Representing the World

From infinitely complex reality to models and representations

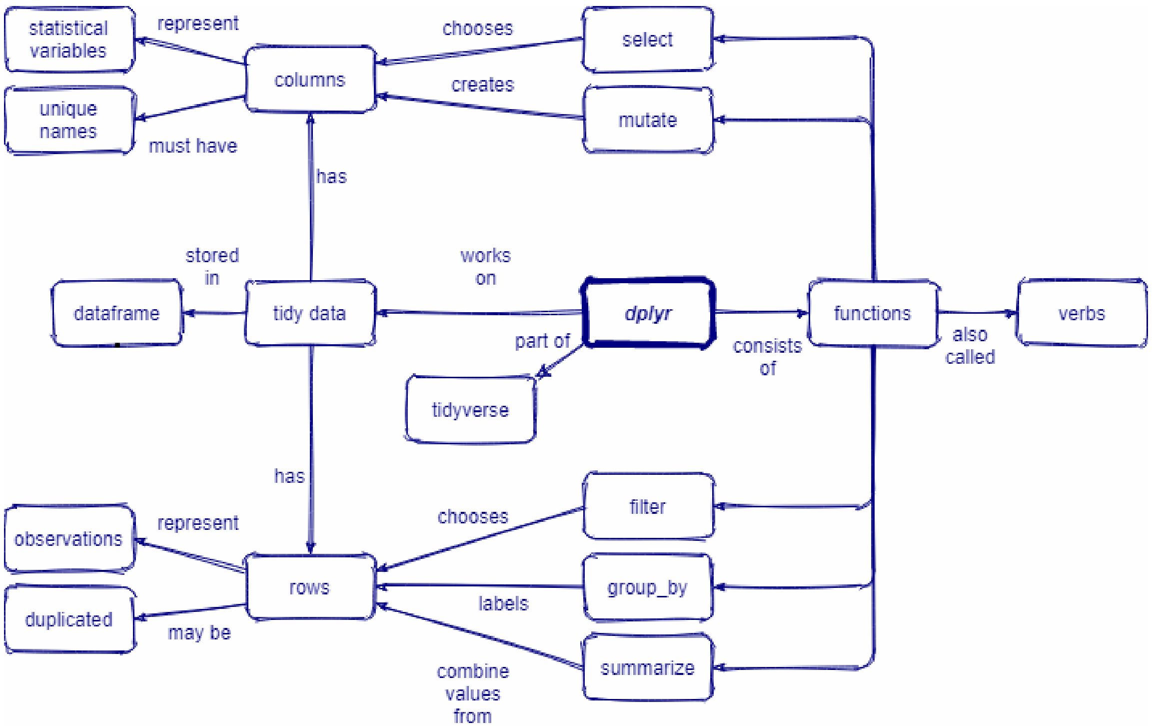
Dr Ruth Neville

(Slides adapted from Claire Dooley, Adam Dennett & Andy MacLachlan's previous years' lectures)

Review of last week

- Familiar with:
  - filter()
  - select()
  - mutate()
  - group_by()
  - summarise()
  - clean_names()
  - str_detect()

https://r4ds.had.co.nz/strings.html



Review of last week

- Piping structure
- Alternative ways to write code, e.g. select col

```r

df %>%
  dplyr::select(column1)

df$column1

df[["column1"]]

```

Review of last week

- Piping structure
- One way to write code
- %>% = ‘and then’
- Takes result of one step and passes it on to the next

Hide

```
library(janitor)

LondonBoroughs <- LondonBoroughs %>%
  dplyr::rename(Borough=`Ward name`)%>%
  clean_names()
```

Take data from ‘LondonBoroughs’

And then.. rename column ‘Ward Name’ to Borough

And then.. standardise column names

Review of last week

- When we need to use a '.' when using the %>% operator and when we can do without it.....
  - Use '.' where there are extra arguments… **you don't have to use it** though…and you can use it where there are no arguments…
  - Infact you don't need the comma either!

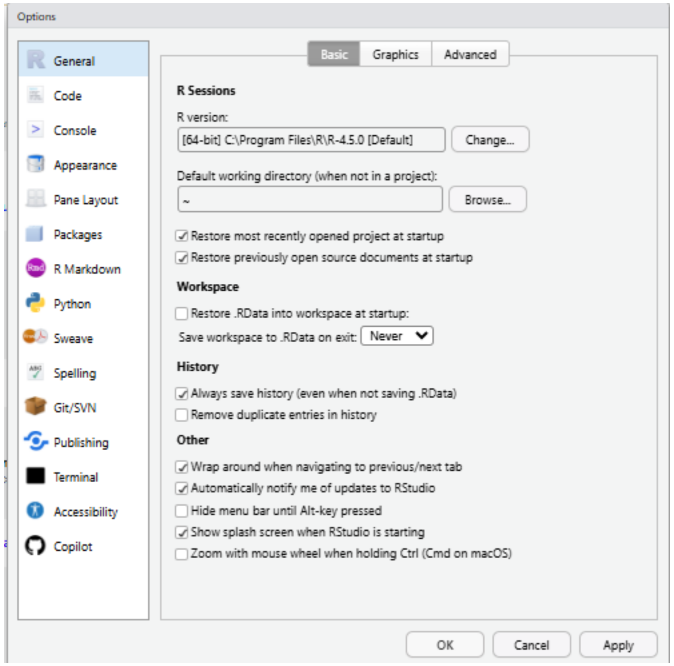
Hide

```
LondonBoroughs <- LondonBoroughs %>%
  #here the ., means all data
  clean_names(., case="big_camel")
```

. Is LondonBoroughs

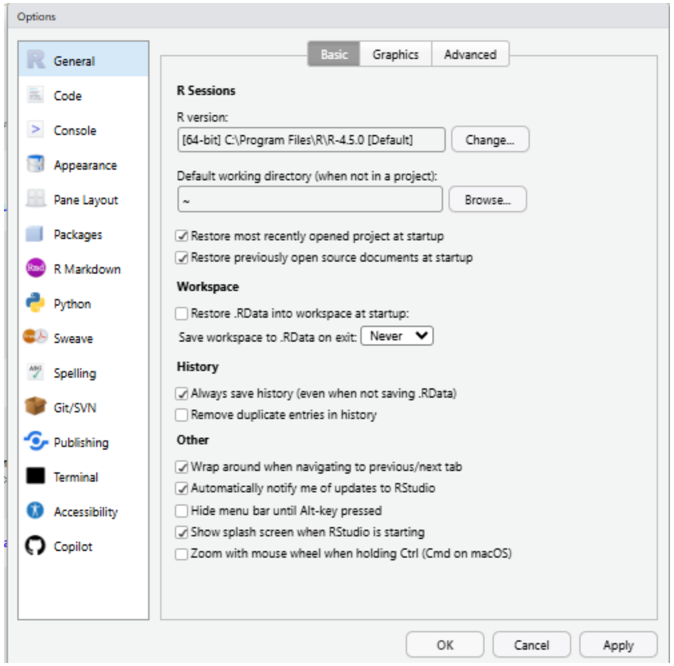
Review of last week

- Keeping environment tidy
- In RStudio Tools -> Global Options -> General



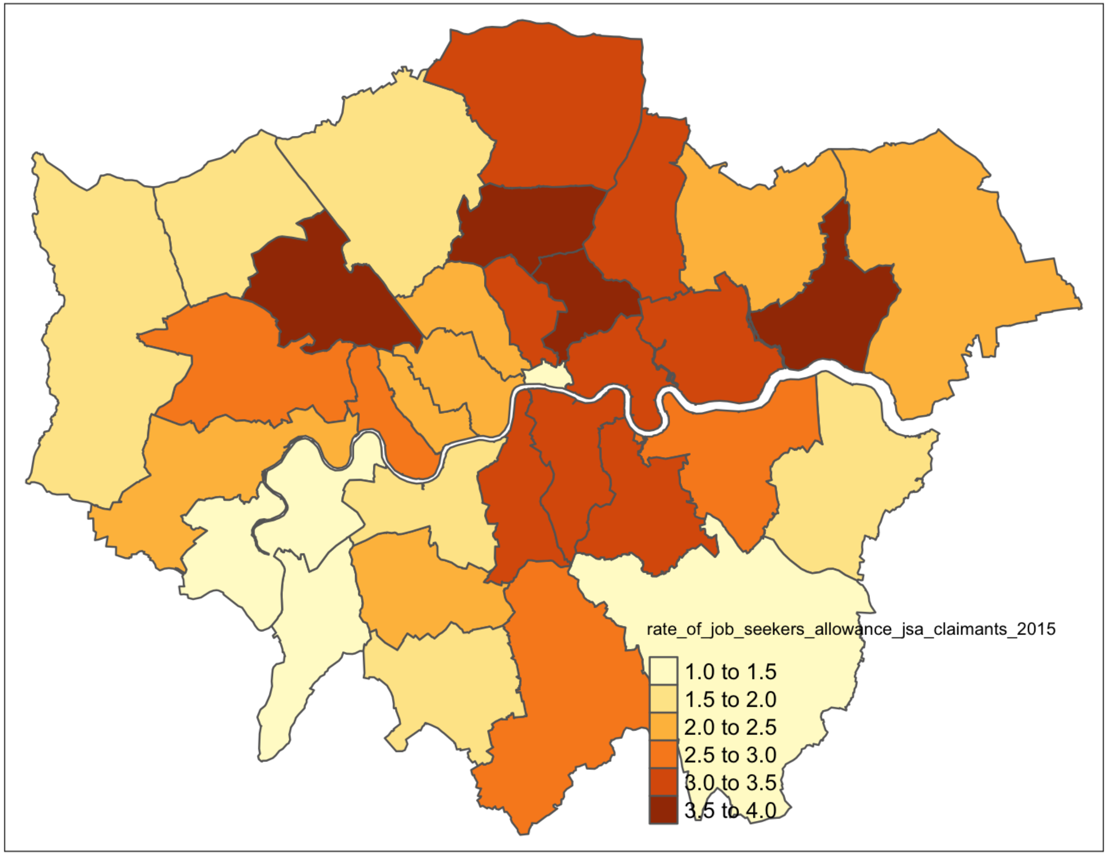
Review of last week

- Keeping environment tidy
- In RStudio Tools -> Global Options -> General



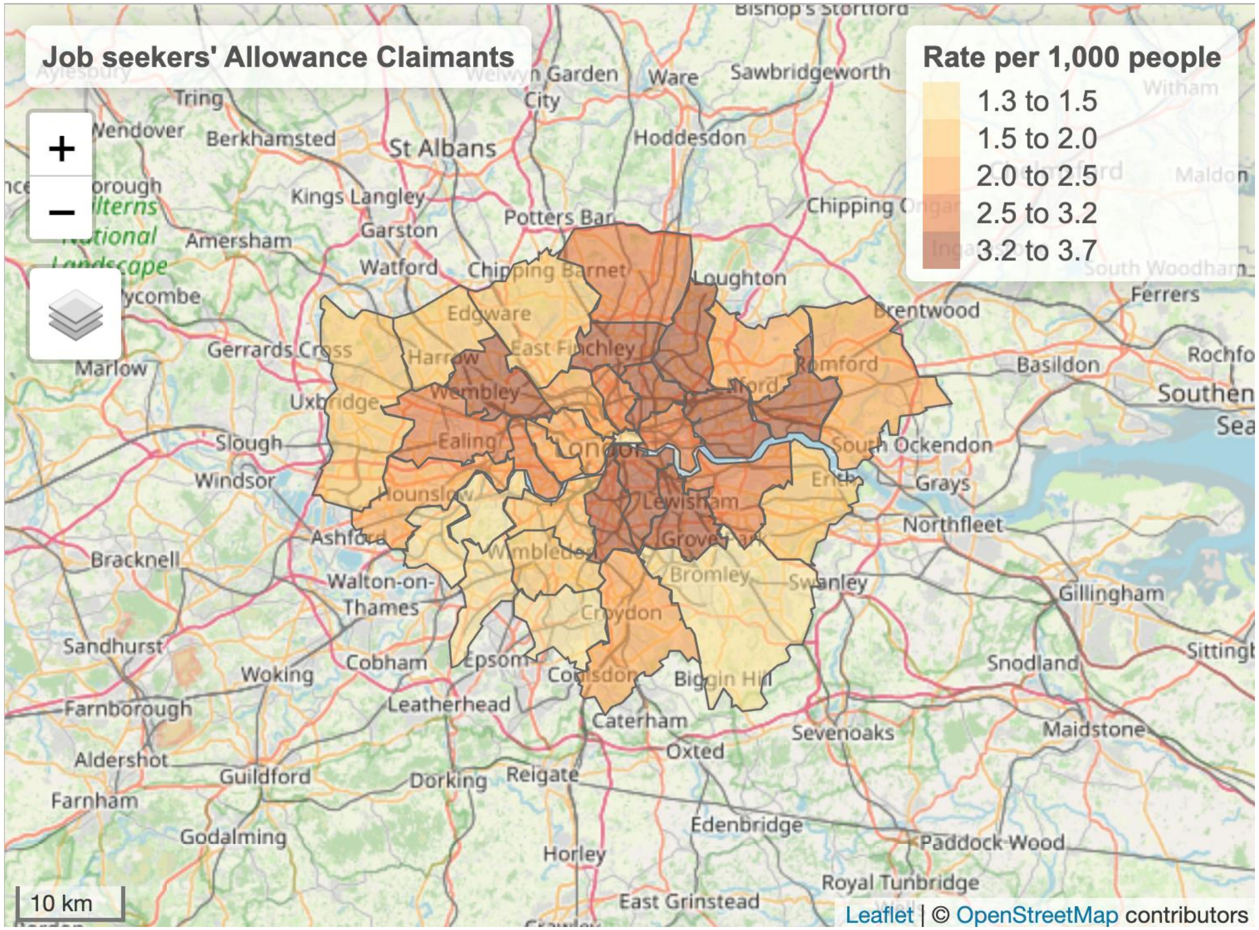
Review of last week

Mapping



Review of last week

Mapping



Review of last week

Reminders

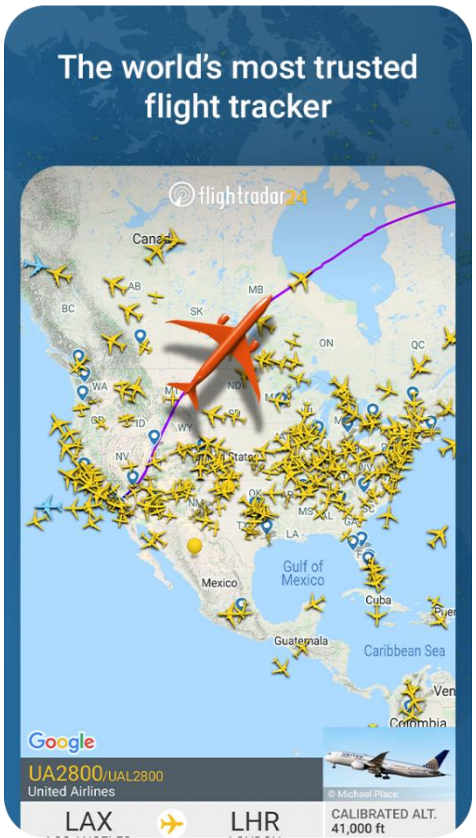
- Check your packages are loaded
- Check your column names – regularly check your data
- Comment your code
- Don’t need to memorise all the functions and packages
- It is very normal to get stuck, often for hours, then you realise it’s something simple (ish)
- Try to think through each step
- Read the error messages
- Force yourself to use R to wrangle – not excel!
- Google is your friend!
- Engage with the slack channel

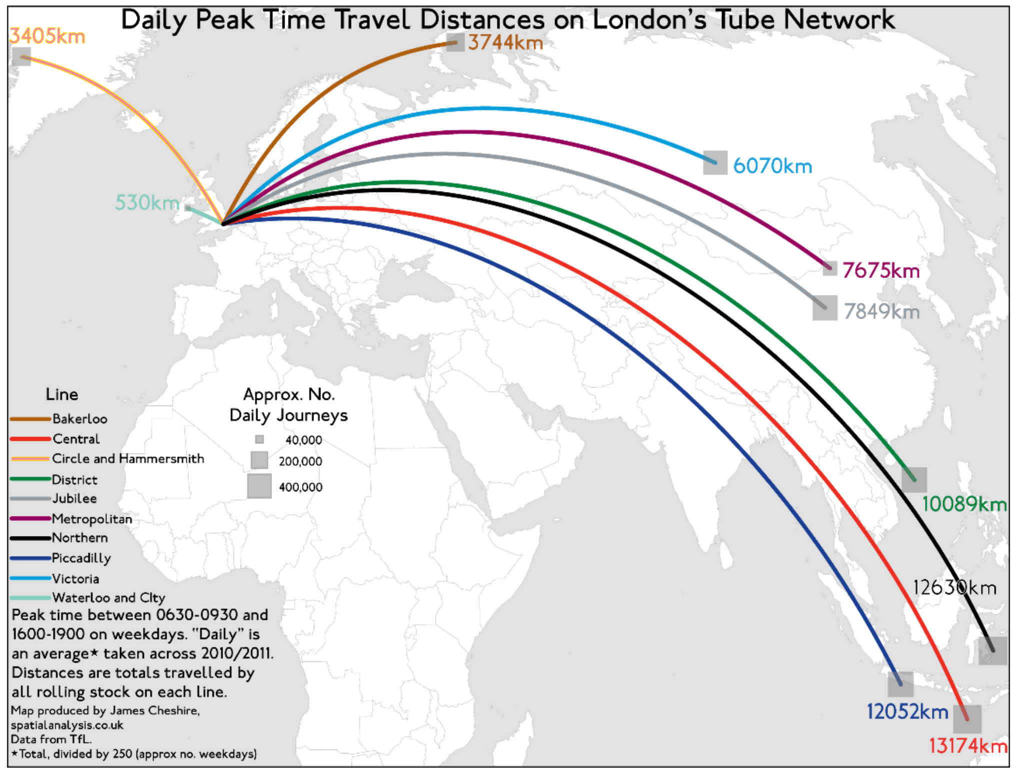
Outline / Learning outcomes

- Explain representation
- Describe discrete objects and continuous fields
- Evaluate raster and vector data
- Summarise projections

