



**EUPHEM**

# Introduction to Mapping for Outbreak Investigation

Ente Rood

Adapted from previous lectures: Frantiska Hrubá

,Rebeca Ramis, Héloïse Lucaccioni, Christina Frank, and Matthias Eckardt

Berlin

9 December 2022

# Outline

Once upon a time in Soho...

## Geographic Information Systems Bootcamp

- What are geographic information systems?
- Spatial data types and formats
- Coordinate reference systems

## Key principles of cartography and disease mapping

- Mapping different data types
- Map layout and elements

## Spatial epidemiology

- Issues in disease mapping
- Considerations in the analysis of disease data



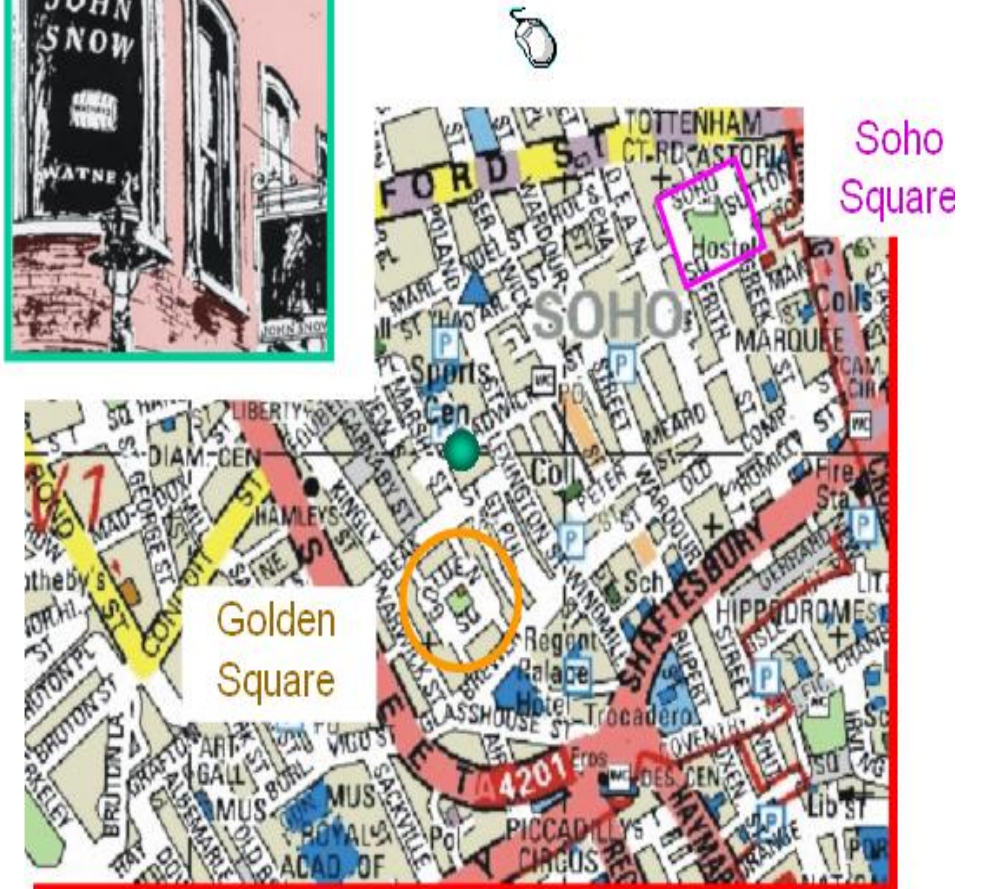
Once upon a time in Soho

Yes,.. it is John Snow,.. with a spatial twist.

# Once upon a time in London in 1854...

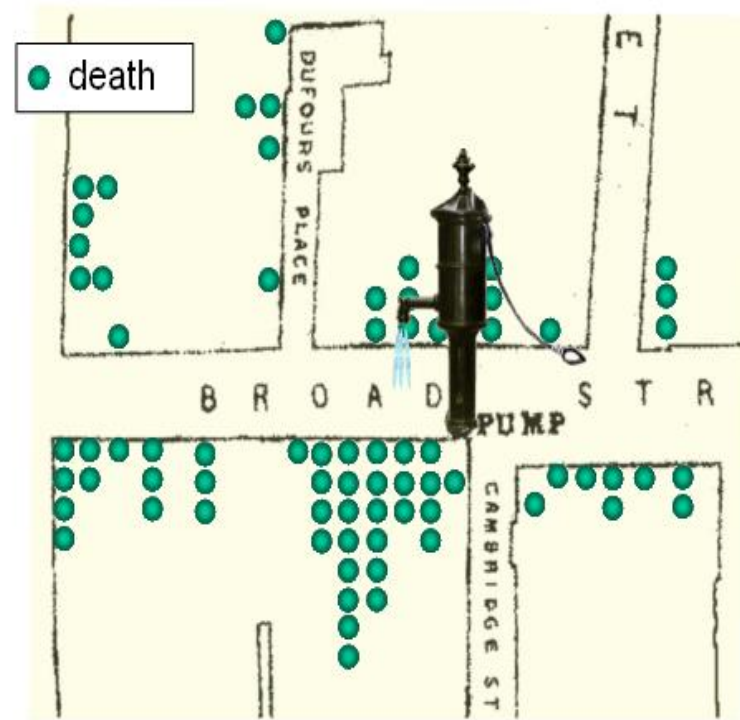
"The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which took place in Broad Street, Golden Square, and the adjoining streets, a few weeks ago."

