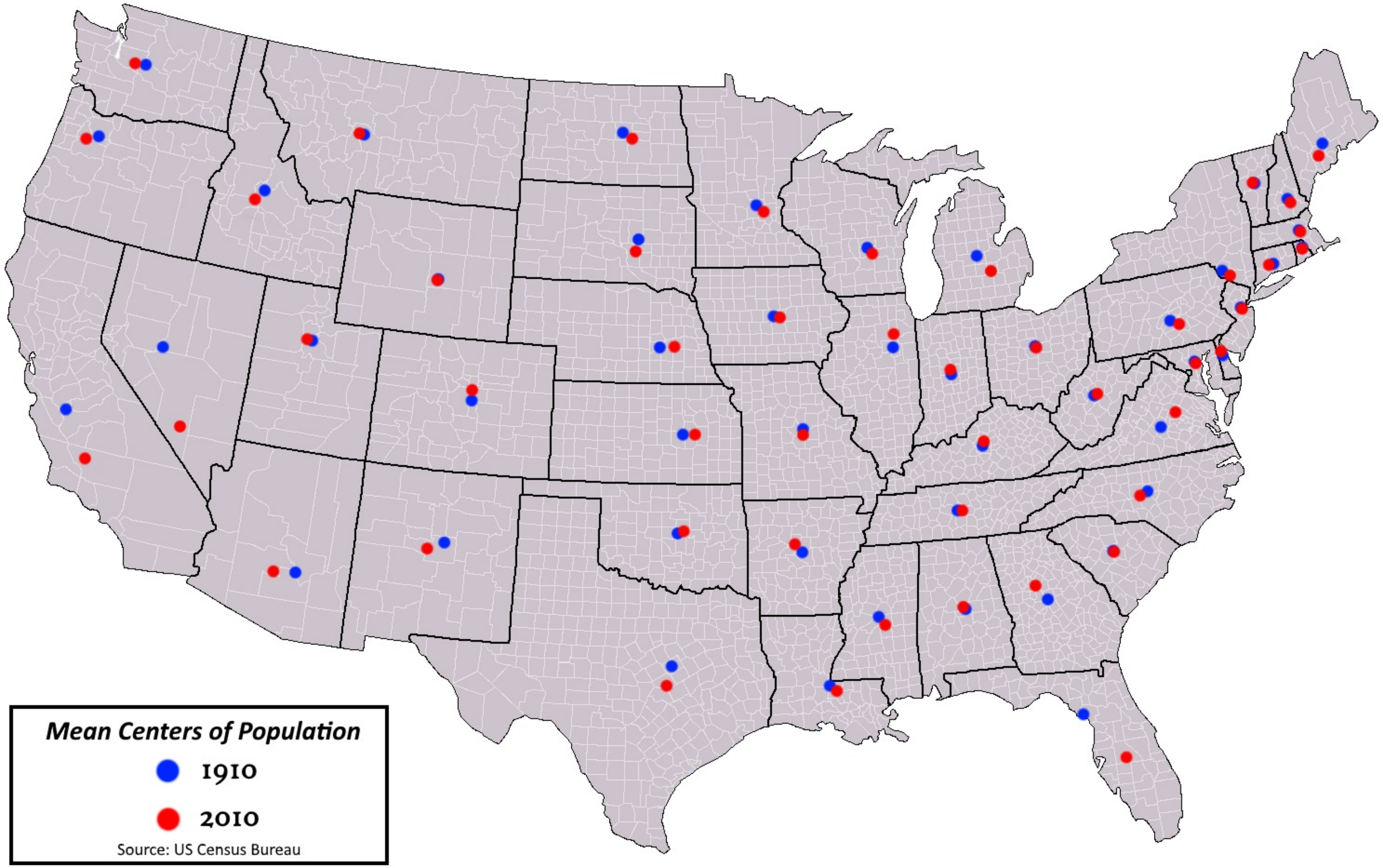
Shape I

- Identify and characterise a shape, e.g. following a process of spatially clustering individual objects or geometries.
- Quantifiable with a compactness ratio or perimeter/area ratio.
- Shape can be important to consider when calculating geometric centroids.

Shape II



Shape II



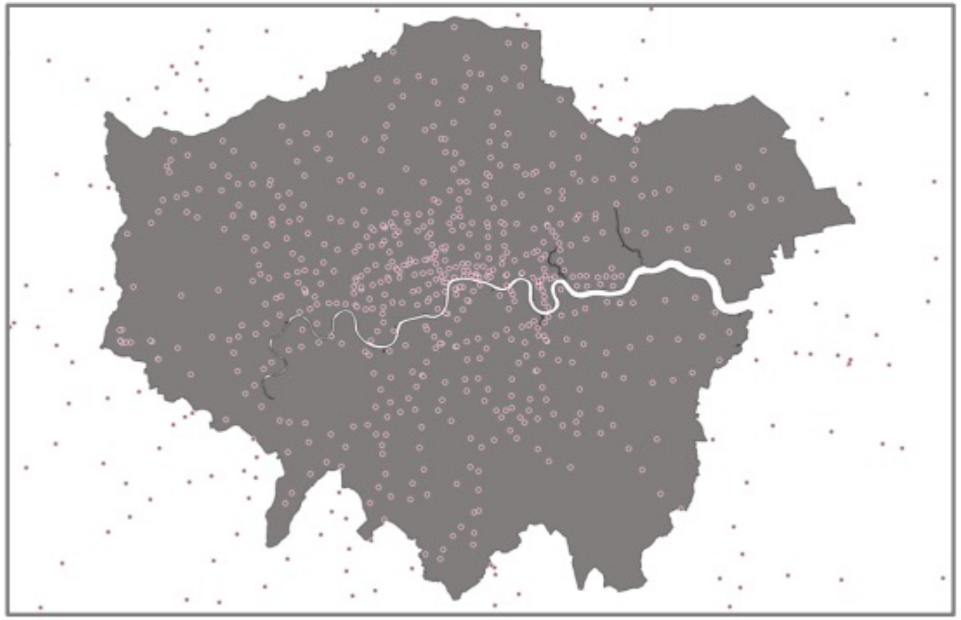
Mean Centers of Population
1910
2010
Source: US Census Bureau

Spatial operations I

- Building blocks of spatial data analysis: select, filter, reduce, remove, join, merge, union different geometries.
- Using spatial properties like distance, area, and shape.

Spatial operations II

?