Representations - Sensing the World

- Aside from real life, everything we see is a representation …





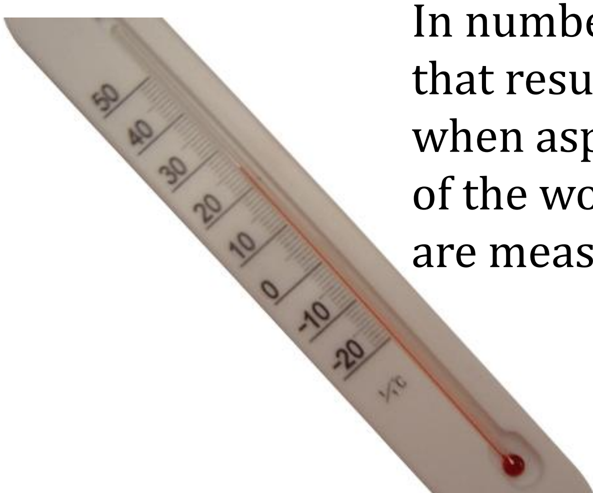
James Cheshire, UCL geog

Representations - Sensing the World

- When we don’t sense our world directly, additional knowledge needs to come from elsewhere:
  - books,
  - the media,
  - movies,
  - maps,
  - images,
  - and other information sources
- This indirect knowledge is necessarily a representation (simplification) of true reality

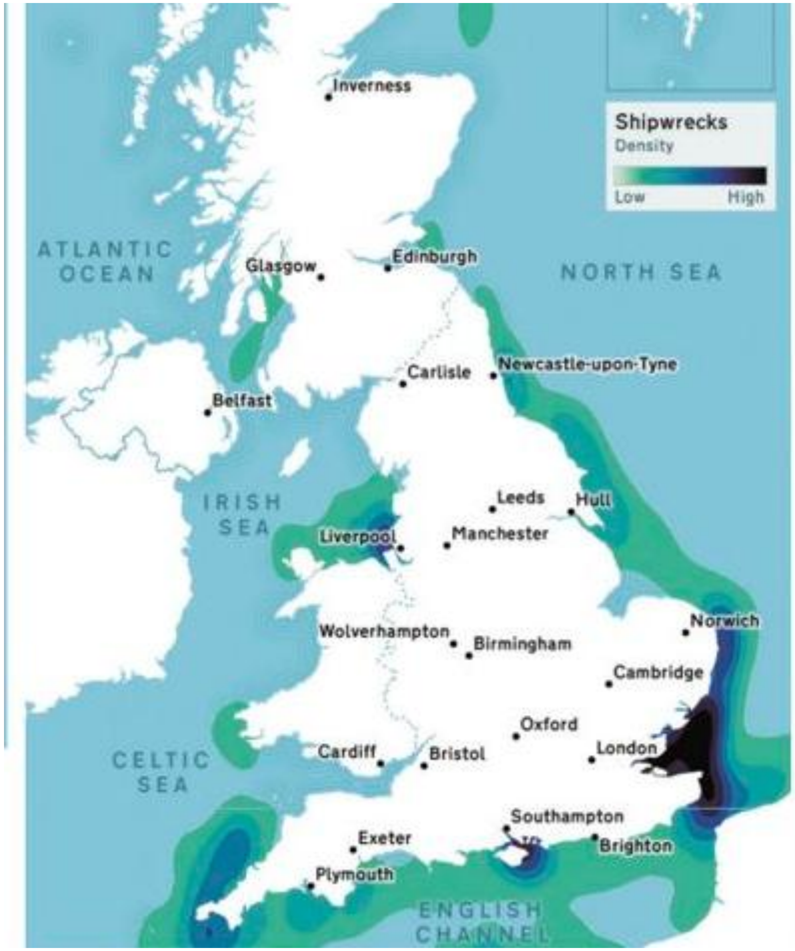
Representations Occur:

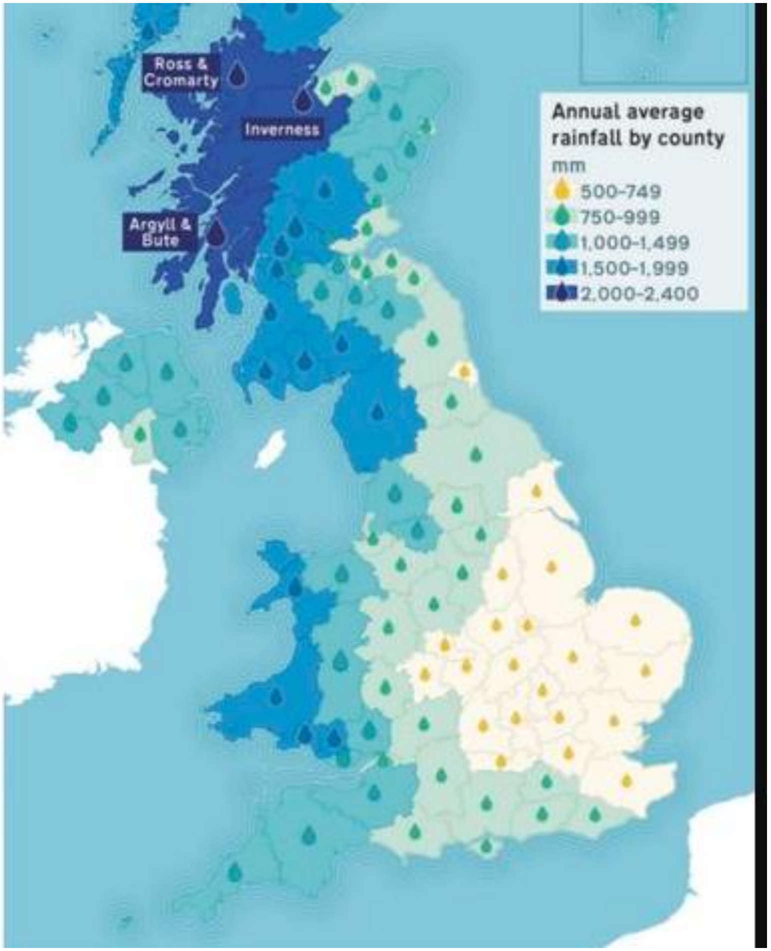
In photographs, which are two-dimensional models of light received by the camera



In numbers that result when aspects of the world are measured

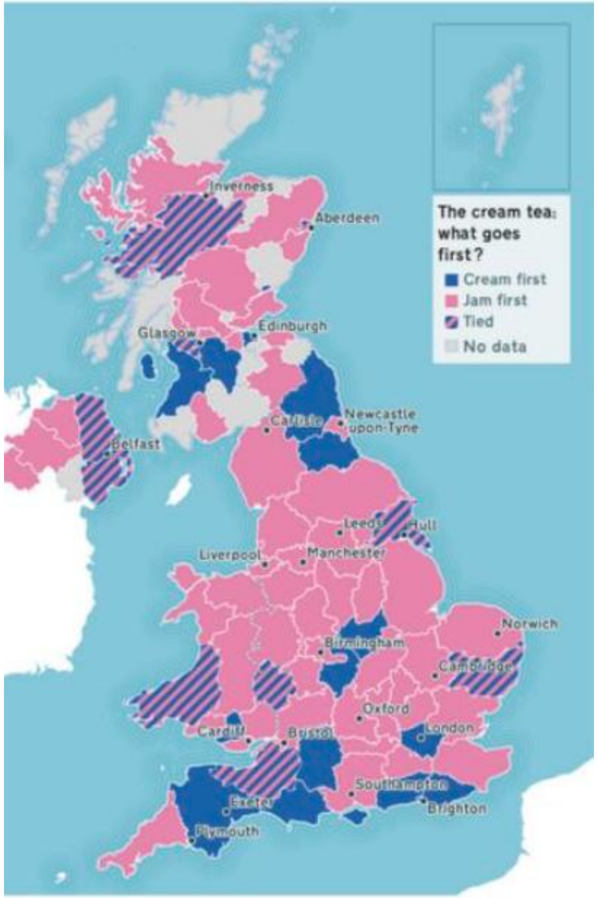
In written text, when information is expressed in words

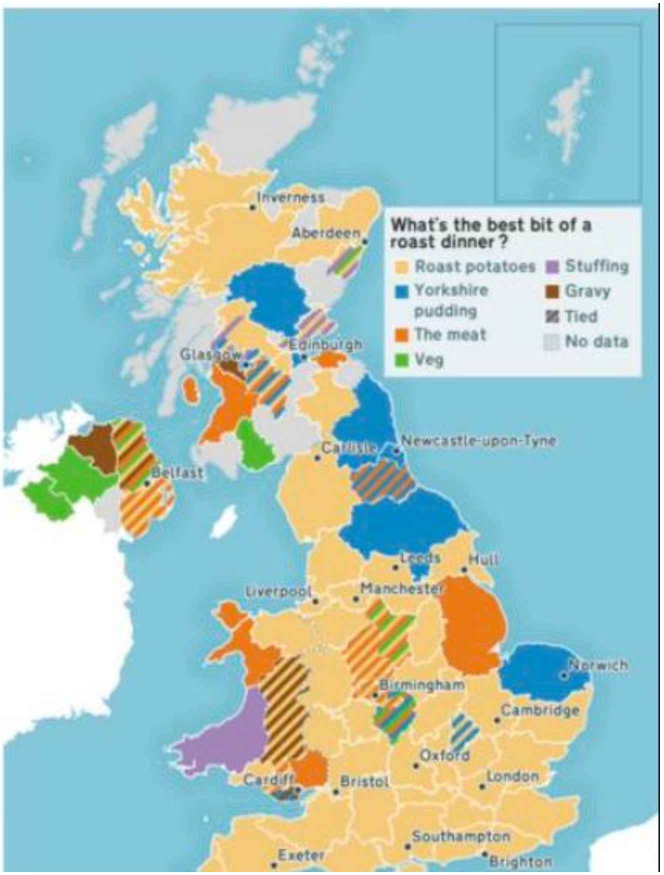




https://www.linkedin.com/posts/helenmckenzie003_my-book-utterly-british-maps-is-live-activity-7376577437263998976-0-Fo?utm_source=social_share_send&utm_medium=member_desktop_web&rcm=ACoAABuSjkgBjpMu5zZSSTZZyoyp7wTf--GKN-8

https://harpercollins.co.uk/products/utterly-british-maps-national-trust-helen-mckenzie?variant=55073569177979



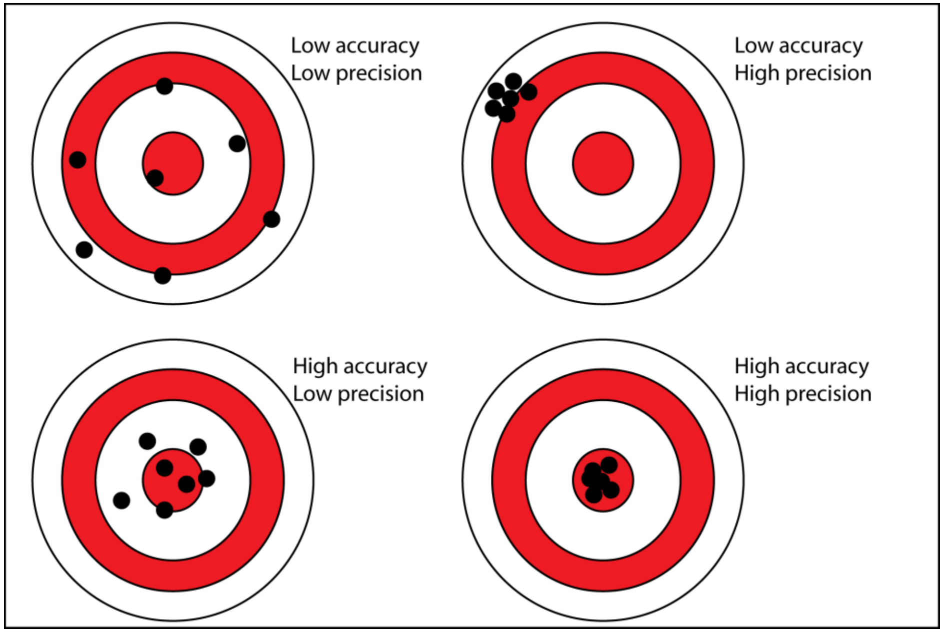


https://www.linkedin.com/posts/helenmckenzie003_my-book-utterly-british-maps-is-live-activity-7376577437263998976-0-Fo?utm_source=social_share_send&utm_medium=member_desktop_web&rcm=ACoAABuSjkgBjpMu5zZSSTZZyoyp7wTf--GKN-8

https://harpercollins.co.uk/products/utterly-british-maps-national-trust-helen-mckenzie?variant=55073569177979

Accuracy of Representations

- Representations can rarely be perfect
  - Details can be irrelevant
  - or too expensive and voluminous to record
- It's **important to know what is missing** in a representation
  - Representations can leave us **uncertain** about the real world
  - Accuracy = close to value
  - Precision = to each other



Digital Representation

Digital spatial data standards

- Help information systems share data
  - Particularly on the Internet
- In Geographic Information Systems, a variety of standards
  - GeoTIFF
  - GeoRSS
  - GeoJSON
  - GeoPackage
- In GIS, the Open Geospatial Consortium
  (OGC) sets the standards for geospatial
  Data – free and open.

Digital spatial data standards

Better Quality Data – Characteristics

Accuracy

Accurate enough for intended purposes:
- Balance with use, cost, effort, timeliness
- Capture close to point of activity
- Make accuracy compromises clear

Validity

Compliance with requirements:
- Application of definitions
- Consistency over time
- Consistency with others

Reliability

Collection processes consistent:
- ...over time
- ... for multiple collection points
- ... between collection systems

Timeliness

To Influence Decisions:
- Capture quickly after the event
- Available quickly enough
- Available frequently enough

Relevance

Data relevant to intended purpose:
- Periodic review of requirements
- Quality assurance and feedback process
- Use carefully for other purposes

Completeness

Monitor quality to match data needs:
- Missing data
- Invalid data
- Incomplete data

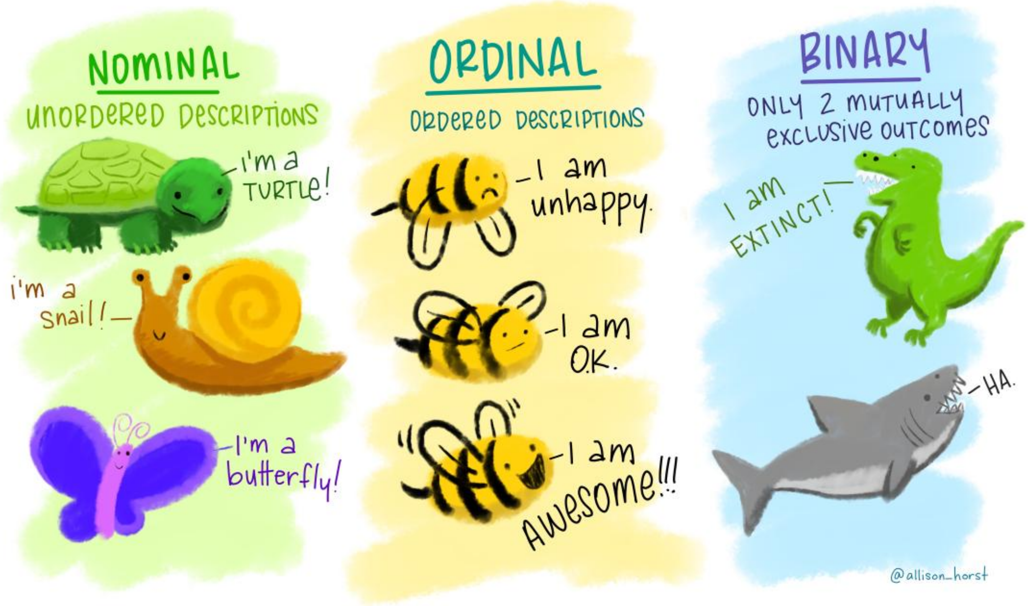
Heuvelink, G.B.M., 1999. Propagation of error in spatial modelling with GIS.

Storing Digital Geographic Data

- Geographic information links a place, and often a time, with some property of that place (and time)
  - "The temperature at 34 N, 120 W at noon local time on 28/9/05 was 18 Celsius"
- The potential number of properties is vast
  - In GIS we term them ***attributes***
  - Attributes can be physical, social, economic, demographic, environmental, etc

Storing Digital Geographic Data

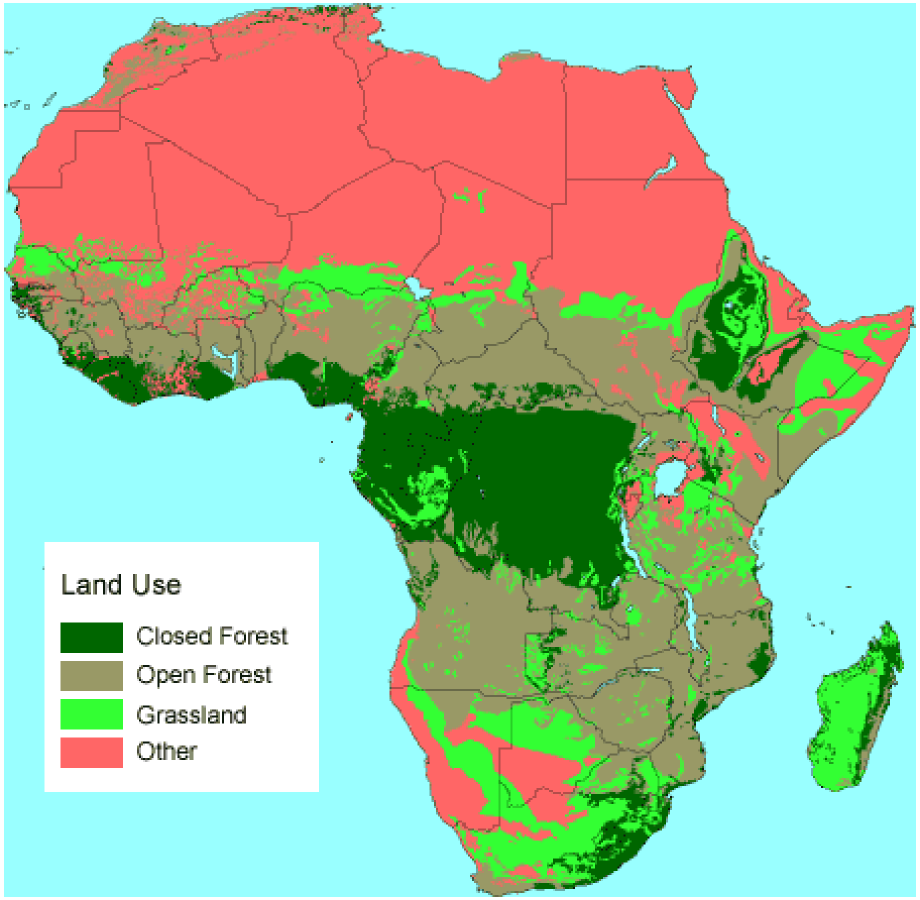
- Attributes (data) can be classified by their statistical data **type**…(binary, nominal, continuous…)
- This is different to the column data type (e.g. numeric, float, character).