- John Snow, September 1854



# Describing illness by time, place, person

"...I found that nearly all the deaths had taken place within a short distance of the pump."



10 deaths lived closer to another pump

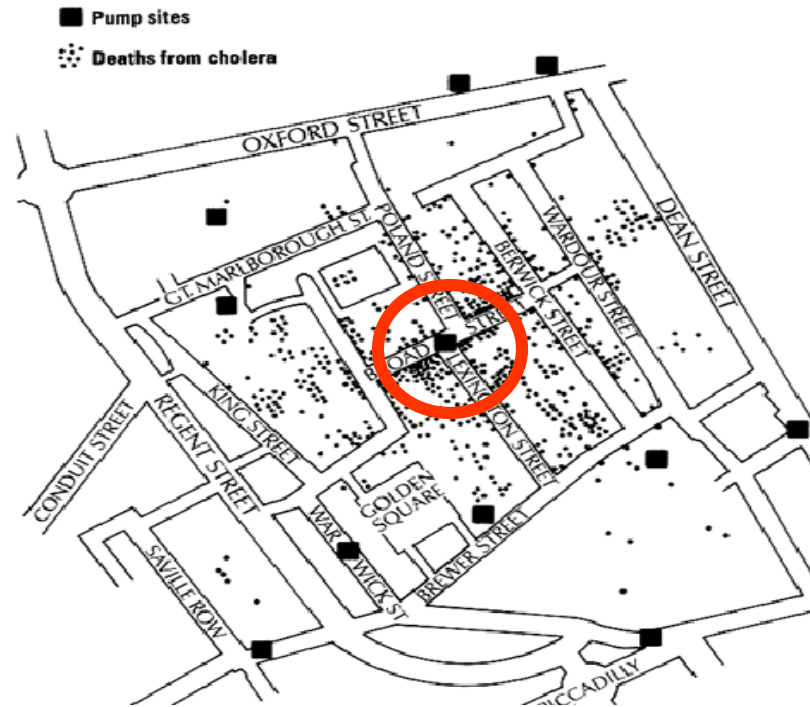
5 deaths always sent to pump in Broad Street

3 deaths were children attending school near the Broad Street pump

Only 6 deaths were not users of the Broad Street pump



All death around the pump had used the water from the pump



Most likely cause is that pump water was contaminated with cholera

# GIS in public health

## What are typical applications of GIS?

- **Manage** process and visualize data
- **Communicating** complex data using maps
- **Analytical** tool to assess geographical processes (distance, dispersion, dependencies/patterns)

## Questions a GIS can help to answer:

- **Location** What is at...?
- **Trend** Where does it concentrate?
- **Inference** Why does it concentrate there?
- **Projection** What if...?

What makes a Geographic  
Information System ???



# What is a GIS?

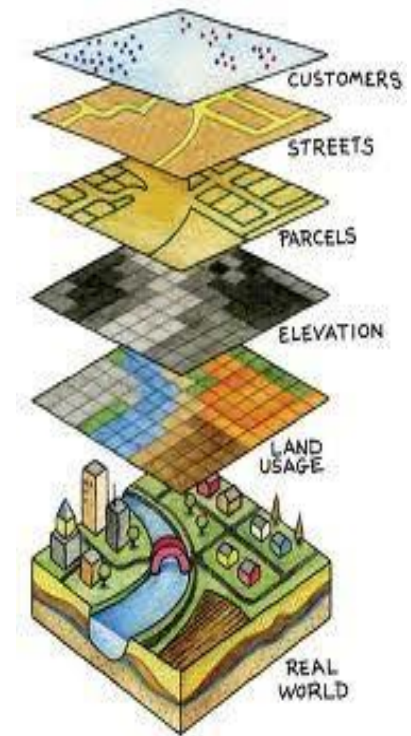
A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, process and present all types of geographical data.