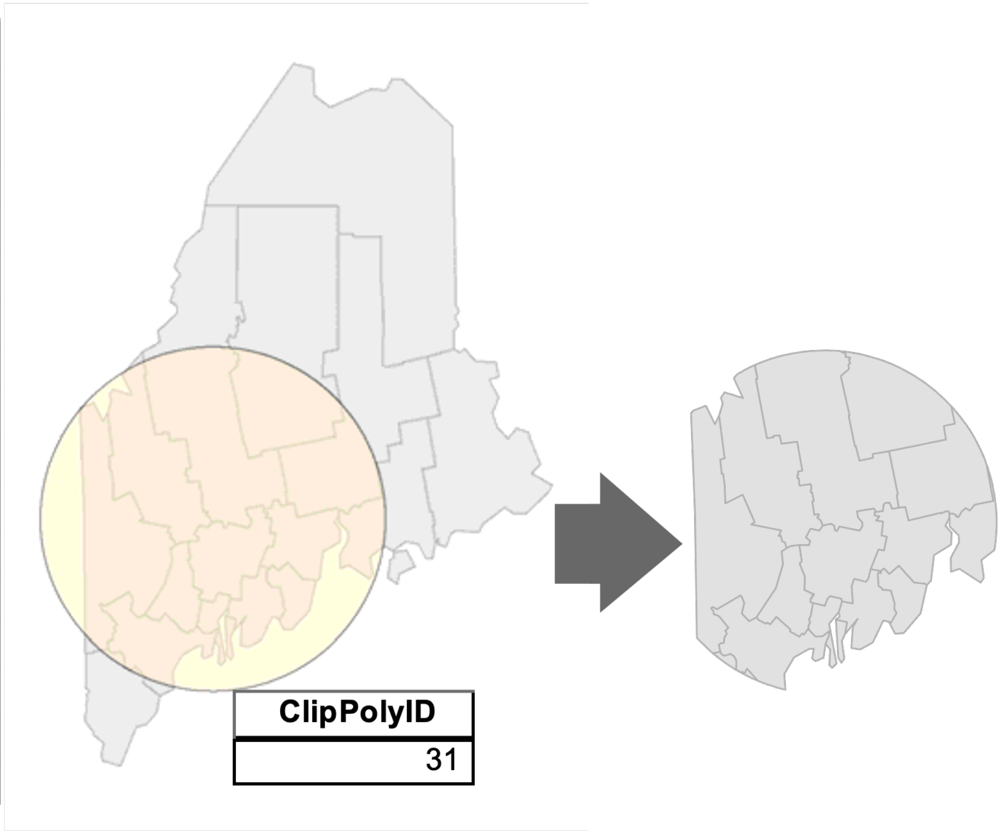


Spatial operations III

Use of topological relationships and geometry functions to take data as an input, analyse the data and then produce output data that is a derivative of the analysis performed on the input data.

Vector operations I - Clip

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



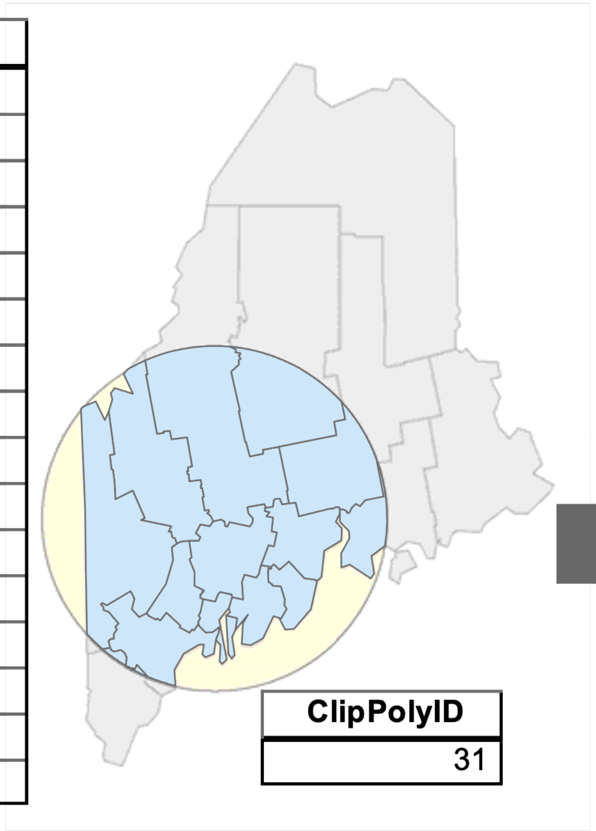
ClipPolyID
31

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

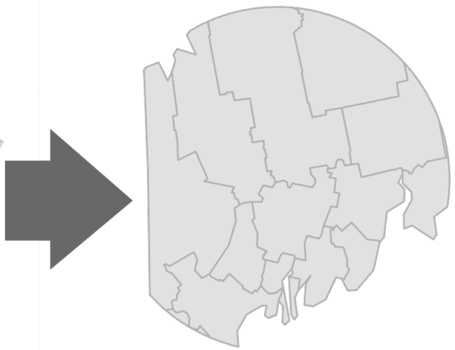
Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Vector operations II - Intersect

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



ClipPolyID
31

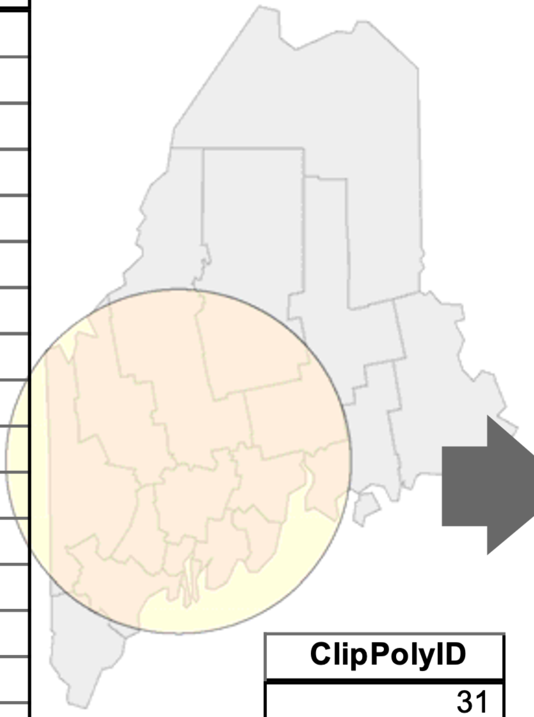


NAME	ClipPolyID
Androscoggin	31
Cumberland	31
Franklin	31
Hancock	31
Kennebec	31
Knox	31
Lincoln	31
Oxford	31
Penobscot	31
Piscataquis	31
Sagadahoc	31
Somerset	31
Waldo	31
York	31