In this case we are thinking about data representation not storage

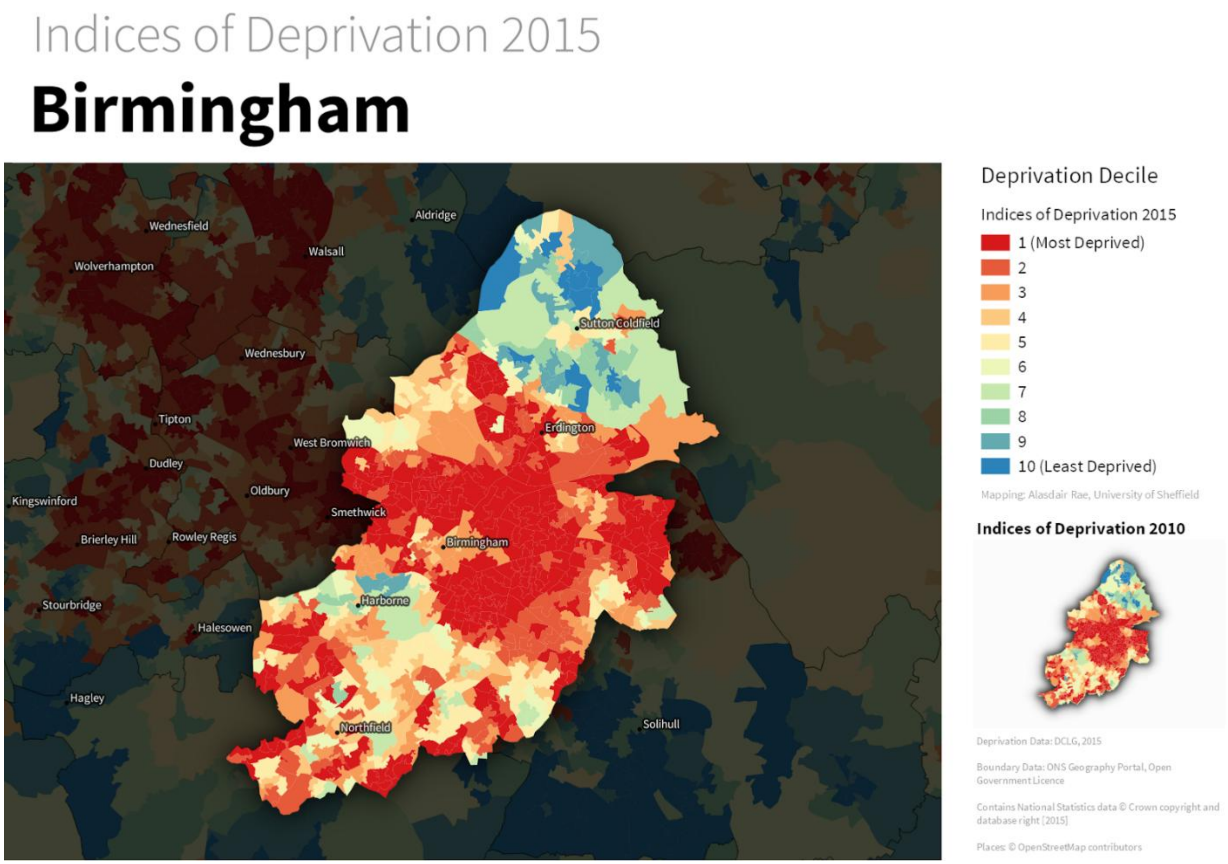
Nominal / Categorical Attributes = names/categories

- Most basic data type
- Simply names or **categories**
- Like a land use map
- Not ordered



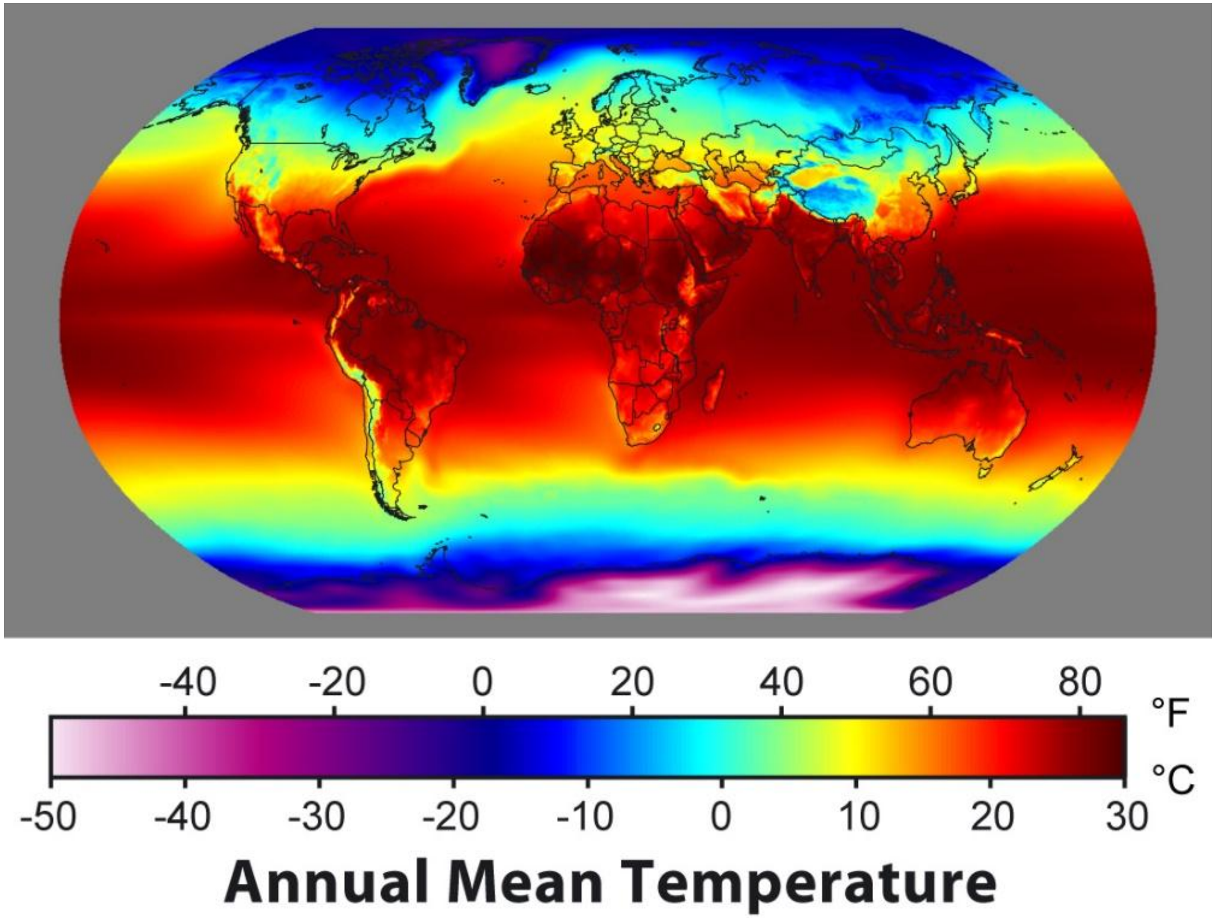
Ordinal Attributes = ordered

- Ordinal attributes have an **ordered or ranked** relationship
- They contain more information than nominal/categorical attributes as each attribute relates to others in the rank



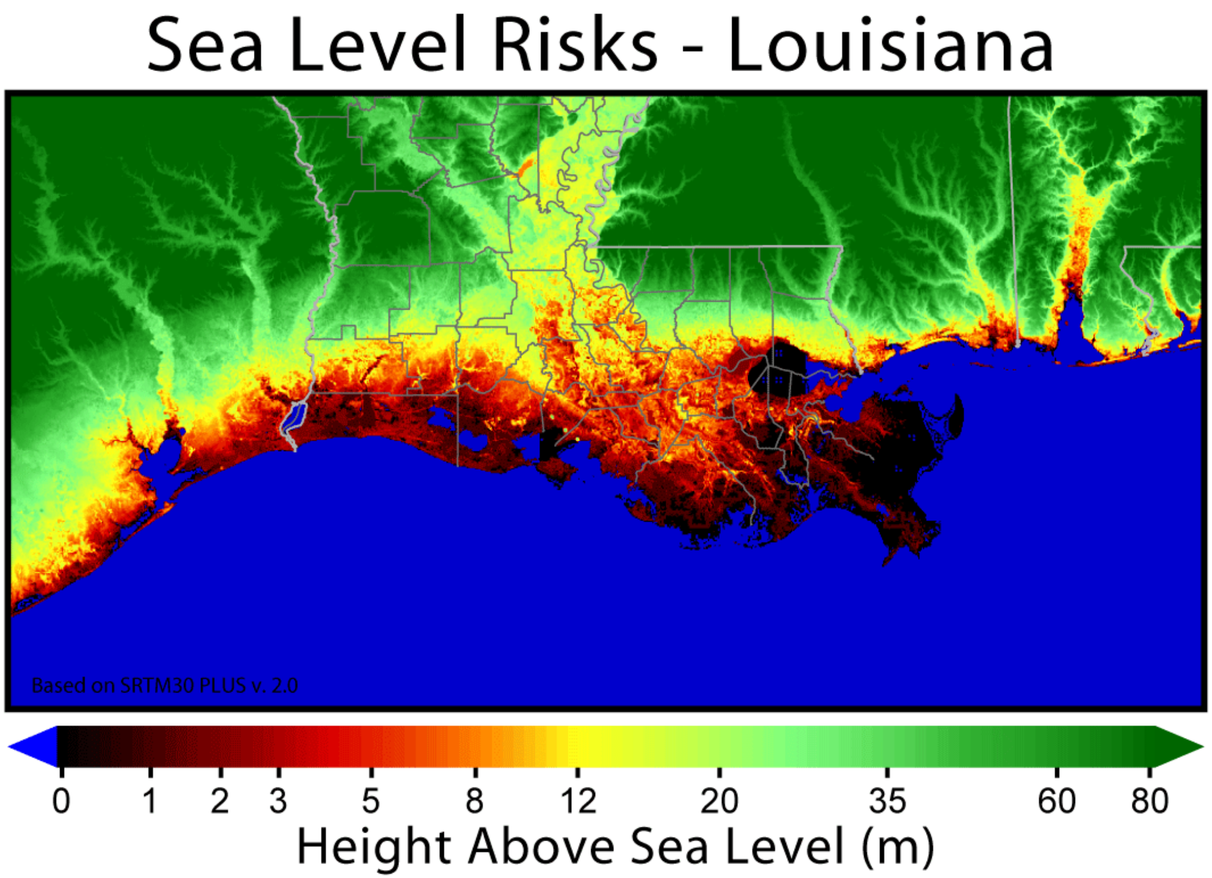
Continuous, unbounded Attributes

Very common, e.g. temperature



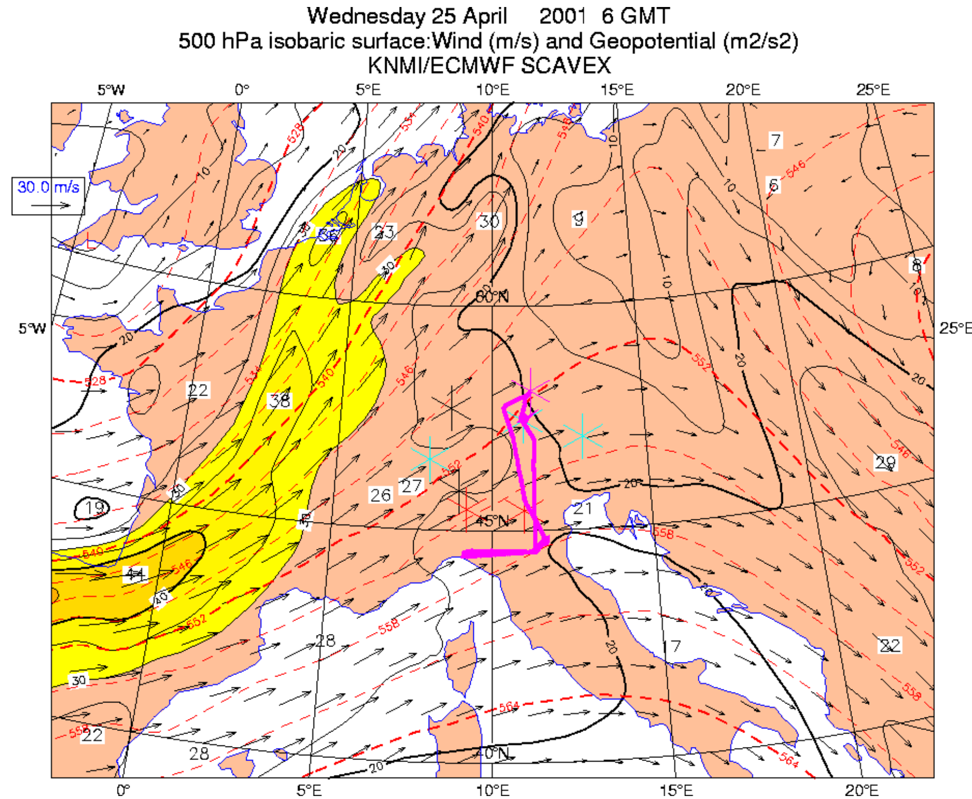
Continuous, bounded Attributes, e.g. Ratio

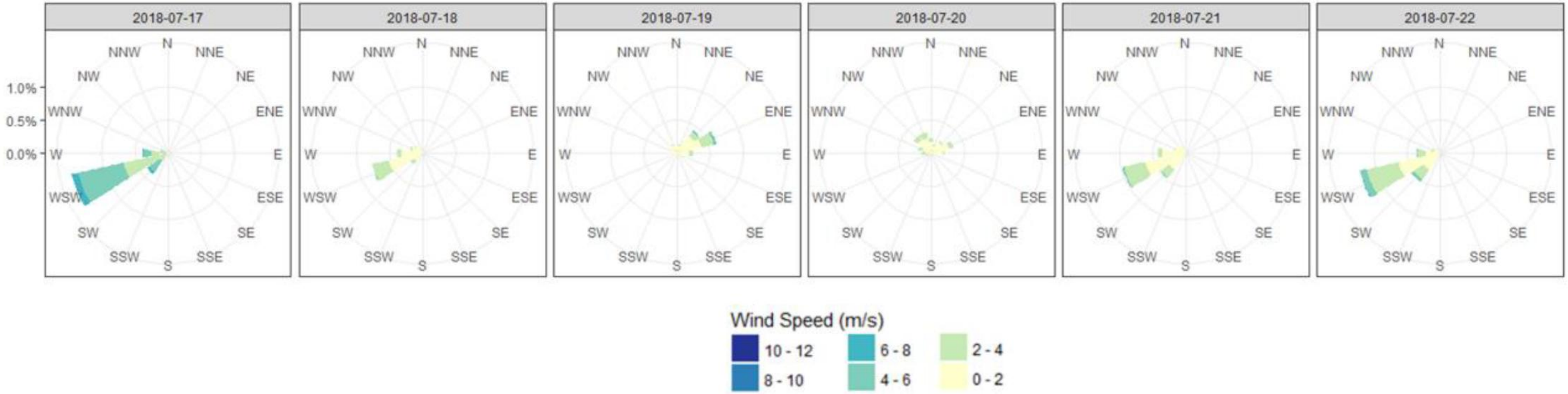
- Ratio attributes have a meaningful reference point
- Height above sea level is a classic ratio attribute



Cyclic Attributes = cycles

- Continuous
- Classic example is wind direction
- Also calendar dates
- Not a data type we will encounter too often…





Geographic Representation

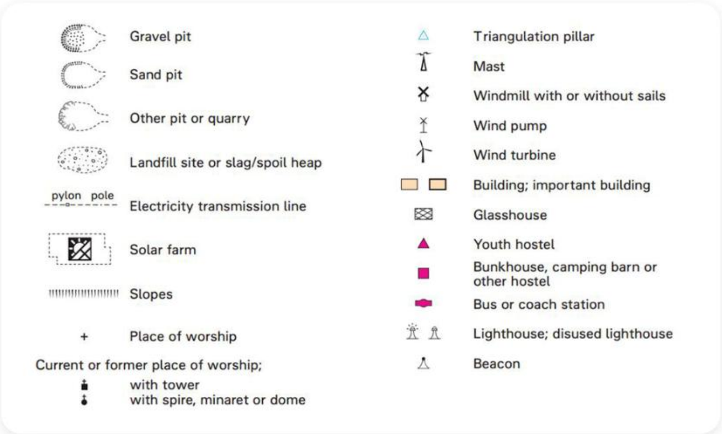
How do we show these objects in a GIS

Two ways of conceptualising geographic representation:

- **Discrete objects** & **Continuous fields -** The most fundamental distinction in geographic representation