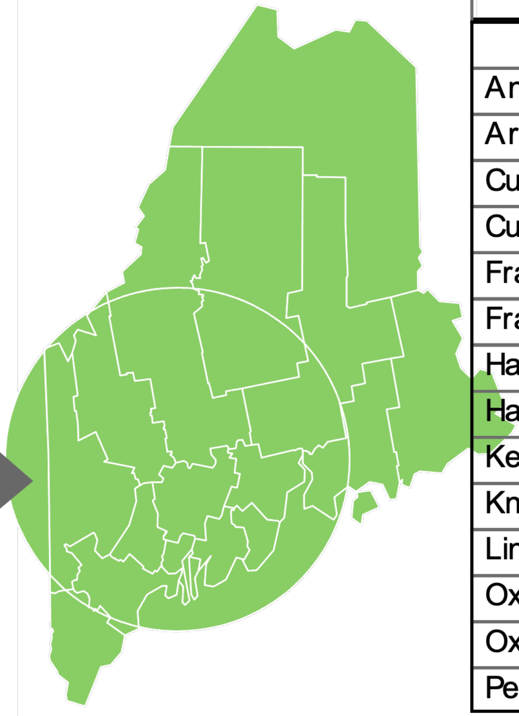
Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Vector operations III - Union

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



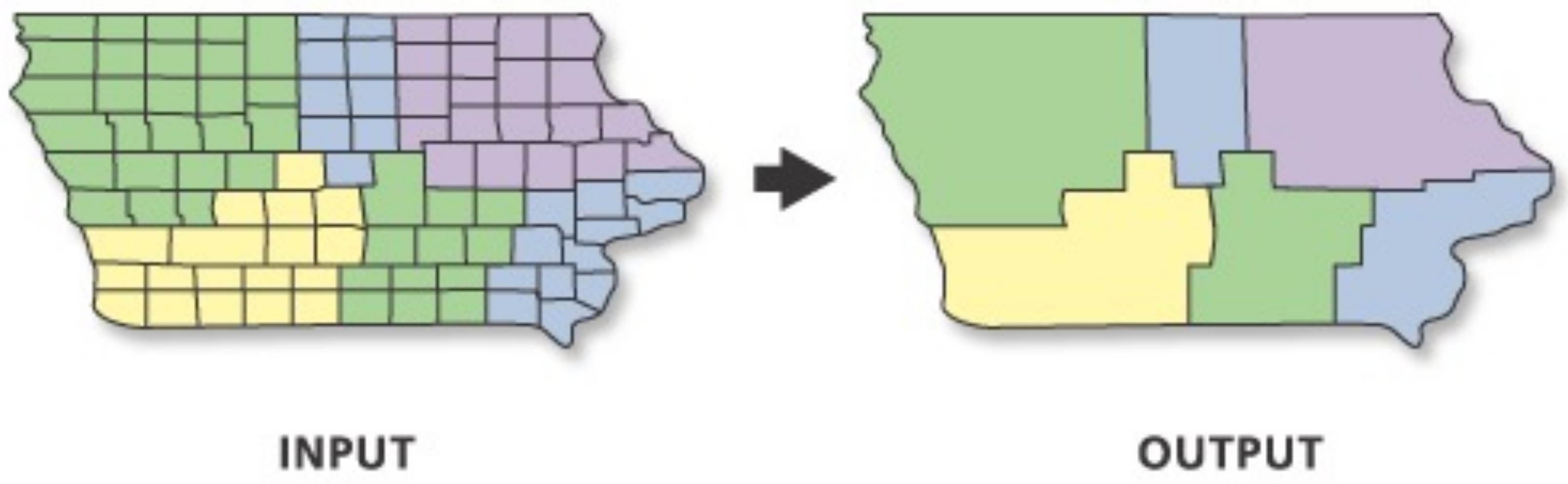
ClipPolyID
31



NAME	ClipPolyID
	31
Androscoggin	31
Aroostook	0
Cumberland	0
Cumberland	31
Franklin	0
Franklin	31
Hancock	0
Hancock	31
Kennebec	31
Knox	31
Lincoln	31
Oxford	0
Oxford	31
Penobscot	0

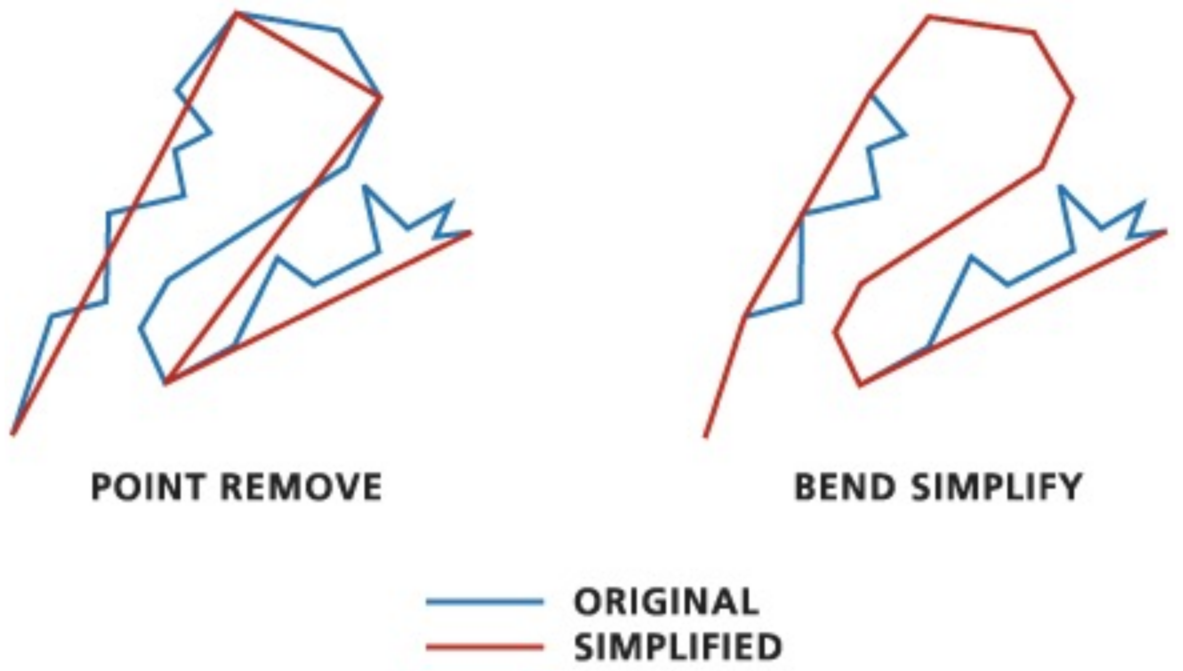
Gimdond, M. 2021. Intro to GIS and Spatial Analysis. [online]
https://mgimond.github.io/Spatial/introGIS.html

Vector operations IV - Dissolve



ESRI. 2021. Dissolve. [online]
https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/dissolve.htm

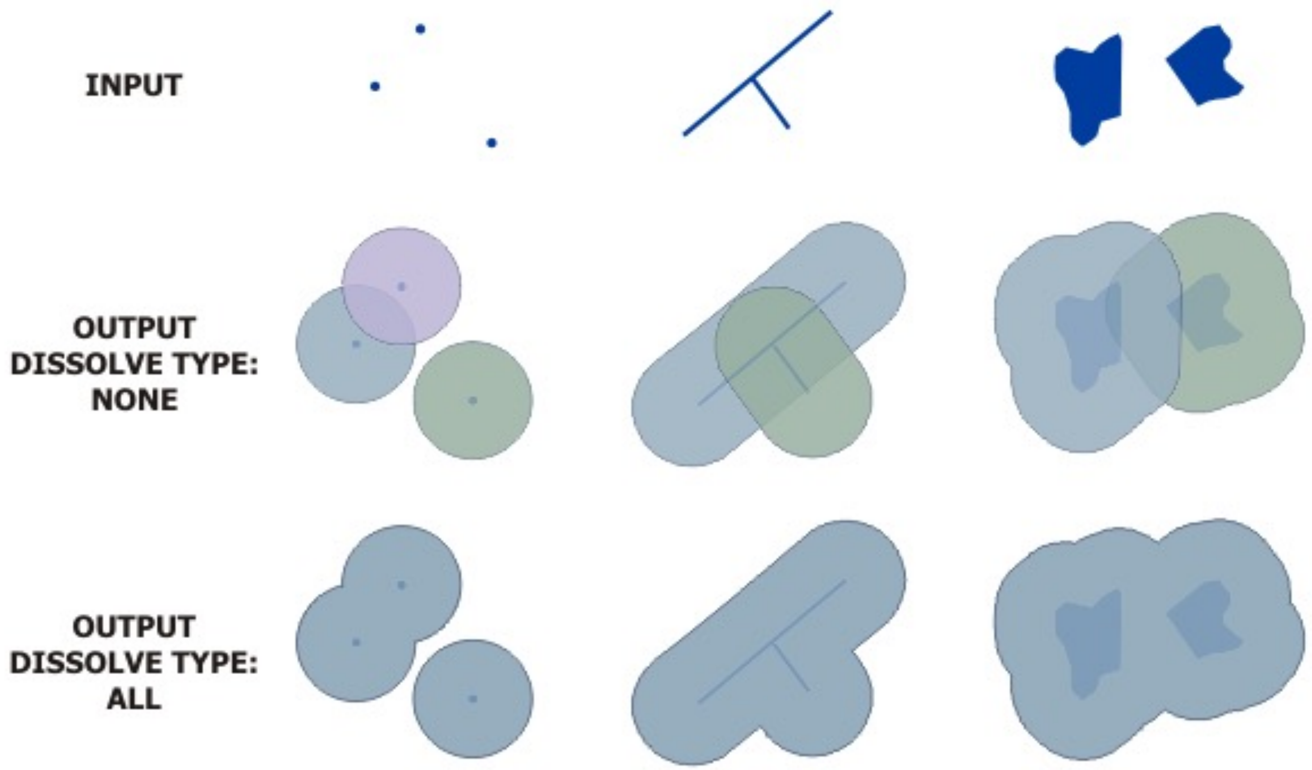
Vector operations V - Simplify



ESRI. 2021. Simplify line. [online]
https://desktop.arcgis.com/en/arcmap/10.3/tools/cartography-toolbox/simplify-line.htm