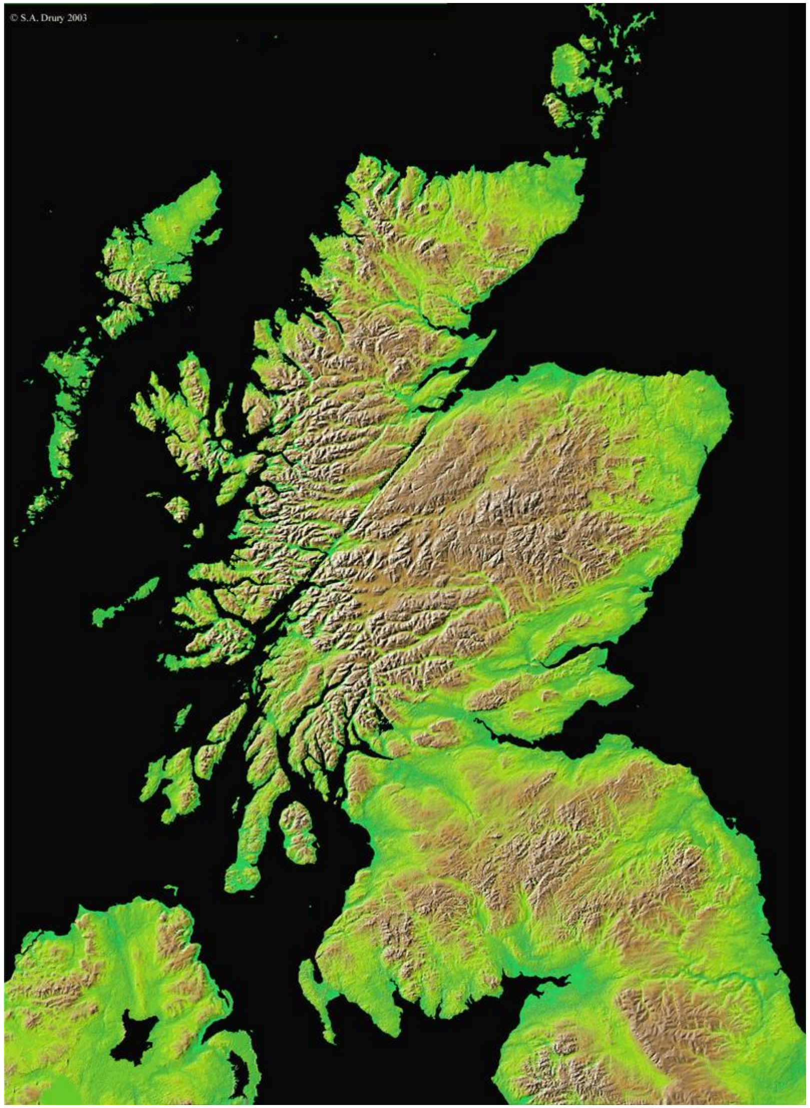
Discrete Objects

- Represent *things you can count and locate exactly* on the map — individual, separate features that exist in specific places.
- Points, lines, and polygons
- Houses, water points, etc.
- Countable
  - State Minnesota – 10,000 lakes
- Persistent through time, perhaps mobile, e.g. vehicles



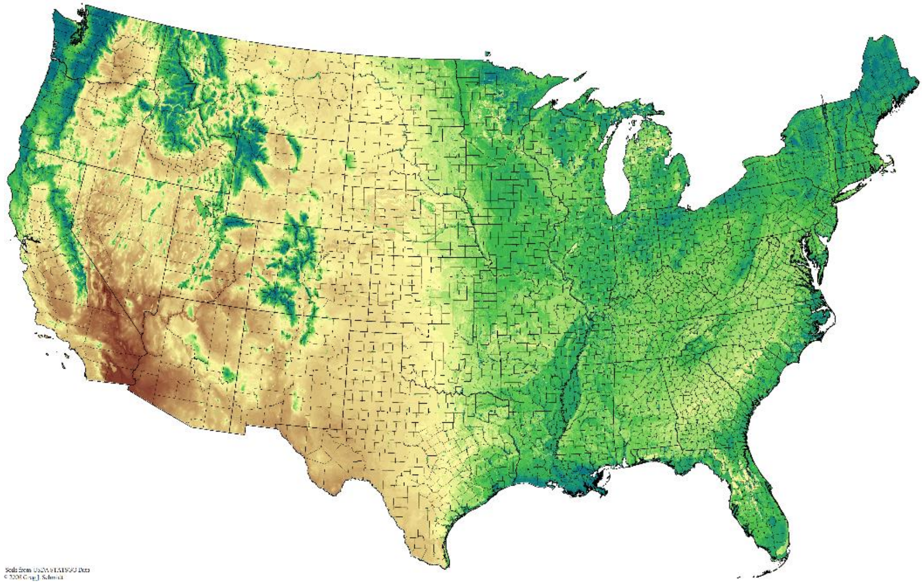
Continuous fields

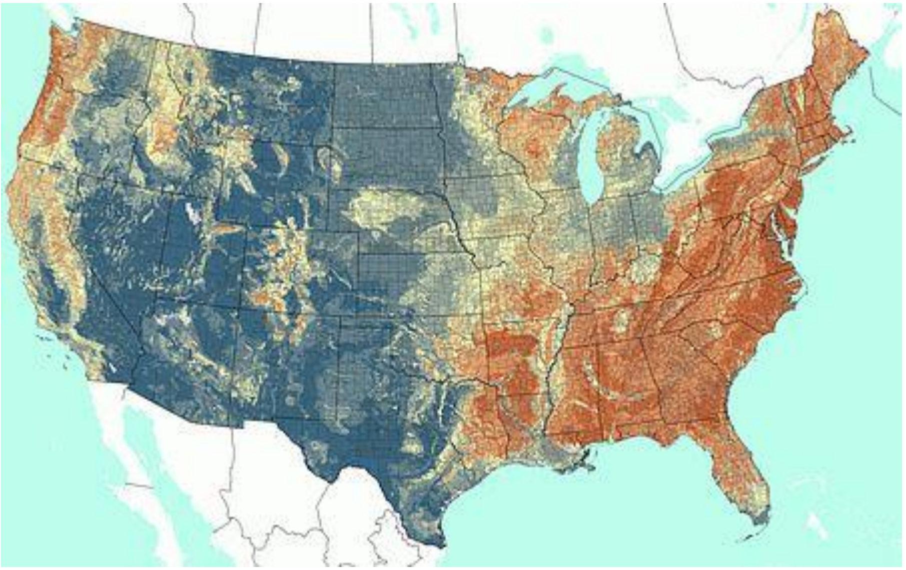
- Properties that **vary continuously over space**
- Example: Elevation
    - A single value at every point on the Earth's surface
    - Any locations can have slope, gradient, peaks, pits



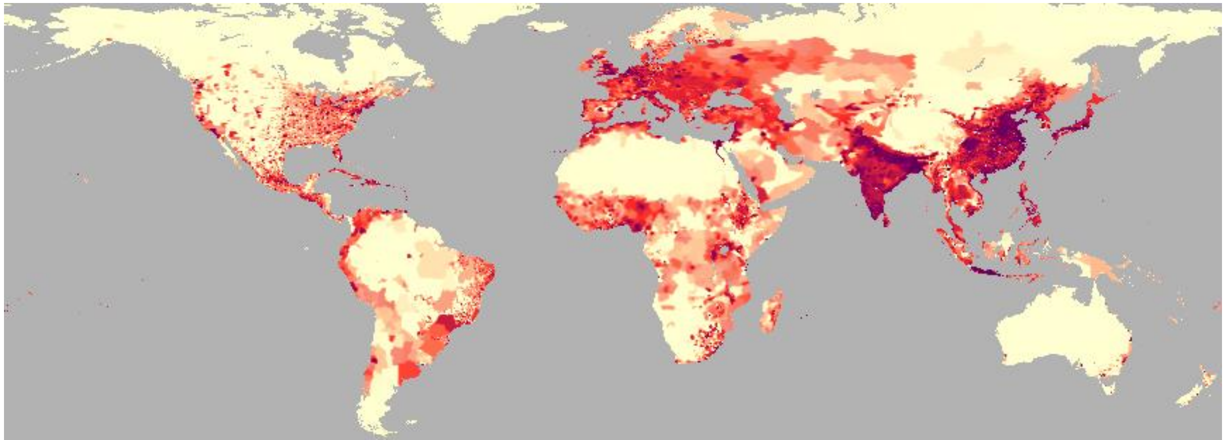
Continuous fields

- Soil properties, e.g. pH, soil moisture





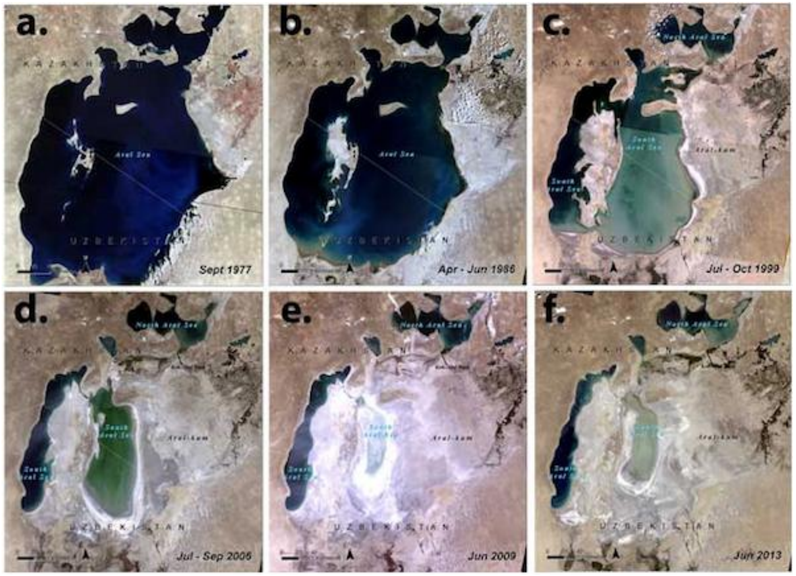
- Population density



Difficult Cases - continuum

- Lakes and other natural phenomena
  - Often conceived as objects, but difficult to define or count precisely
    - Urban area?
    - Water bodies that dry up?
- Weather forecasting
  - Forecasts originate in models of fields, but are presented in terms of discrete objects
    - Highs, lows, fronts

Aral sea: decreasing in size



How you are thinking about your data will influence how you measure it…

Is a lake discrete (object) or continuous

Is urban area discrete (object) or continuous

Getting Around Difficult Cases

In field view you may assign a scale of "lakeness"

"Lakeness"	Definition
1	Location is always dry under all circumstances (not a lake!)
2	Location is sometimes flooded in Spring (sometimes a lake!)
3	Location supports marshy vegetation (edge of a lake!)
4	Water is always present to a depth of less than 1m (shallow lake!)
5	Water is always present to a depth of more than 1m (deep lake!)

Getting Around Difficult Cases

Degree of urbanisation - https://human-settlement.emergency.copernicus.eu/degurba.php

GHSL - Global Human Settlement Layer

Open and free data and tools for assessing the human presence on the planet

Home Copernicus Data and tools Visual analytics Degree of urbanisation Knowledge and training News

Home > Degree of urbanisation

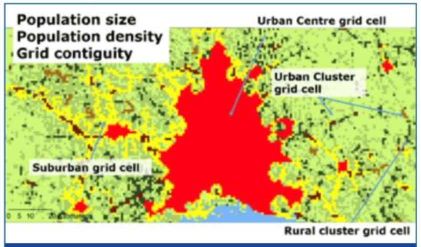
Harmonised global definition of cities and settlements

The Degree of Urbanisation, a new global definition of cities, urban and rural areas

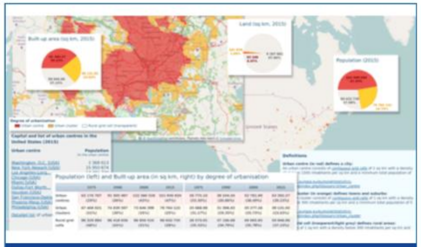
In its 51st session, the United Nations Statistical Commission endorsed the methodology for delineation of cities and urban and rural areas for international and regional statistical comparison purposes. The method was proposed by a consortium of international organisations (EU, OECD, World Bank, FAO, UN-Habitat, ILO) led by the EU.

This page presents key resources needed for information, visualisation and implementation of the **Degree of Urbanisation**.

Why a Global Definition?
The rationale behind the approach



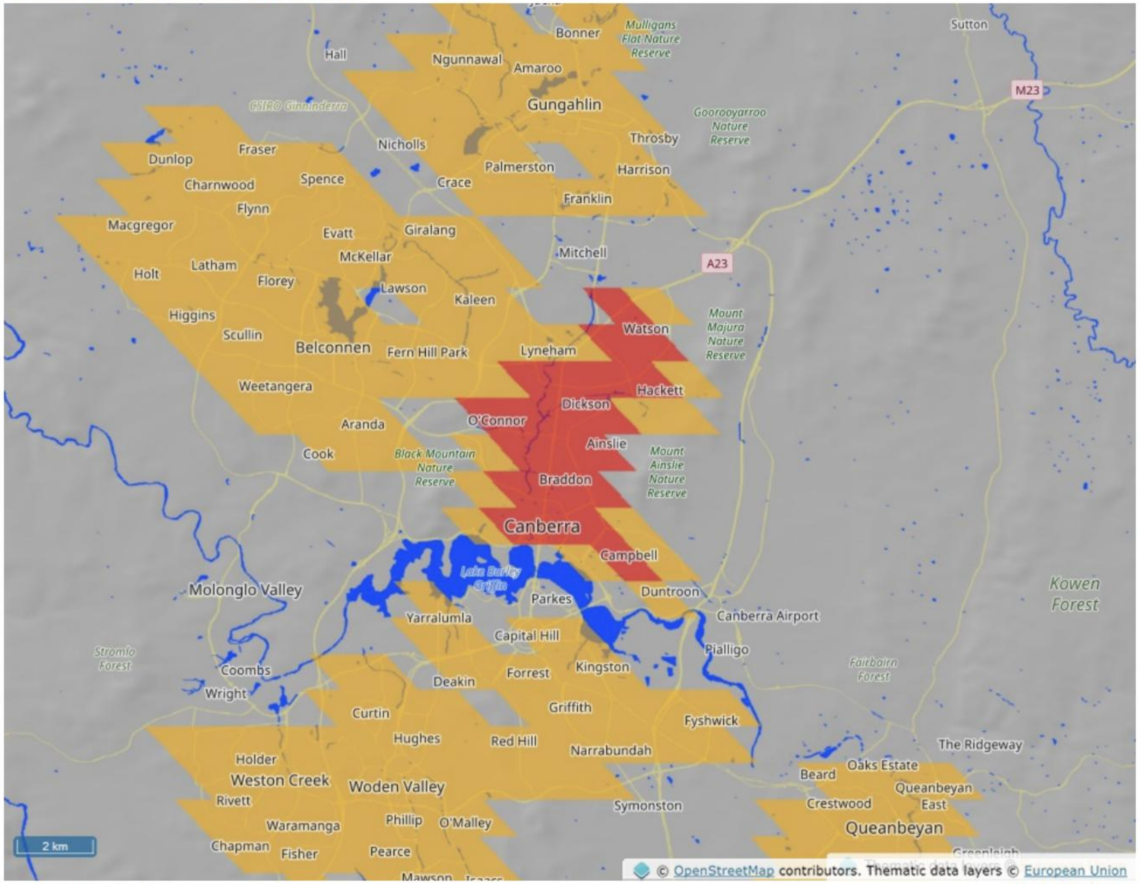
Definitions
The classes of the Degree of Urbanisation



Country Fact Sheets
Degree of Urbanisation by country

Getting Around Difficult Cases

North Canberra [Canberra], Australia



Degree of urbanisation

Urban centre Urban cluster Rural grid cell (transparent)

The QR Code on the right opens an interactive version of the map above (minimum screen width resolution required: 600 pixels).

Getting Around Difficult Cases

https://read.dukeupress.edu/demography/article/62/4/1171/402145/Beyond-the-Immediate-Impacts-of-COVID-19-on

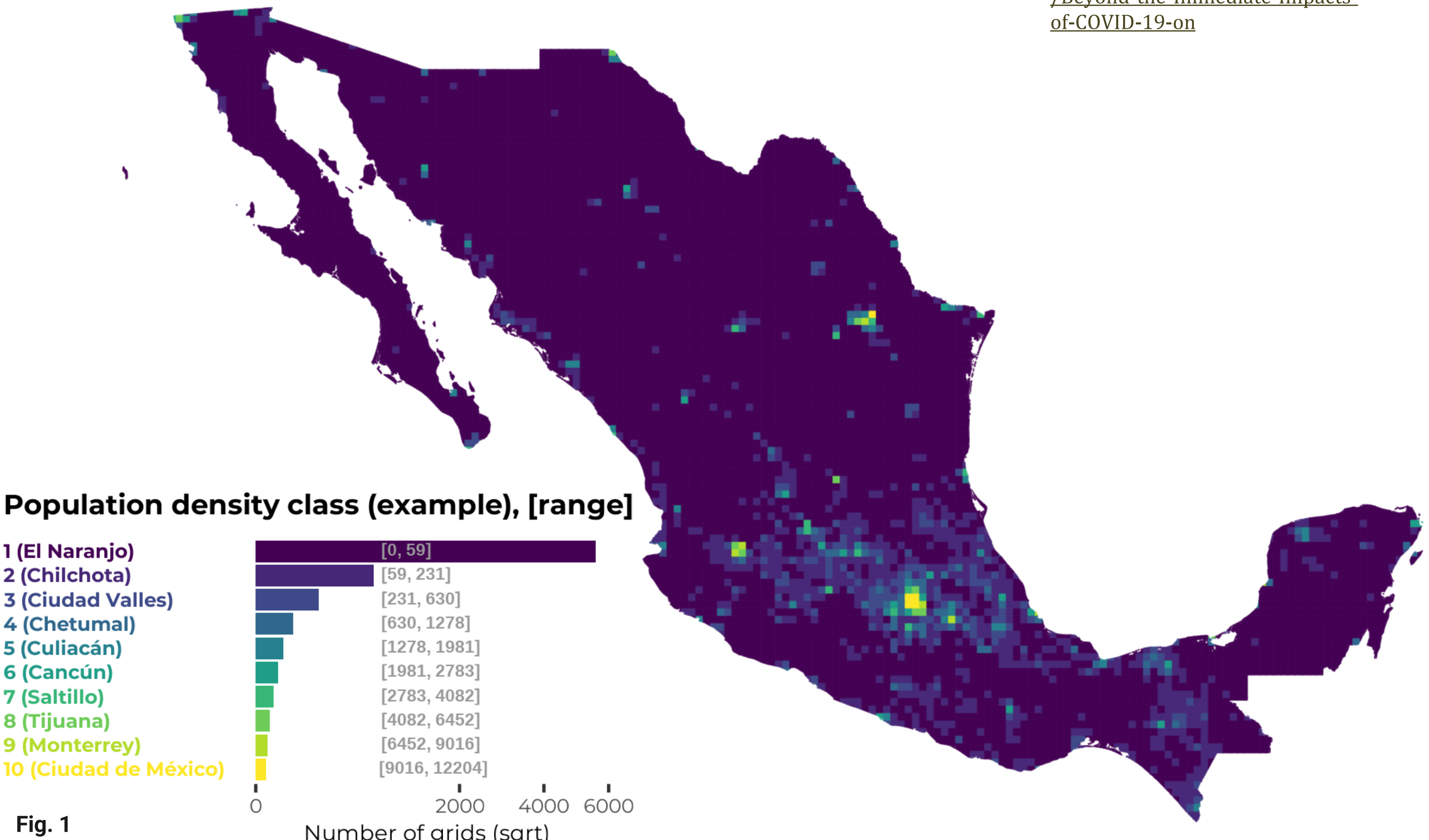
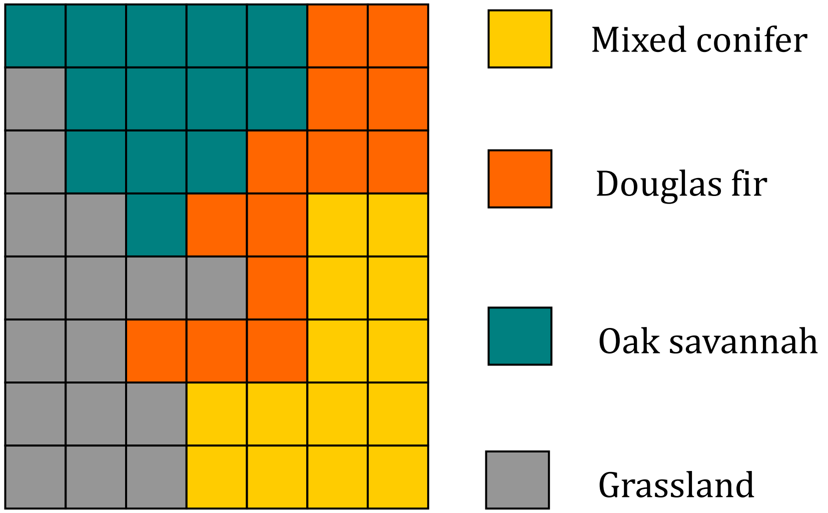