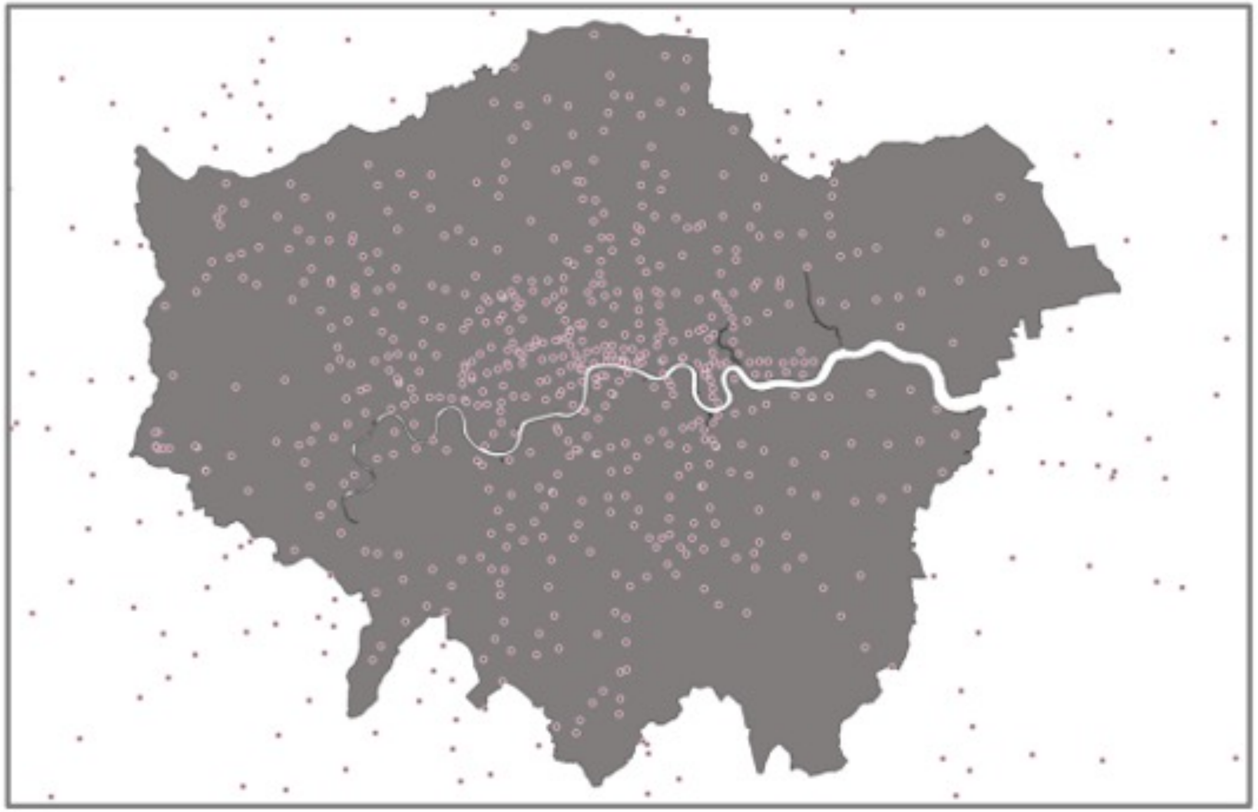
Vector operations V - Simplify

Vector operations VI - Buffer

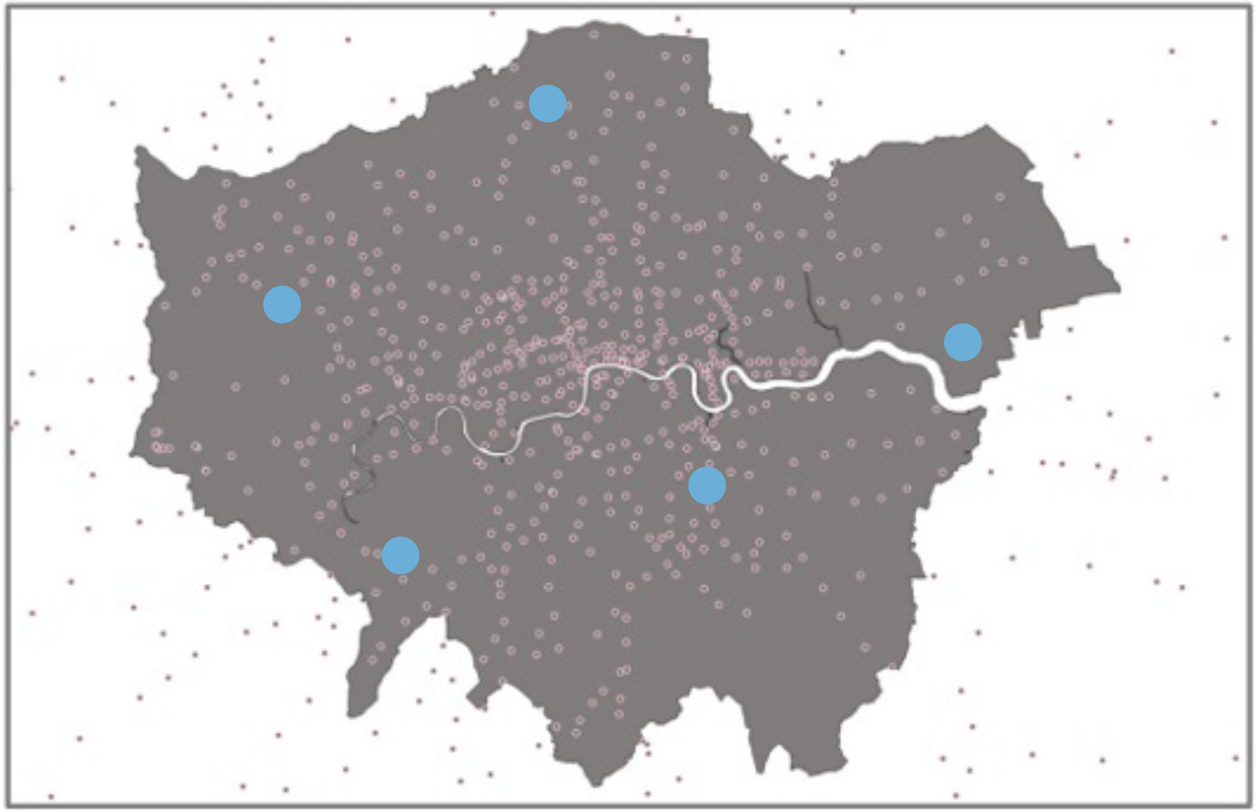


ESRI. 2021. Buffer. [online]
https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/buffer.htm

Vector operations VII – Spatial query



Vector operations VIII – Attribute query



Spatial relationships I

- Spatial relationships define how exteriors, interiors, and boundaries of different geometries interact with one another.
- Known as topological relationship.
- Evaluates adjacency, connectivity, and / or containment.
- Used often with spatial data processing to join data sets together.
- Need to be sure of how you want to capture your spatial relationship between your two features when using a spatial join.

Spatial relationships II

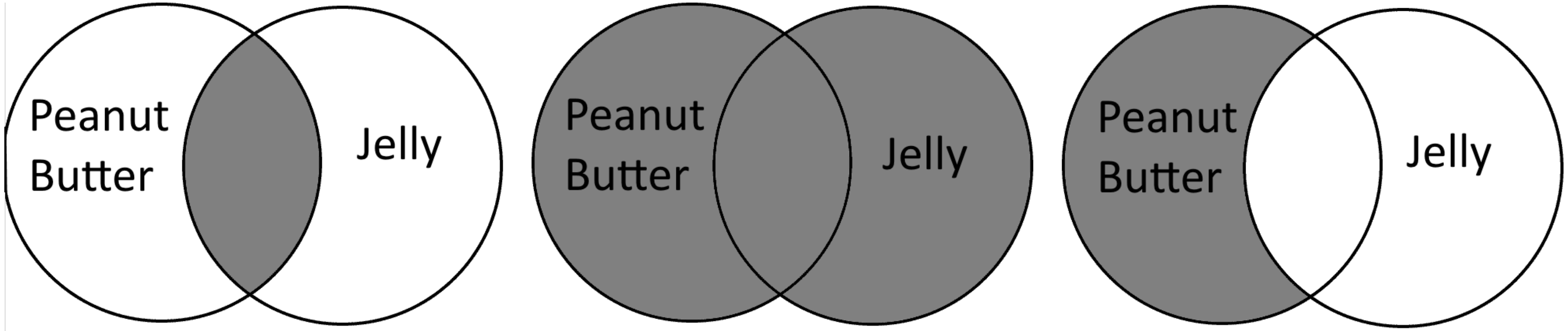
Equals A is the same as B	
Touches A touches B	
Overlaps A and B have multiple points in common	
Contains A contains B	
Disjoint A shares nothing with B	
Covers A covers B (or vice versa)	
Crosses A and B have at least one point in common	

Spatial relationships III

- There is some specific maths behind calculating these topological relationships between spatial objects, which is then enacted by the computer via code.
- *"Does an object contains or overlap another?"*
  1 Establish exterior, interior and boundaries of the geometries of each the object.
  2 Calculate the number of times these three properties intersect with one another.
  3 Follow the requirements of the function to understand if it is TRUE or FALSE.

Spatial relationships IV

Truth evaluation using Boolean operators



AND

Using AND, this search would only retrieve results with Peanut Butter and Jelly.

OR

Using OR, this search would retrieve results with peanut butter, with jelly, and with both.

NOT

Using NOT, this search would retrieve results with peanut butter, and exclude those with jelly or PB with jelly.

Inside your machine I

GDAL/OGR or GEOS do this for you – these libraries contain the 'definitions' of all these type of geometric operations. PROJ deals with projections. In short:

- GDAL/OGR is a translator library (C++) for raster and vector geospatial data formats.
- GEOS is a (C++) port of the JTS Topology Suite (JTS).
- PROJ is a generic coordinate transformation software that transforms geospatial coordinates from one coordinate reference system (CRS) to another

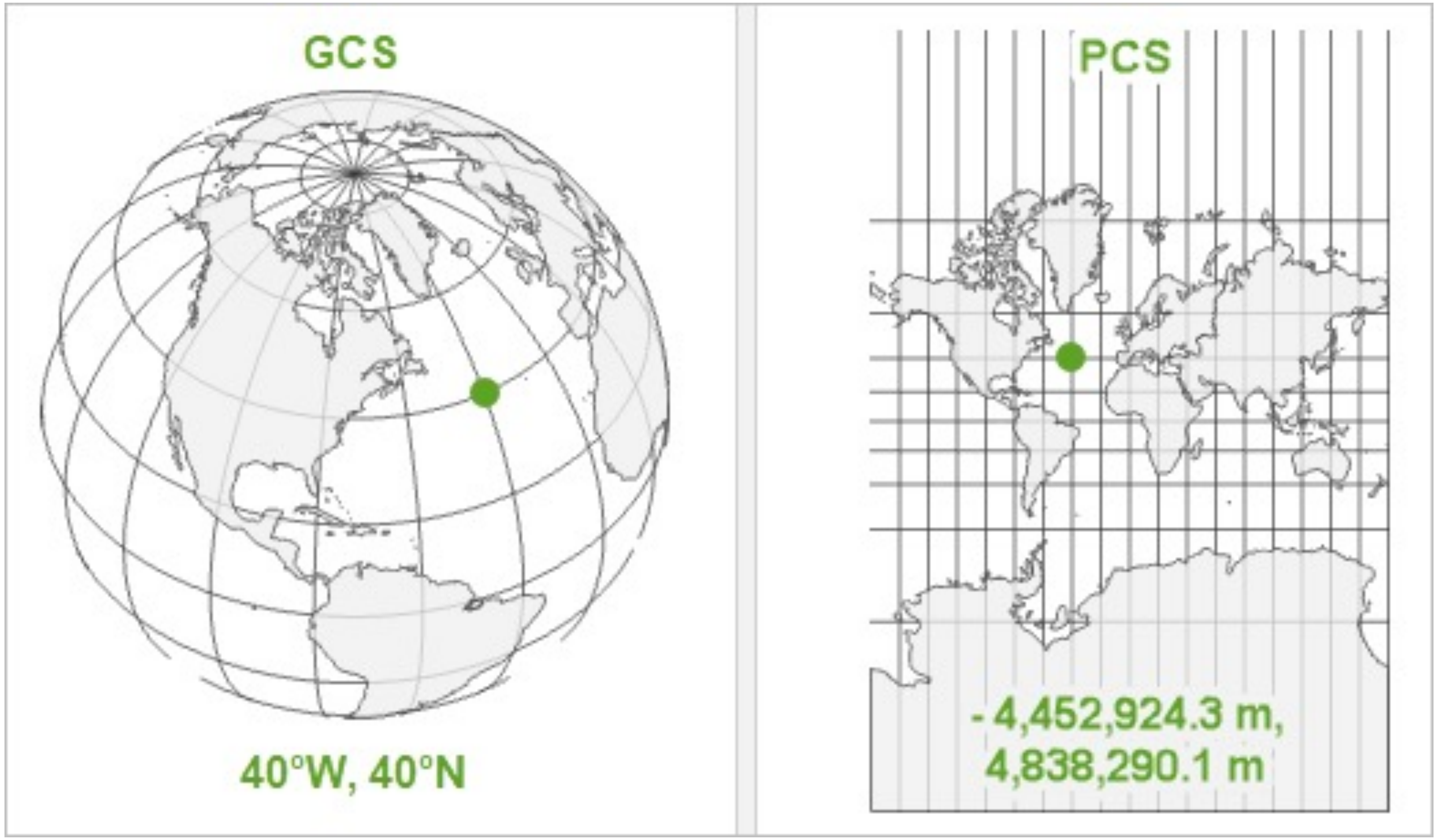
Inside your machine II

- R / Python / QGIS in fact ‘call’ to the GDAL/OGR and PROJ libraries (‘bindings’).
- For example: `pandas` depends on `Shapely`/`PyGEOS` (interface to GEOS), `Fiona` (interface to GDAL), and `pyproj` (interface to PROJ).
- Similar in the R environment: `sf` interfaces to GEOS, GDAL and PROJ.

Coordinate systems and projections I

- 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.
- 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).

Coordinate systems and projections II



Coordinate systems and projections III

- 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 plan 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).
- A GCS is defined by an ellipsoid, geoid and datum.