Fig. 1
Population density classes in Mexico from the Microsoft Bing Maps Tile System, showing the number of grids and range of population density in each class, 2020

Digitally Representing Discrete Objects and Fields

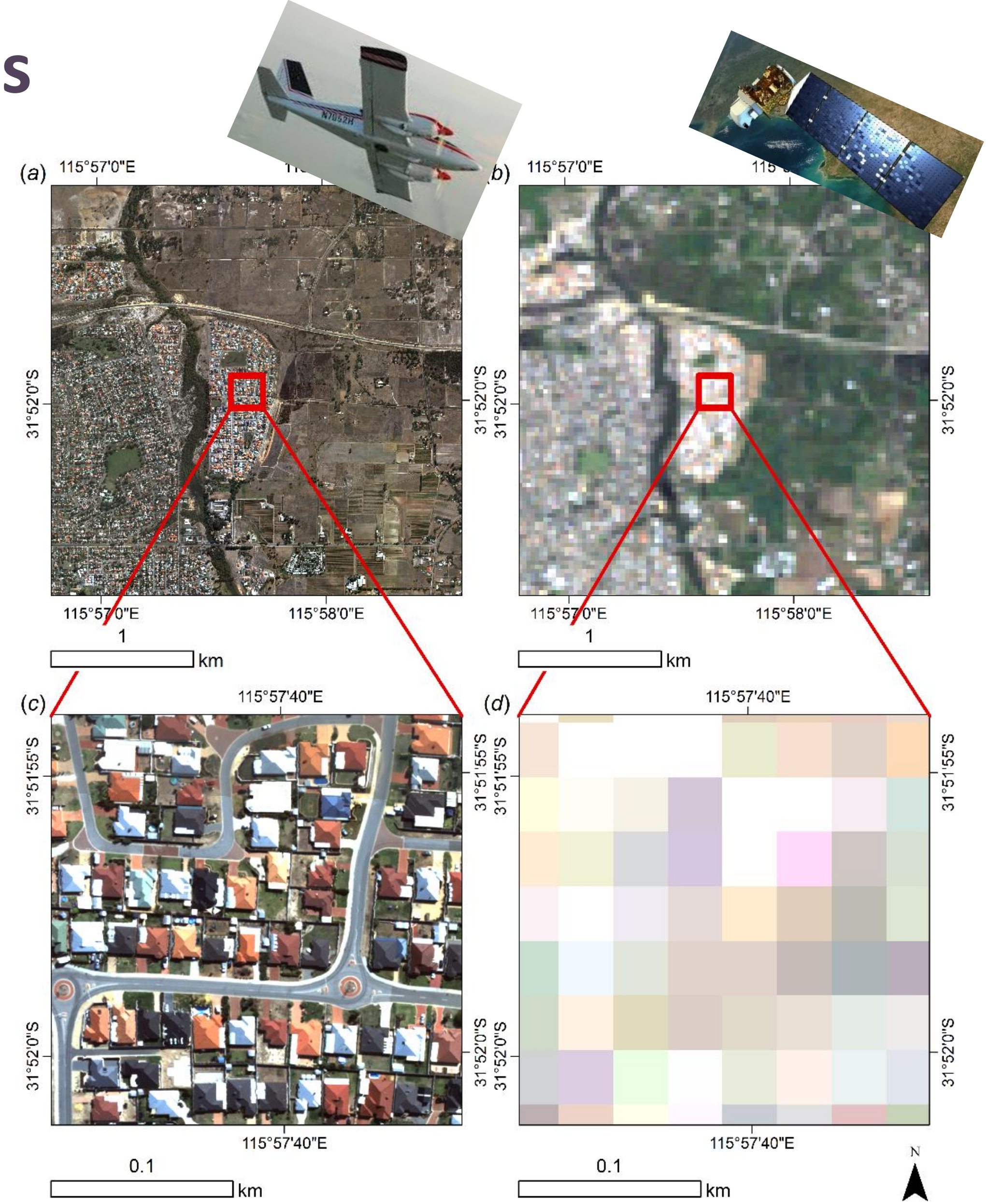
Raster

- Divide the world into square cells
- Represent discrete objects as collections of one or more cells
- Represent fields by assigning attribute values to cells
- **More commonly used to represent continuous fields than discrete objects**

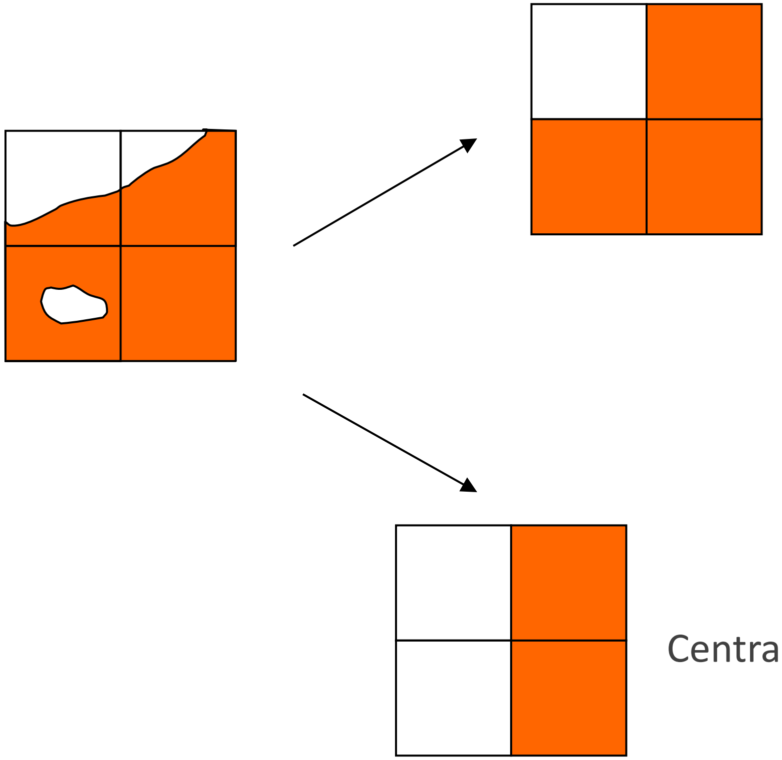


Characteristics of Rasters

- Pixel size
- Assigning reality into pixels (Assignment scheme)
  - The **value of a cell may be an average** over the cell, or a total within the cell, or the commonest value in the cell
  - It may also be the value found at the **cell's central point**
  - Could be **interpolated**



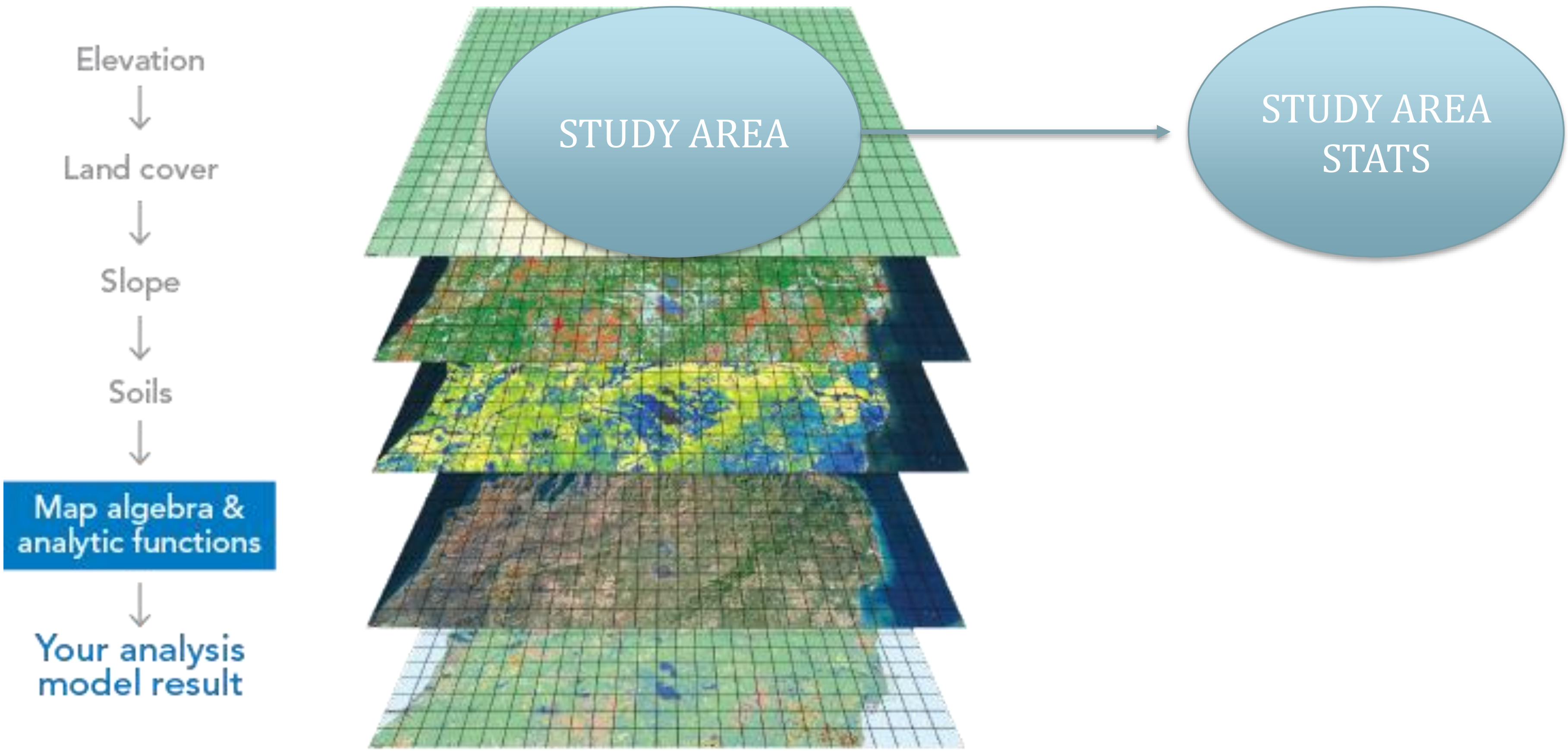
Assigning cell values



Largest share rule

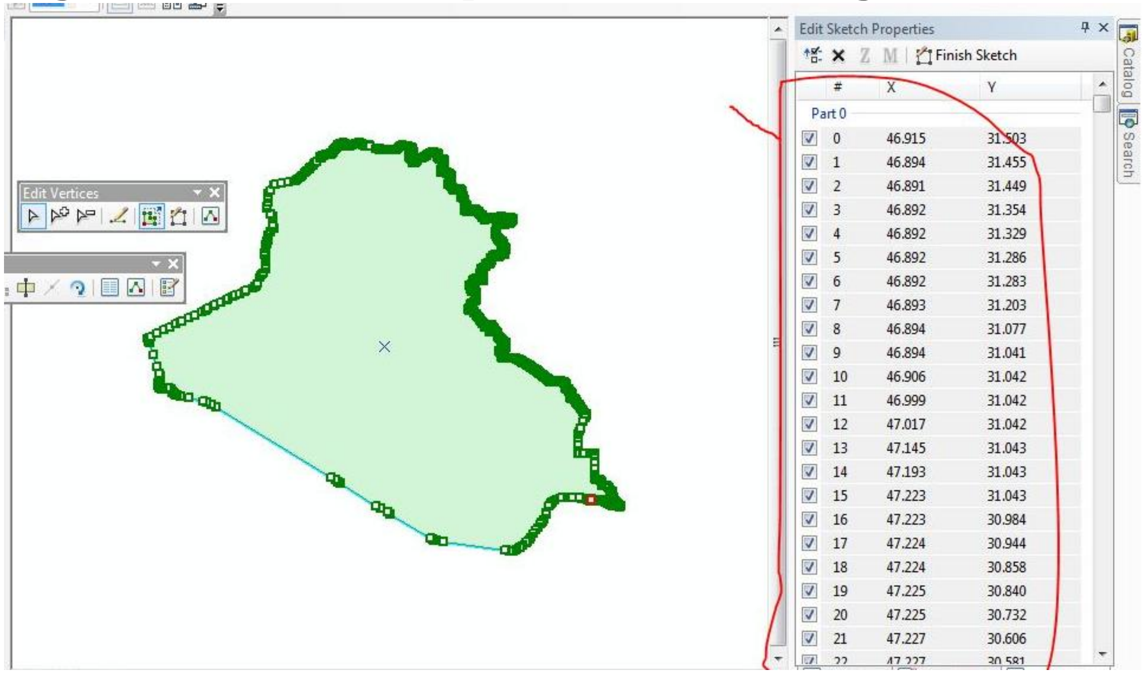
Central Point Rule

Stacking rasters and processing



Vector Data

- Used to represent **points**, **lines**, and **polygons**
- All are represented using coordinates
  - Lines as *polylines*
  - Areas as polygons
    - Straight lines between points, connecting back to the start



Raster vs Vector

- **Volume** of data
  - Raster becomes more voluminous as cell size decreases
- **Source** of data
  - Remote sensing, elevation data come in raster form
  - Vector favored for administrative data
- **Analysis**
  - Some analyses are better suited to raster (map calculation, suitability indices etc.), some to vector (route finding, network analysis etc.)

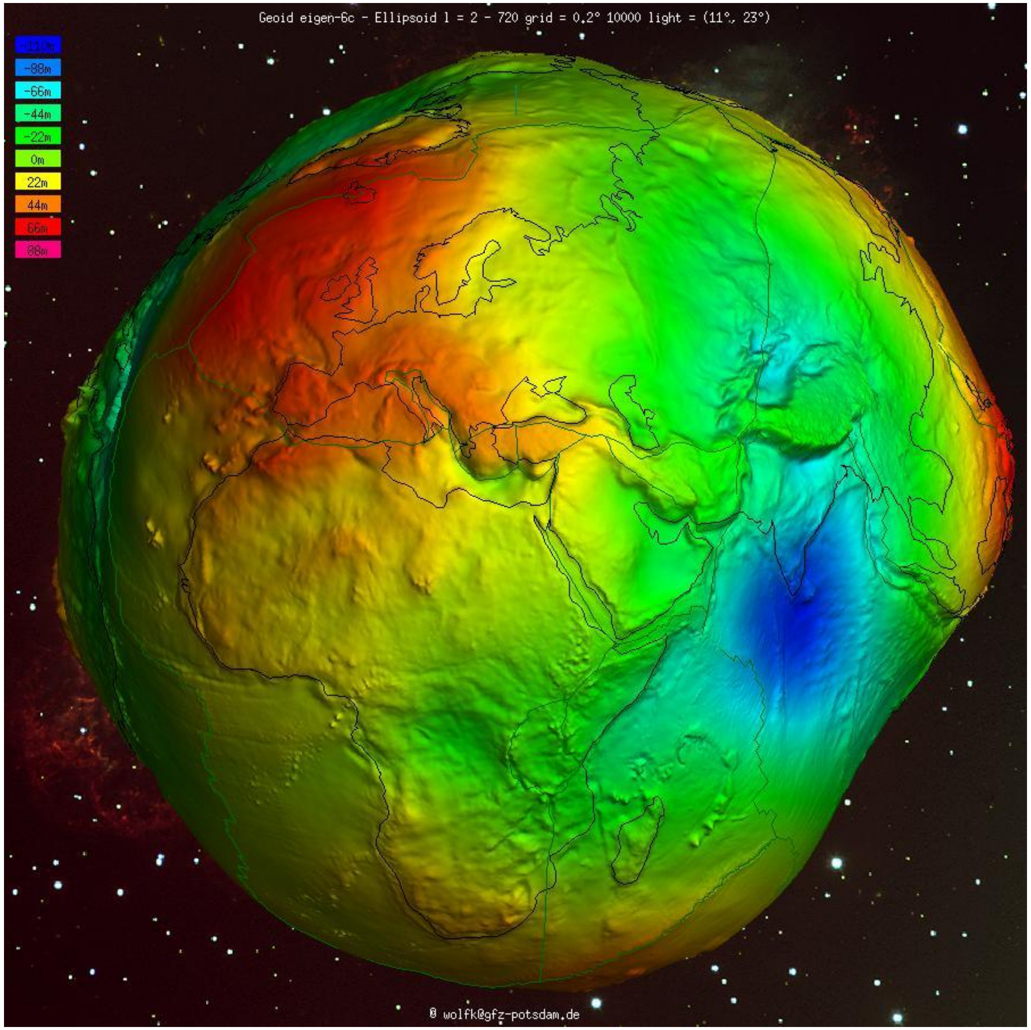
Raster vs Vector

Issue	Raster	Vector
Volume of data	Depends on cell size	Depends on density of vertices
Sources of data	Remote sensing, imagery	Social and environmental data
Applications	Resources, environmental	Social, economic, administrative
Software	Raster GIS, image processing	Vector GIS, automated cartography
Resolution	Fixed	Variable

Map Projections – or how to represent 3D in 2D

Representing the Globe

- The earth is a 3D sphere (well, almost). It's wider than it is tall
- In order to locate a point on the surface of a sphere, we need a set of coordinates
- Coordinates will tell us how near to the top or bottom of the sphere we are, or how far around
- But where do we start?