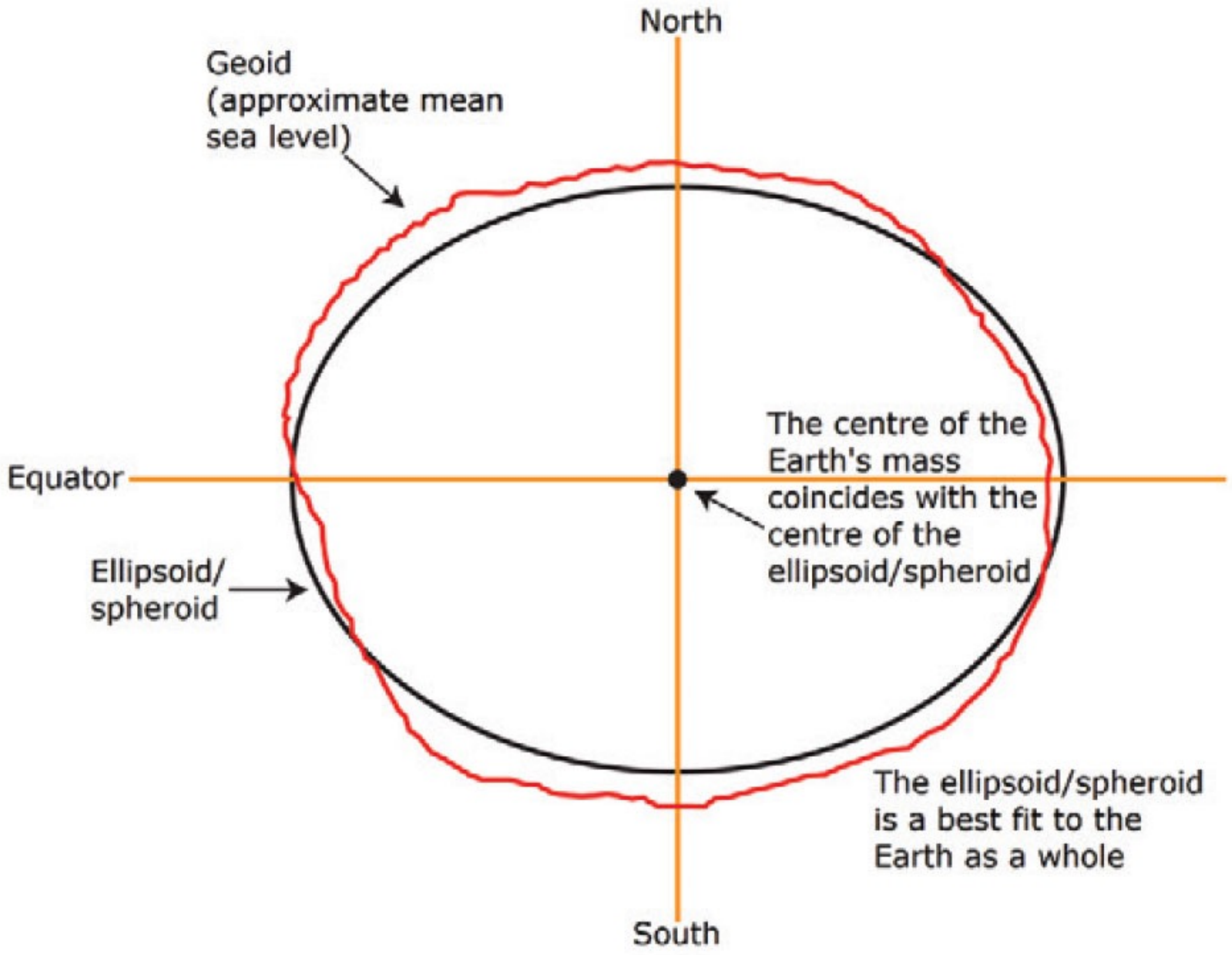
Coordinate systems and projections IV

- 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 – but not perfect.
- We need to reconcile a mathematical model of the earth’s shape with the geoid.

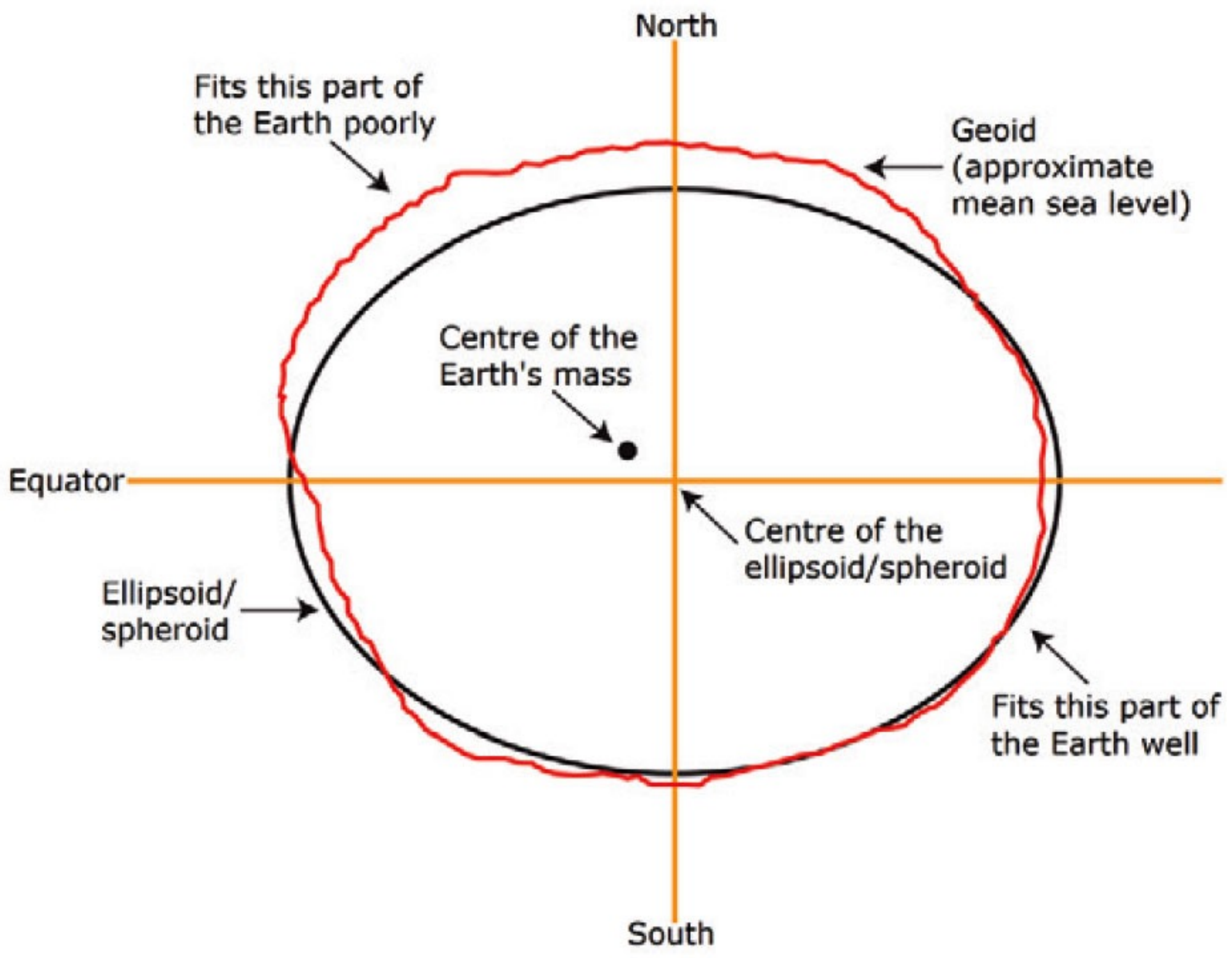
Coordinate systems and projections V

- The solution is to align the geoid with the ellipsoid (or sphere) representation of the earth and to map the earth's surface features onto this ellipsoid/sphere.
- 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.
- A GCS is thus defined by the ellipsoid (or sphere) model and by the way this ellipsoid is aligned with the geoid (defining the datum).

Coordinate systems and projections VI



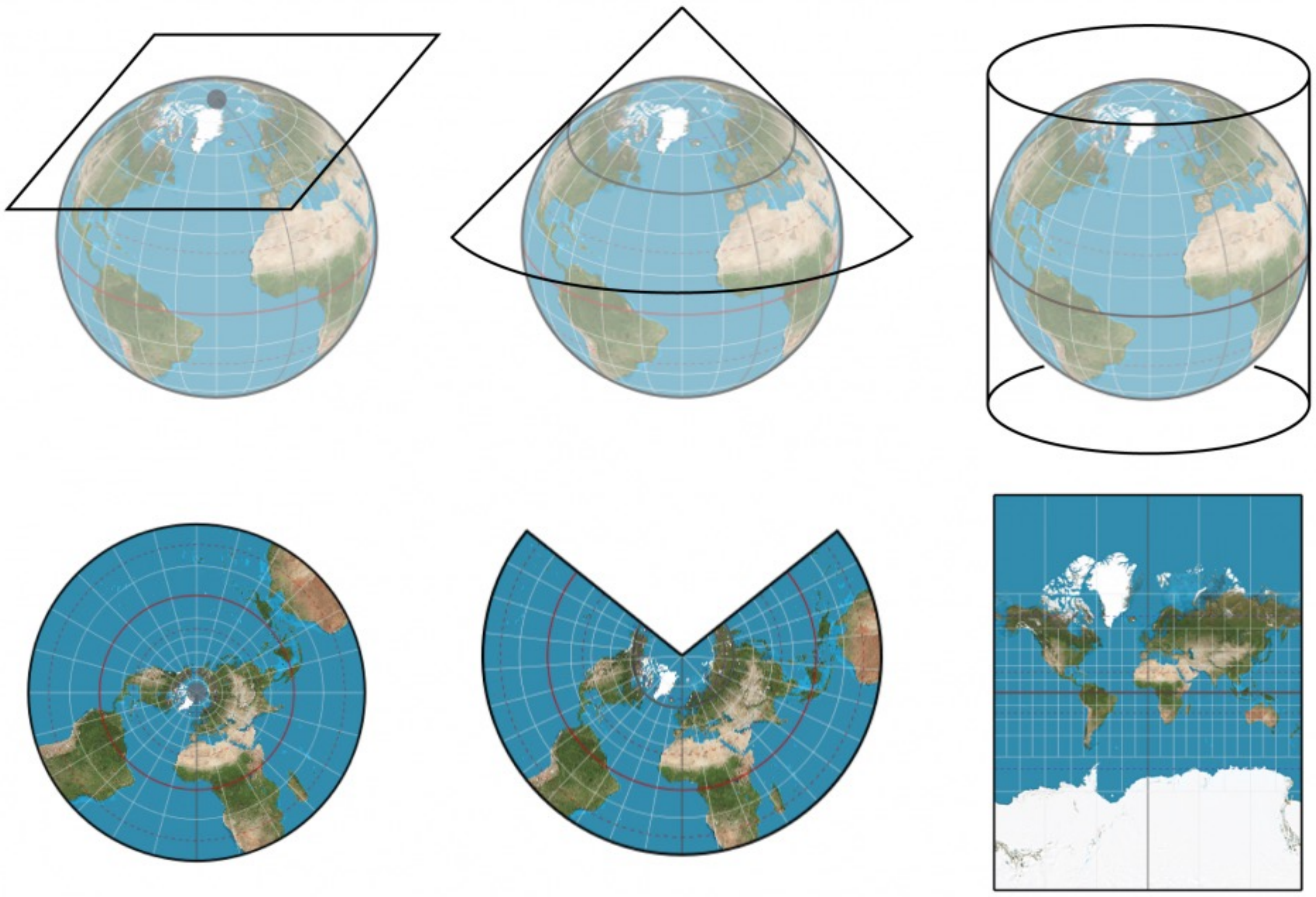
Coordinate systems and projections VII



Coordinate systems and projections VIII

- 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, cylindrical and conical.

Coordinate systems and projections IX



Coordinate systems and projections X

- 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.

Coordinate systems and projections XI

- A very popular projection to represent the world on a flat surface is the Mercator projection.
- However: it exaggerates the size of countries near the poles, while downplaying the size of those near the equator.
- Very nice visualisation of the effects: https://thetruesize.com/

Coordinate systems and projections XII

- 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.
- 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.

Coordinate systems and projections XIII

- Details 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.
- The data set was originally created by European Petroleum Survey Group (EPSG) in 1985 and was made public in 1993.
- EPSG: 4326 – WGS84 (often used for GPS coordinates)
- EPSG: 27000 – British National Grid
- EPSG: 3857 – Web Mercator

Coordinate systems and projections XIV

`.prj`

```
PROJCS["British_National_Grid",GEOGCS["GCS_OSGB_1936",DATUM["D_OS
GB_1936",SPHEROID["Airy_1830",6377563.396,299.3249646]],PRIMEM["G
reenwich",0.0],UNIT["Degree",0.0174532925199433]],PROJECTION["Tra
nsverse_Mercator"],PARAMETER["False_Easting",400000.0],PARAMETER[
"False_Northing",-100000.0],PARAMETER["Central_Meridian",-
2.0],PARAMETER["Scale_Factor",0.9996012717],PARAMETER["Latitude_O
f_Origin",49.0],UNIT["Meter",1.0]]
```

Conclusion

- Spatial data operations are the main building blocks to spatial data processing and at the heart of spatial analysis.
- For every analysis: try to understand what tool you are using, why, and any technical requirements or issues you may need to address.

Questions

Justin van Dijk

j.t.vandijk@ucl.ac.uk