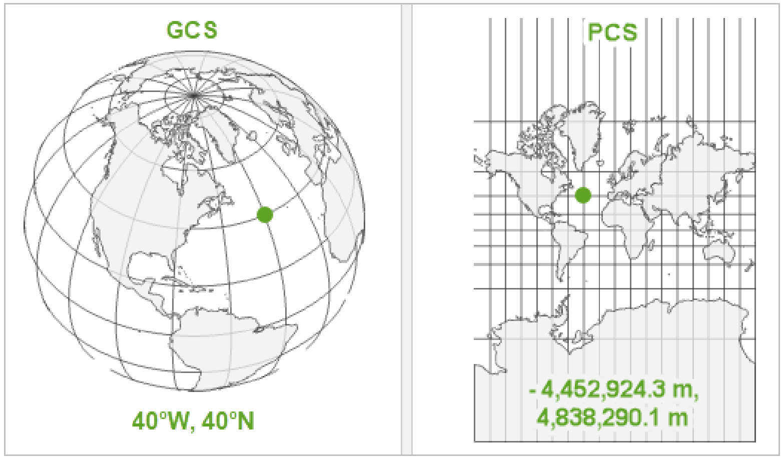
3D into 2D - Projecting

- But what if we want to view a 3D object in 2D?
- 2D planes are easier to deal with (ever tried navigating with a globe?!)
- Projections enable us to represent 3D coordinates on a 2D surface
- But losing a dimension means we lose some information – different projections lose different information

The Mercator Projection

3D into 2D - Projecting

- Geographic coordinate reference system – 3D **WHERE** the place is on earth. It has **ANGULAR** units (degrees). BUT it will draw flat as your screen is flat.
- Projected coordinate reference system – **HOW** to draw that place on a FLAT surface. It has **LINEAR** units (e.g. metres)



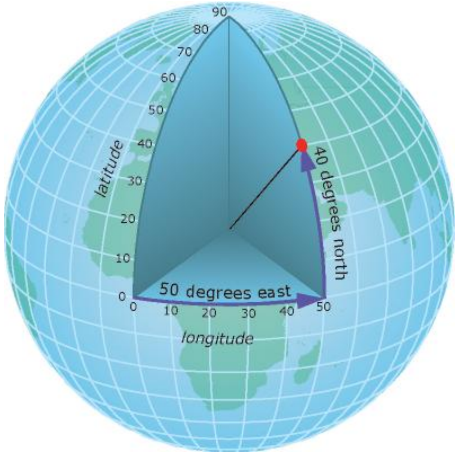
From: https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/gcs_vs_pcs/

- A Projected coordinate reference system **CONTAINS** a geographic coordinate reference system. It's just a geographic one that is projected!

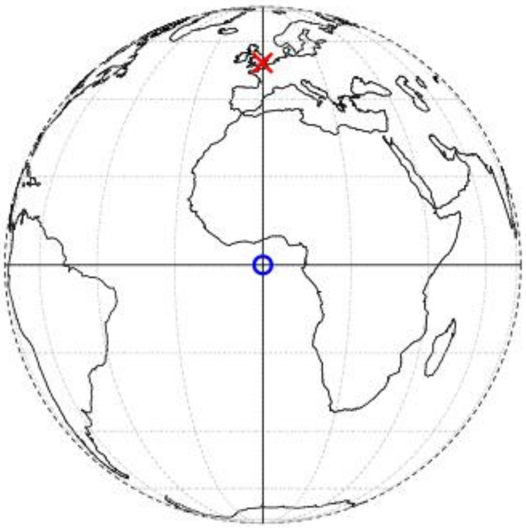
World Geodetic System (WGS 84)

- Standard (3D) method of representing our solid surface (last revision established in 1984)
- Sphere divided into 360 parts called degrees (of longitude).
- Each degree has 60 minutes
- Each minute has 60 seconds
- 51° (degree) 30’ (min) 35.5140’’ (secs) N – degree/minute/second (DMS)
- Example: 0° 7' 5.1312’’ W
- Also commonly given as Decimal degrees (DD) – often seen on GPS
- Latitude (horizontal)
- Longitude (vertical)

https://gisgeography.com/decimal-degrees-dd-minutes-seconds-dms/



Geographic CRS WGS 84 – spheroid (degrees)



Representing the Globe on a computer…

- https://www.youtube.com/watch?v=8I_VpC6IuJs

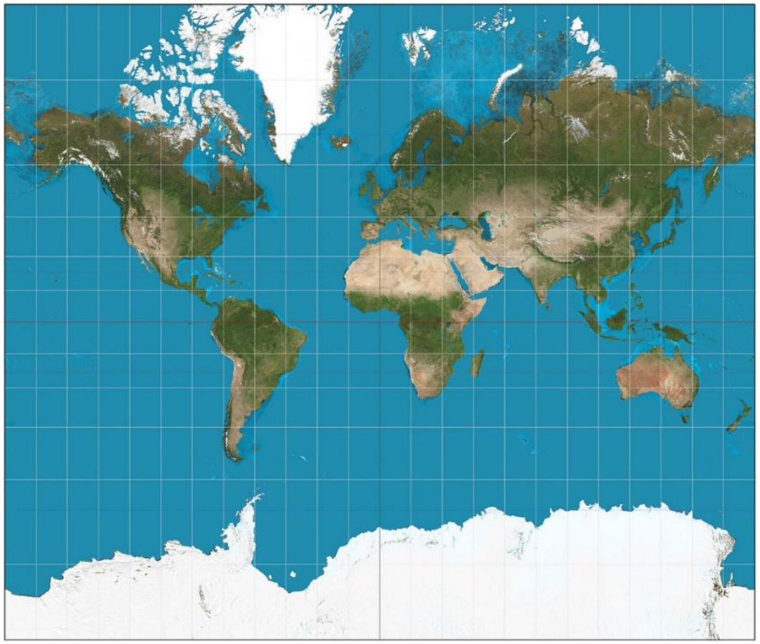
- Geographic CRS = treats data as sphere (angular units)

- Projected CRS = treats data as flat object (metric units)

- To change between them we …**reproject** the data

Mercator Projection

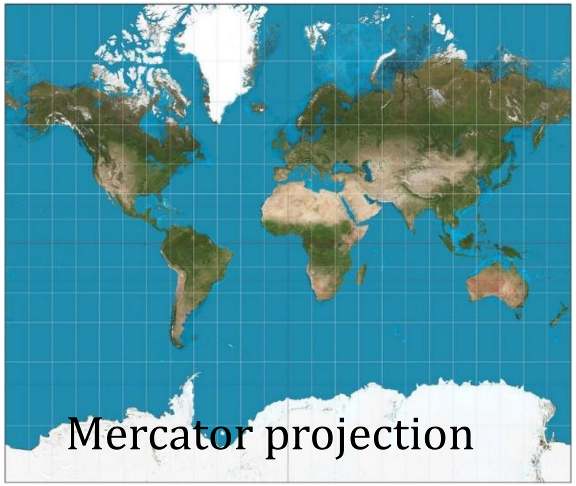
- Invented for navigation purposes by Gerardus Mercator in 1569
- Bearings (angles) are preserved (particularly useful when navigating a ship with a compass!) = **constant true direction**
- Area and distance are not preserved – exaggerated at poles, downplay at equator
- Greenland 2.1 million square km. Africa 30.04 square km

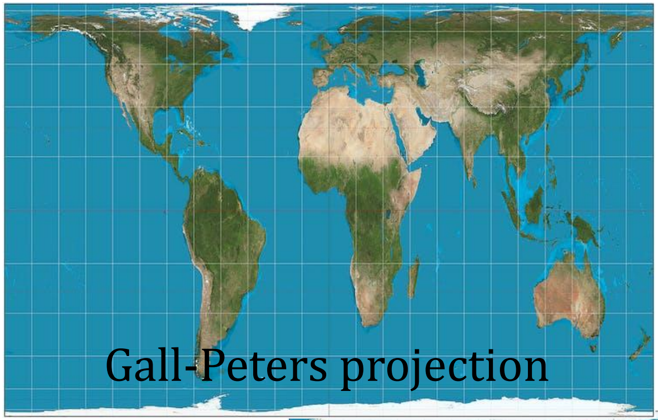


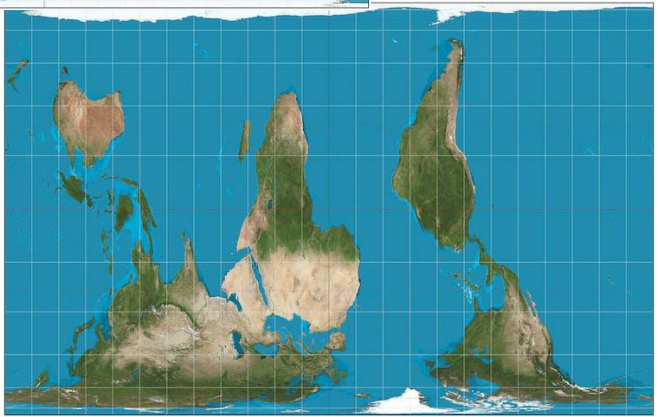
Have a play with this!

Why Mercator?

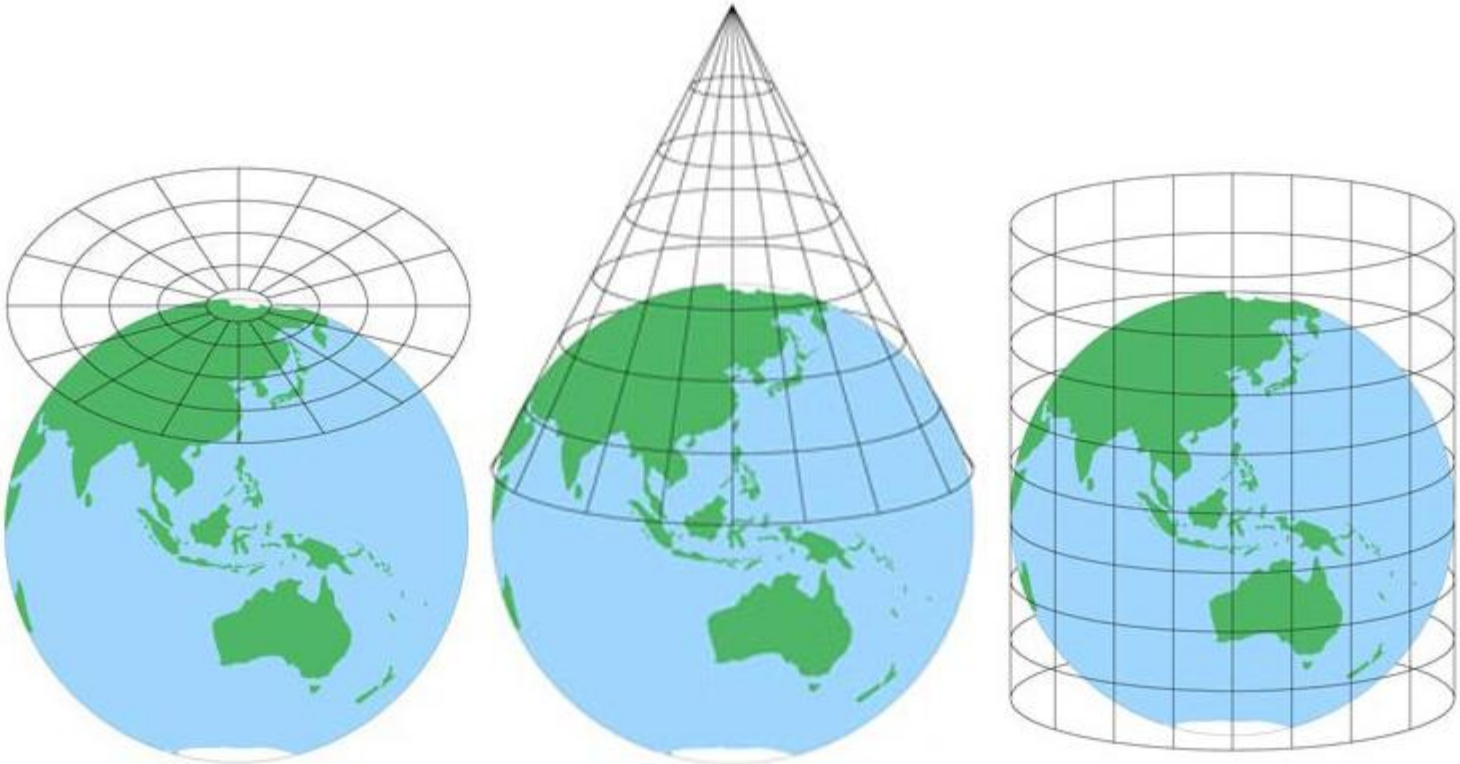
- Mercator:
  - European imperialist attitude
  - Size = POWER (appearance of large land size)
  - Ethnic bias
  - Are you a sailor?
- Gall-Peters projection
  - Right size of countries (areas preserved)
- South Up
  - Which countries are at the top = importance


Mercator projection


Gall-Peters projection



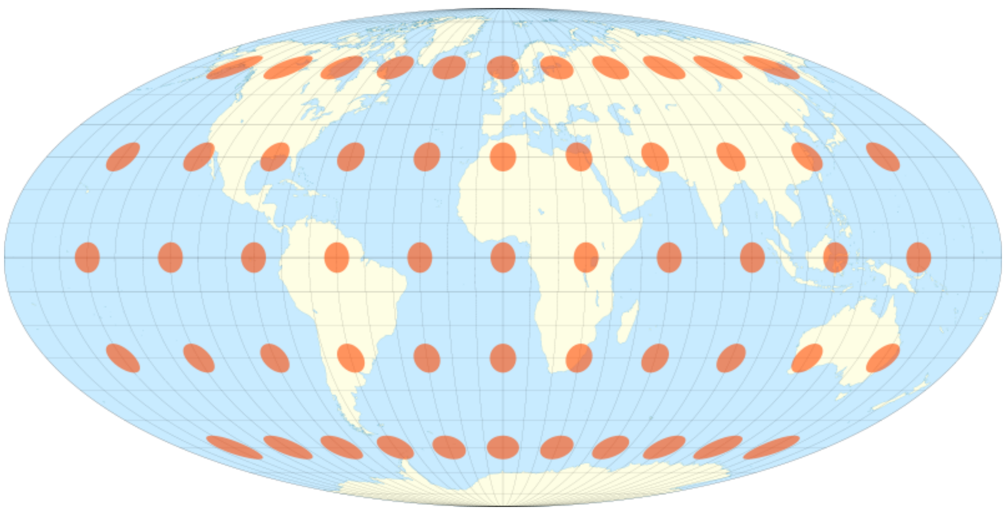
3D into 2D - Projecting



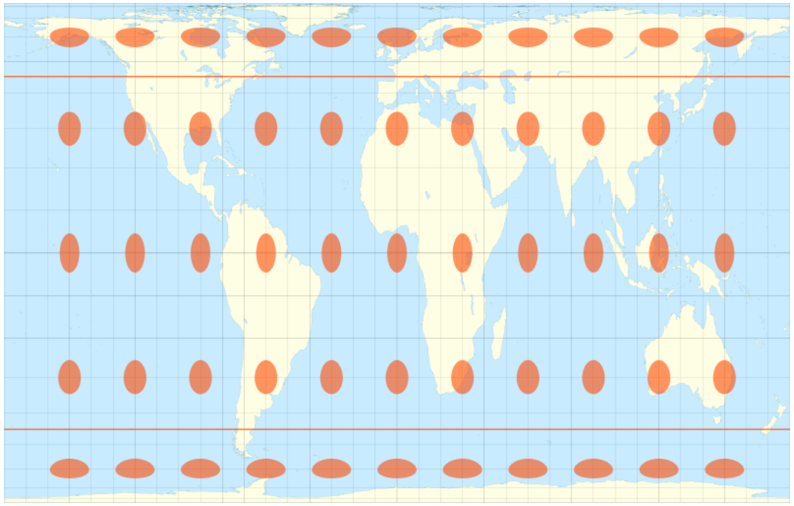
https://www.youtube.com/watch?v=wlfLW1j05Dg

Preserving Size

- Size not shapes (near edge)



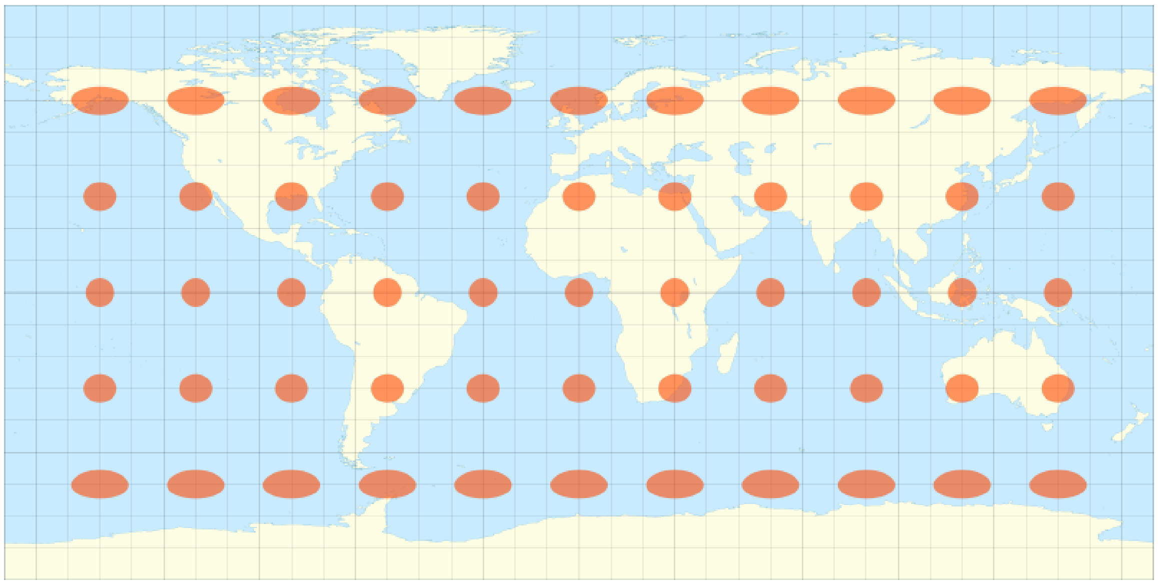
MOLLWEIDE PROJECTION



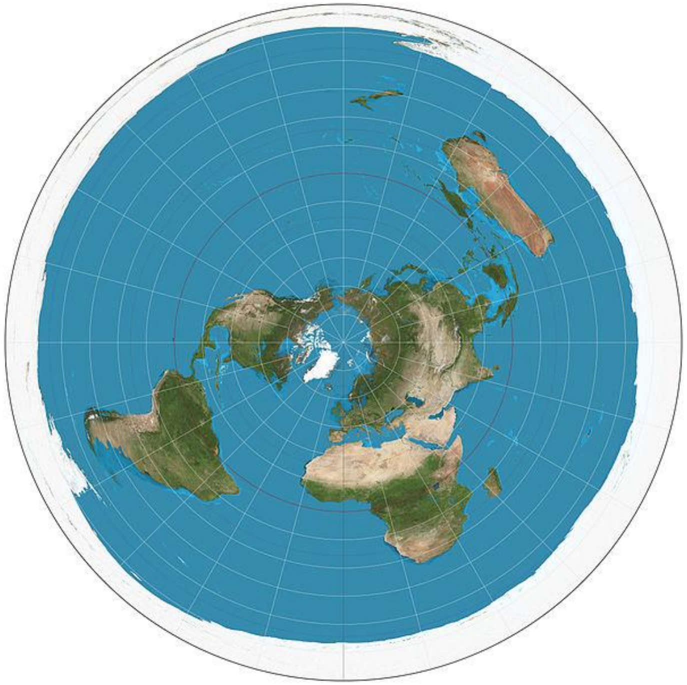
PETERS PROJECTION

Preserving Size / Distance

- Size not shapes (near edge)
- All points correct distance from centre but not between any others



EQUIRECTANGULAR PROJECTION

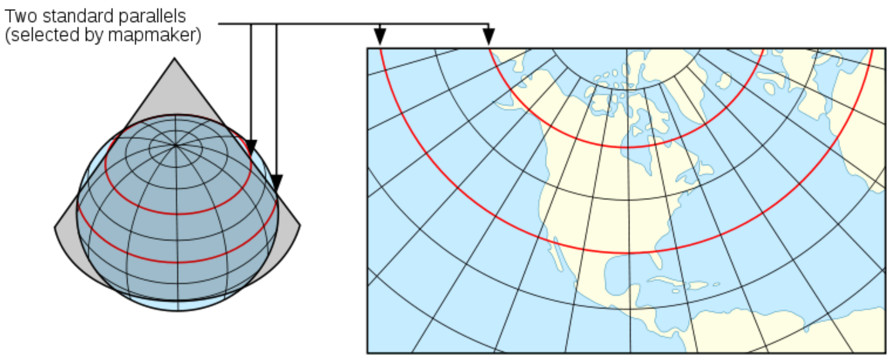


AZIMUTHAL EQUIDISTANT PROJECTION

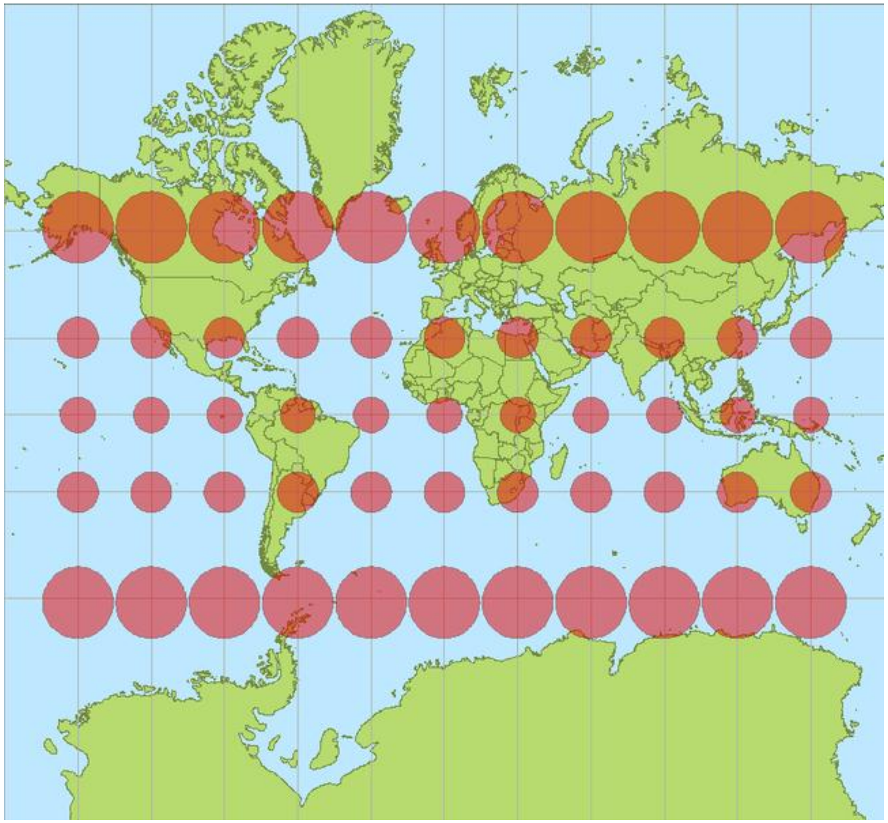
UN LOGO

Preserving Shape

- Shape not angles



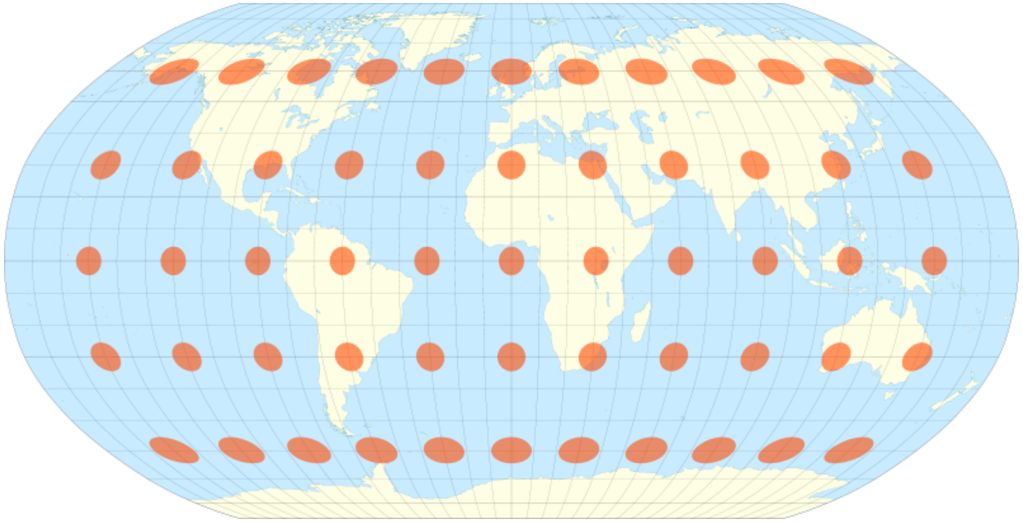
LAMBERT EQUAL AREA PROJECTION



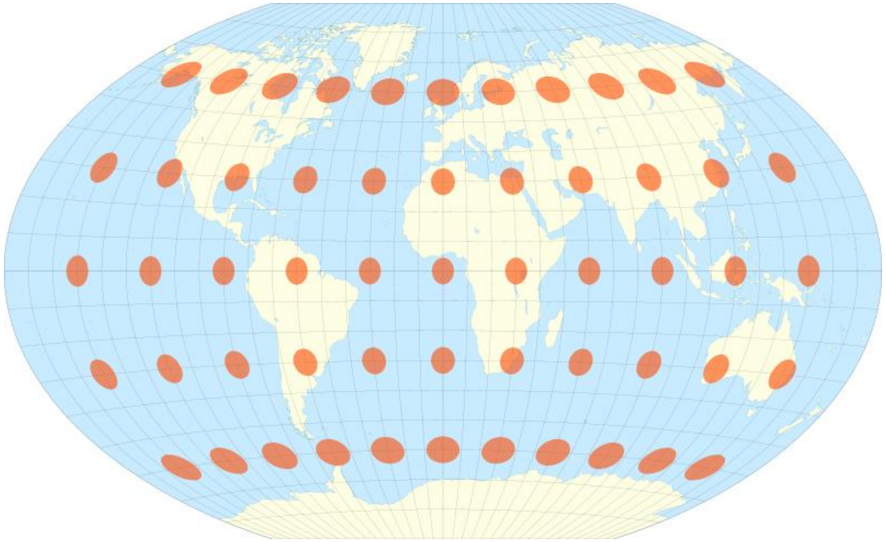
LAMBERT EQUAL AREA

Preserving Everything!! Well, almost…

- Compromise between shape and distance distortions

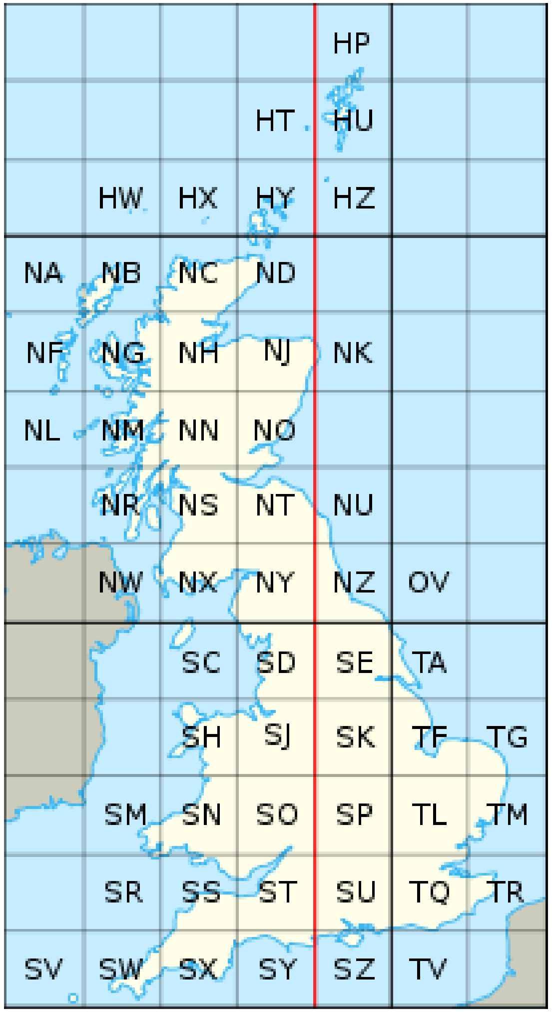


ROBINSON PROJECTION



WINKEL TRIPEL PROJECTION

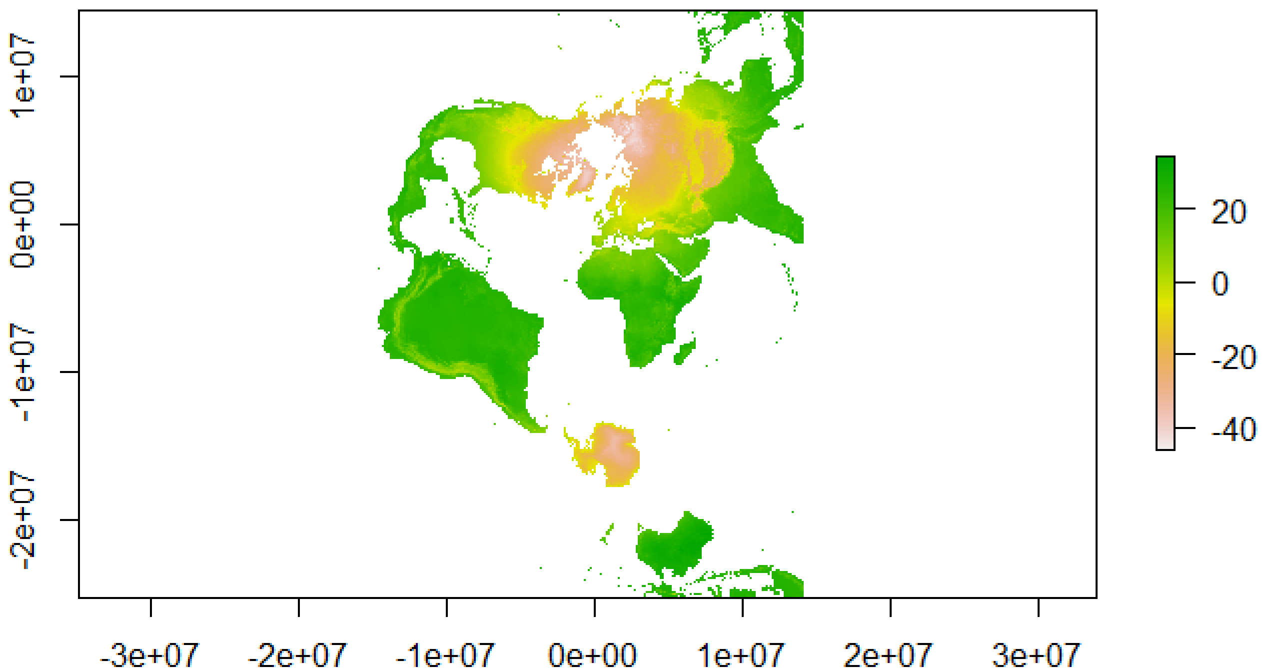
Preserving Everything: British National Grid (BNG)



- BNG is **a local projected coordinate reference system**
- **It has a local datum**
- Coordinates in Eastings and Northings.
- Measurements in metres.
- Point of origin is near the Isles of Scilly.
- Standard for all maps of Britain.
- Most countries have their own version

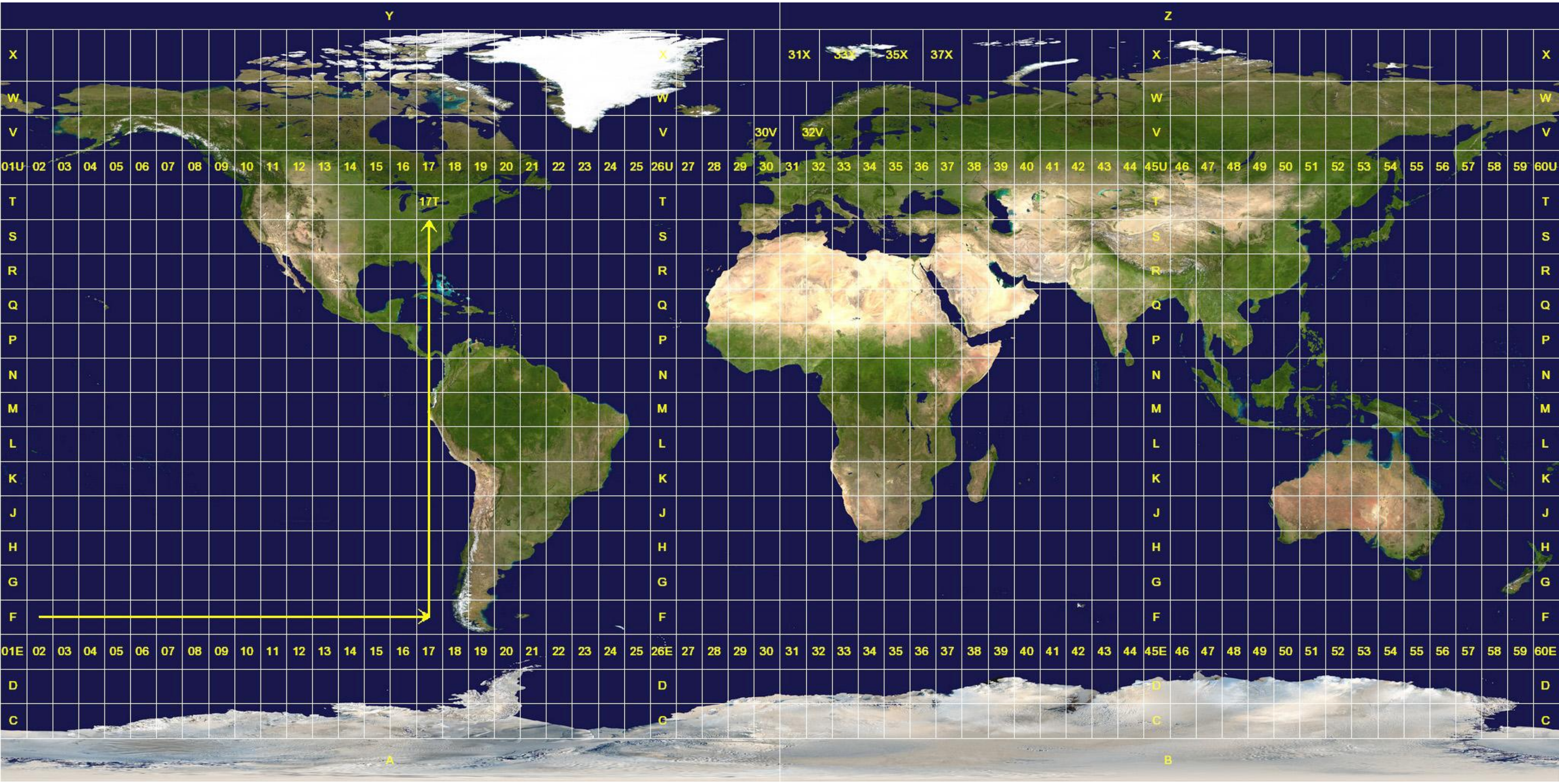
What is this?

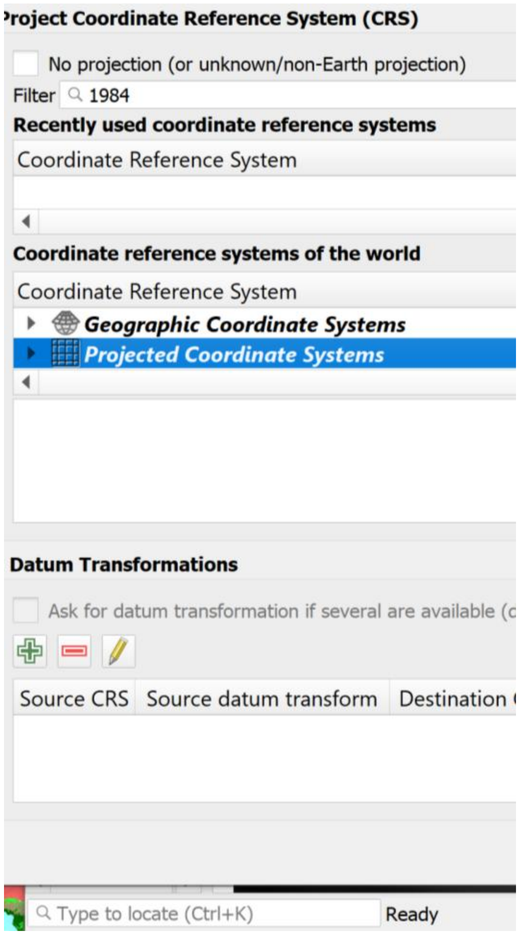
- British National Grid (BNG) applied to the world
- Makes no sense…
- BNG is specific to the UK, other counties will have the same
- Or UTM zones….

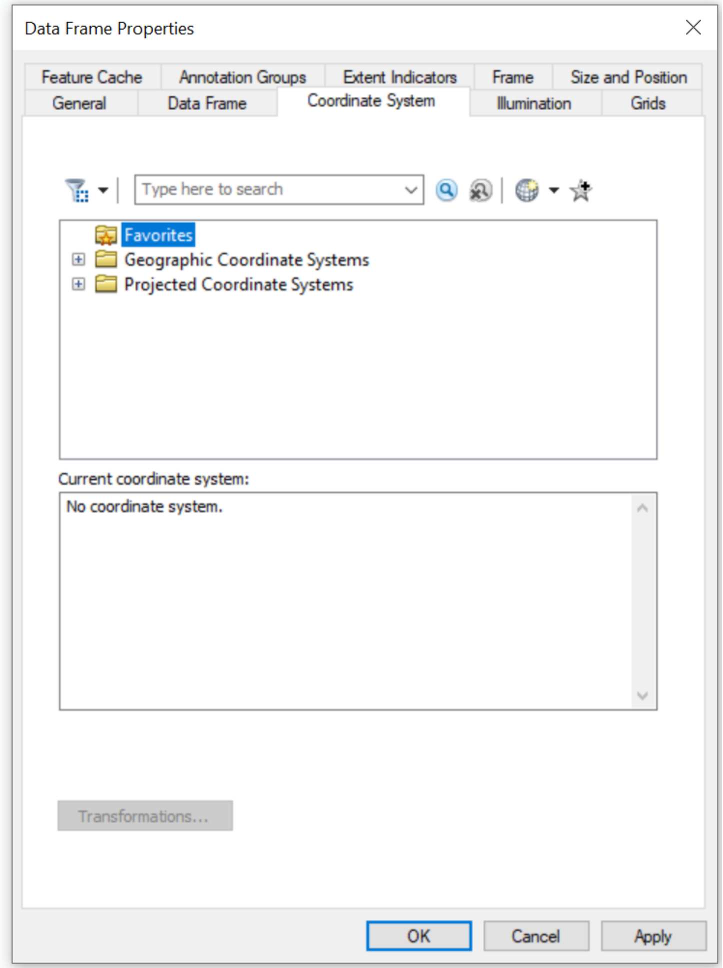


Universal Transverse Mercator (UTM)

- Divides Earth into 60 zones and projects each zone separately
- Projected CRS




Project Coordinate Reference System (CRS)
No projection (or unknown/non-Earth projection)
Filter 1984
Recently used coordinate reference systems
Coordinate Reference System
Coordinate reference systems of the world
Coordinate Reference System
Geographic Coordinate Systems
Projected Coordinate Systems
Datum Transformations
Ask for datum transformation if several are available
Source CRS
Source datum transform
Destination
Type to locate (Ctrl+K)
Ready


Data Frame Properties
Feature Cache
Annotation Groups
Extent Indicators
Frame
Size and Position
General
Data Frame
Coordinate System
Illumination
Grids
Type here to search
Favorites
Geographic Coordinate Systems
Projected Coordinate Systems
Current coordinate system:
No coordinate system.
Transformations...
OK
Cancel
Apply

Interactive Projection Fun!

- http://www.jasondavies.com/maps/transition/

Defining Spatial / Coordinate Reference Systems

- Knowing which coordinate reference system (CRS) your coordinates are in is absolutely **vital** for being able to specify your point on the earth correctly
- Frequently in GIS you will work with data which refer to points on the earth using different CRSs. Therefore in order to compare them, you will need to know which data are in which CRS and how to convert between them – **getting the wrong CRS is one of the most common sources of error in GIS**
- All of the projections described in the slides above (and many more besides) can be identified with a unique Spatial Reference System Identifier (SRID)

SRID, EPSG and Proj.4

- One of the more commonly used sets of SRID (spatial reference identifier) values are maintained by the European Petroleum Survey Group (EPSG). For example:
    - **EPSG:4326** refers to the WGS 84 world geodetic system
    - **EPSG:27700** refers to British National Grid
- Most EPSG identifiers will also have a Proj4 string. For example, the Proj4 string for **EPSG:4326** is:

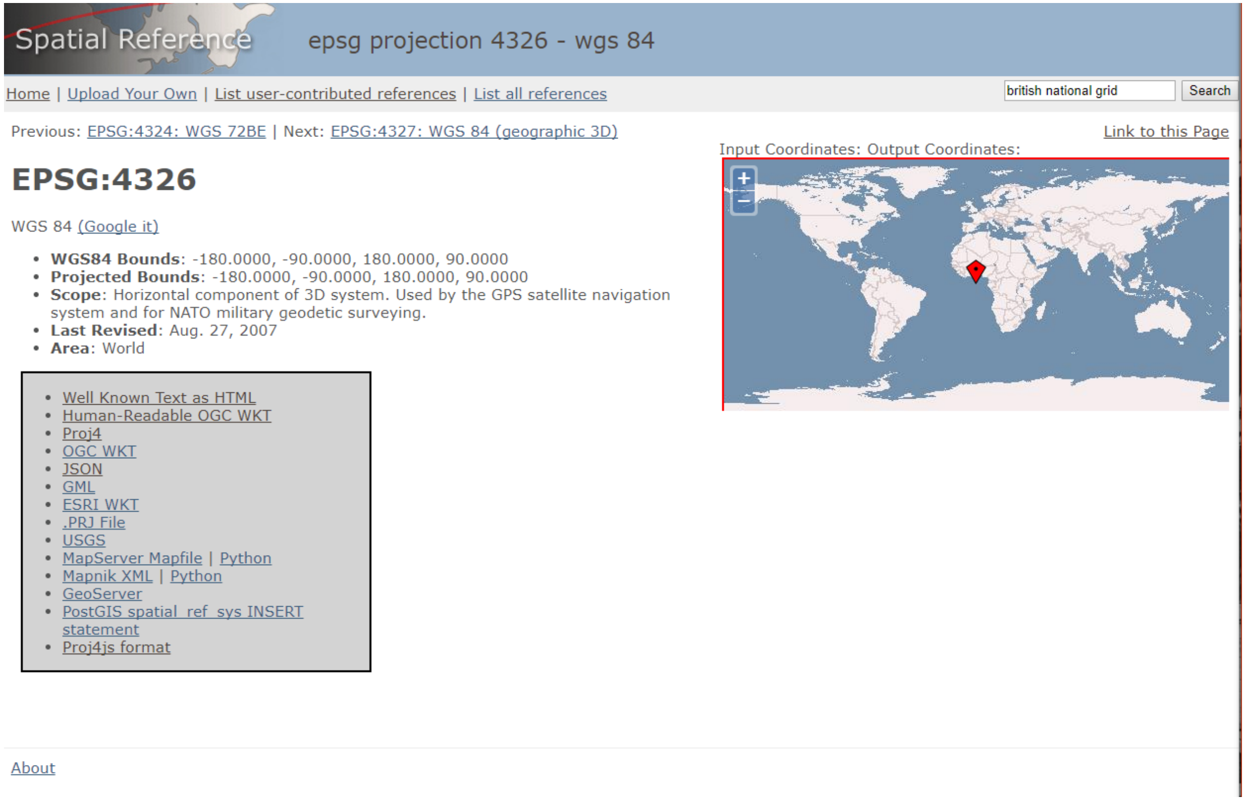
```
+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs
```

- In R there are a few packages that can help retrieve or set the Proj4 string. For example:
    - sp package: after loading sp library, try ?proj4string

```
proj4string(obj) <- value
```

SRID, EPSG and Proj.4

- If you want to find an SRID code for a particular spatial reference system or its related Proj4 string, visit: http://spatialreference.org



BUT WAIT!

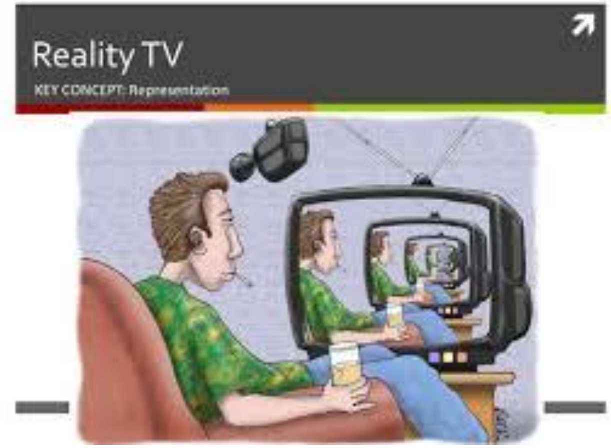
- Some times data won't even have a CRS specified, it's less common now but it does happen.
- In which case it's usually nearly always WGS84
- Setting and transforming a CRS are **different**
- In R:
  - st_crs() - to check CRS of your data
  - st_set_crs() - to assign a CRS
  - st_transform() – transform data to a new CRS
- Setting a CRS won't reproject it, you need to transform it.

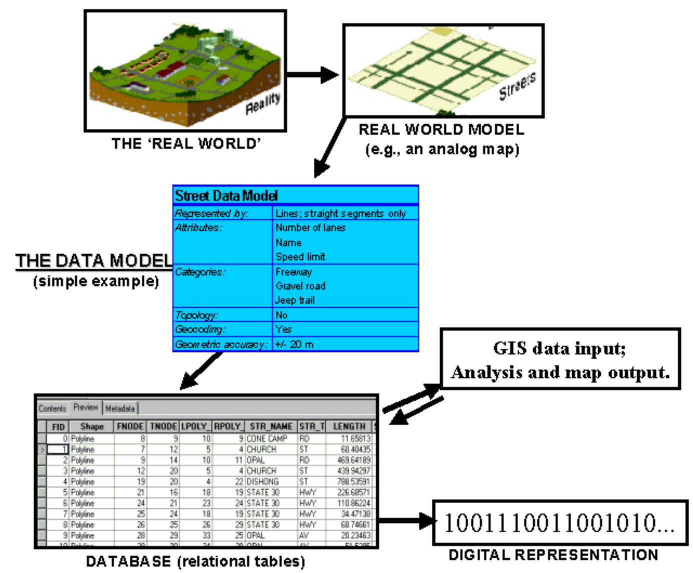
Take home messages

- Know your data
  - Valuable to understand the history of these projections and think critically about data
  - Think about the source of your data and its intended use

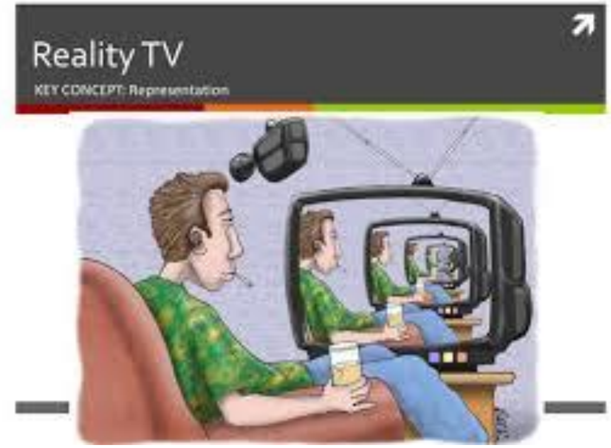
- Know what's appropriate (e.g. data to map / run analysis on)

- Make sure you have set a CRS and all layers have the same CRS

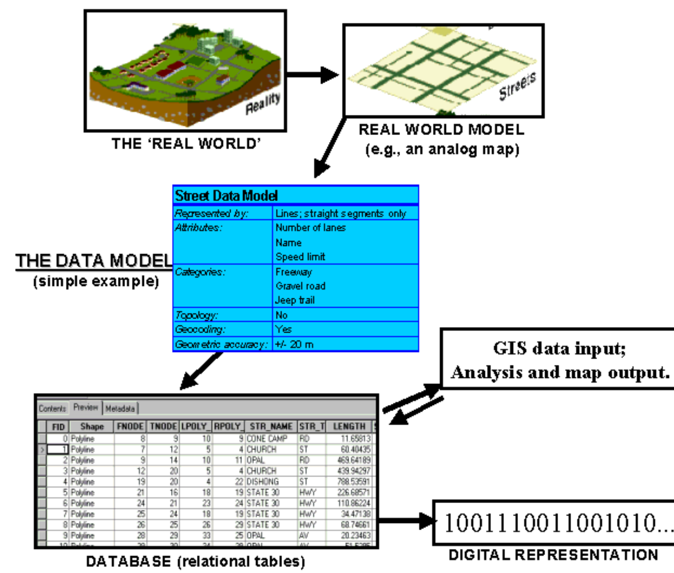




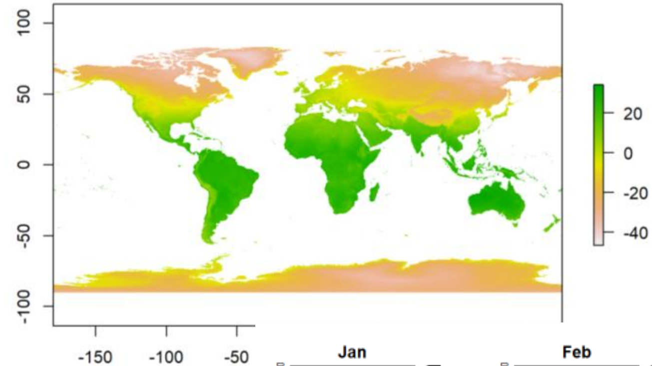
Summary

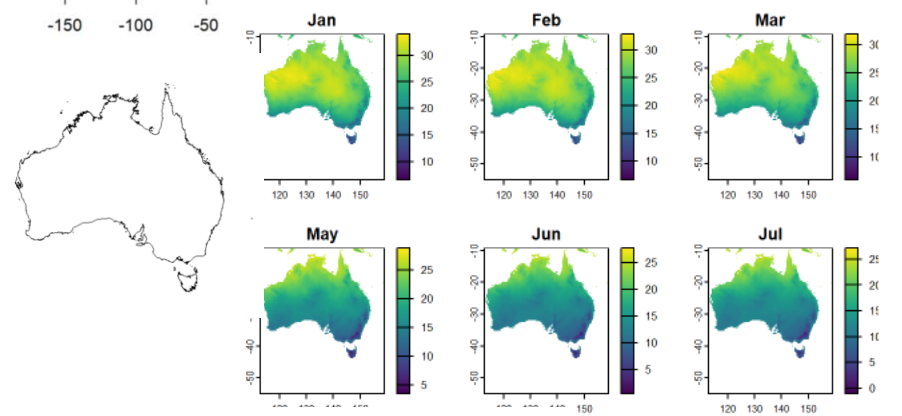


- Reducing the infinite complexity of the read world results in incomplete representations of it.
- Spatial data present a unique and interesting set of challenges.
- There is yet to be a definitive solution to many of them- we will only ever have abstractions from reality!
- We need to decide what is (un)acceptable in our analysis.

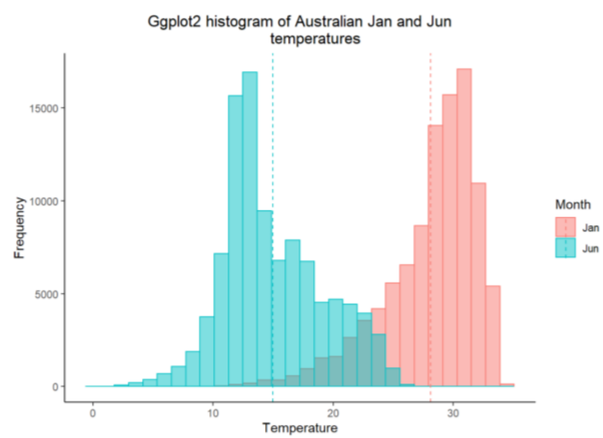


This week's practical – Spatial descriptive statistics (from raster data)





- Work with vector (sf) and raster (terra) data
- Set and transform CRS
- Extract information from raster data and produce basic descriptives
- Understanding cropping, masking, and projecting spatial data



This week's practical – Spatial descriptive statistics (from raster data)

- Pivot_wider vs. pivot_longer

Wide

ID	Site	Jan	Feb	Mar	Apr	May	Jun
1	Brisbane	25.48617	25.23723	24.16489	21.66596	18.79574	15.68511
2	Melbourne	19.73590	20.34103	18.59744	15.59744	13.01795	10.4102
3	Perth	24.44200	24.82400	23.20500	20.06100	16.87400	14.5480
4	Sydney	22.38312	22.65065	21.26104	18.70649	16.00130	13.2000
5	Broome	22.37368	23.05790	24.85789	26.50000	27.56842	27.9526
6	Darwin	28.07302	27.97143	28.05238	28.15397	26.80476	24.82381
7	Orange	19.76100	19.59300	17.03000	12.87400	9.33100	6.2700
8	Bunbury	22.56667	22.90000	21.45555	18.58889	15.76667	13.66667
9	Cairns	27.23152	26.95543	26.28587	24.92283	23.23152	21.17391
10	Adelaide	21.02500	21.63800	19.39500	16.44500	13.75700	11.2600

Each variable (month) has its own column
Traditional GIS attribute table
Good for mapping with tmap

Long

Month	Temp
Jan	27.52000
Feb	27.46000
Mar	27.24000
Apr	27.10000
May	26.72000
Jun	25.78000
Jul	25.14000
Aug	25.14000
Sep	25.90000
Oct	26.76000

One column for variable name, one for value
Each row is one observation
Better for ggplot2 (histograms, bar charts, faceting)

Next week

- To prepare for next week please watch git for humans before the lecture: https://www.youtube.com/watch?v=eWxxfttcMts