## 1. Information presented in layers

- Point data (patients, Health facilities)
- Line data (infrastructure)
- Lattice data (Administrative units)
- Gridded (imagery, continuous surfaces)

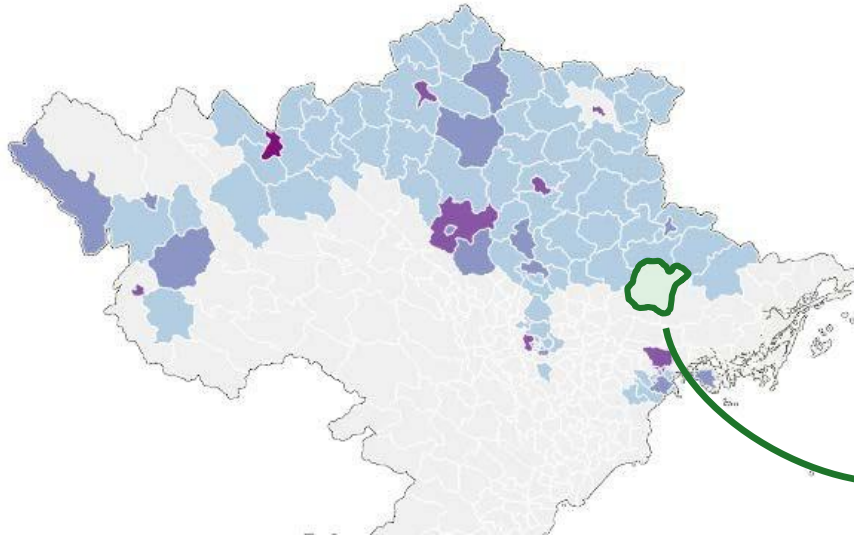
## 2. Analytical use

- Describe patterns
- Compare patterns
- Quantify (spatial) relations
- Predict outcomes



# GIS data components

(vocabulary)



## SPATIAL FEATURES

Graphic spatial representation of real-world physical features

FEATURE ATTRIBUTES (variables)  
Non-spatial data describing the features

Attribute table - Population Density / km2 :: Features total: 148, filtered: 148, selected: 0

	R_NAME	R_CODE	Z_NAME	Z_CODE	W_NAME	W_CODE	COUNT	Pop2007	Area
0	SNIP	7	Gurage	701	Kebena	70101	23	64627.00000000...	297.61073652600
1	SNIP	7	Gurage	701	Abeshege	70102	27	75948.00000000...	559.04725893100
2	SNIP	7	Gurage	701	Ezha	70103	28	106694.00000000...	334.43700377100
3	SNIP	7	Gurage	701	Kokir Gedebano	70104	34	118615.00000000...	548.40098208600
4	SNIP	7	Gurage	701	SODO WEREDA	70105	55	169230.00000000...	881.50248053100
5	SNIP	7	Gurage	701	Meskan	70106	41	200969.00000000...	446.70878981900
6	SNIP	7	Gurage	701	Marego	70107	24	79737.00000000...	251.98999788700
7	SNIP	7	Gurage	701	Gumer	70109	19	100762.00000000...	233.03729960900
8	SNIP	7	Gurage	701	Cheha	70110	42	145705.00000000...	573.15061695200
9	SNIP	7	Gurage	701	ENEMOR ENER	70111	66	210850.00000000...	915.01945298100
10	SNIP	7	Gurage	701	Muhur NA Akil	70112	30	109083.00000000...	472.91334288500
11	SNIP	7	Gurage	701	Geta	70113	16	87276.00000000...	202.00715061800
12	SNIP	7	Gurage	701	WELEKITE TOWN	70114	1	36271.00000000...	10.25990108250
13	SNIP	7	Gurage	701	Butajira Town	70115	1	41974.00000000...	16.12635104590
14	SNIP	7	Hadiya	702	Misha	70201	33	159406.00000000...	362.97967579000
15	SNIP	7	Hadiya	702	Gibe	70202	19	137276.00000000...	447.83265361100
16	SNIP	7	Hadiya	702	LEMMO	70203	32	149048.00000000...	354.36876847700
17	SNIP	7	Hadiya	702	Shashago	70204	31	128794.00000000...	314.88189686500
18	SNIP	7	Hadiya	702	West BADA WOC.	70205	31	104865.00000000...	307.65732036400
19	SNIP	7	Hadiya	702	Soro	70206	47	247236.00000000...	705.89617489900
20	SNIP	7	Hadiya	702	Duna	70207	29	158252.00000000...	222.47243661000
21	SNIP	7	Hadiya	702	Analemno	70208	21	92020.00000000...	224.14245722800

One table row = one spatial FEATURE

One table column = one attribute

# GIS software and Apps

- GIS applications  
free and paid

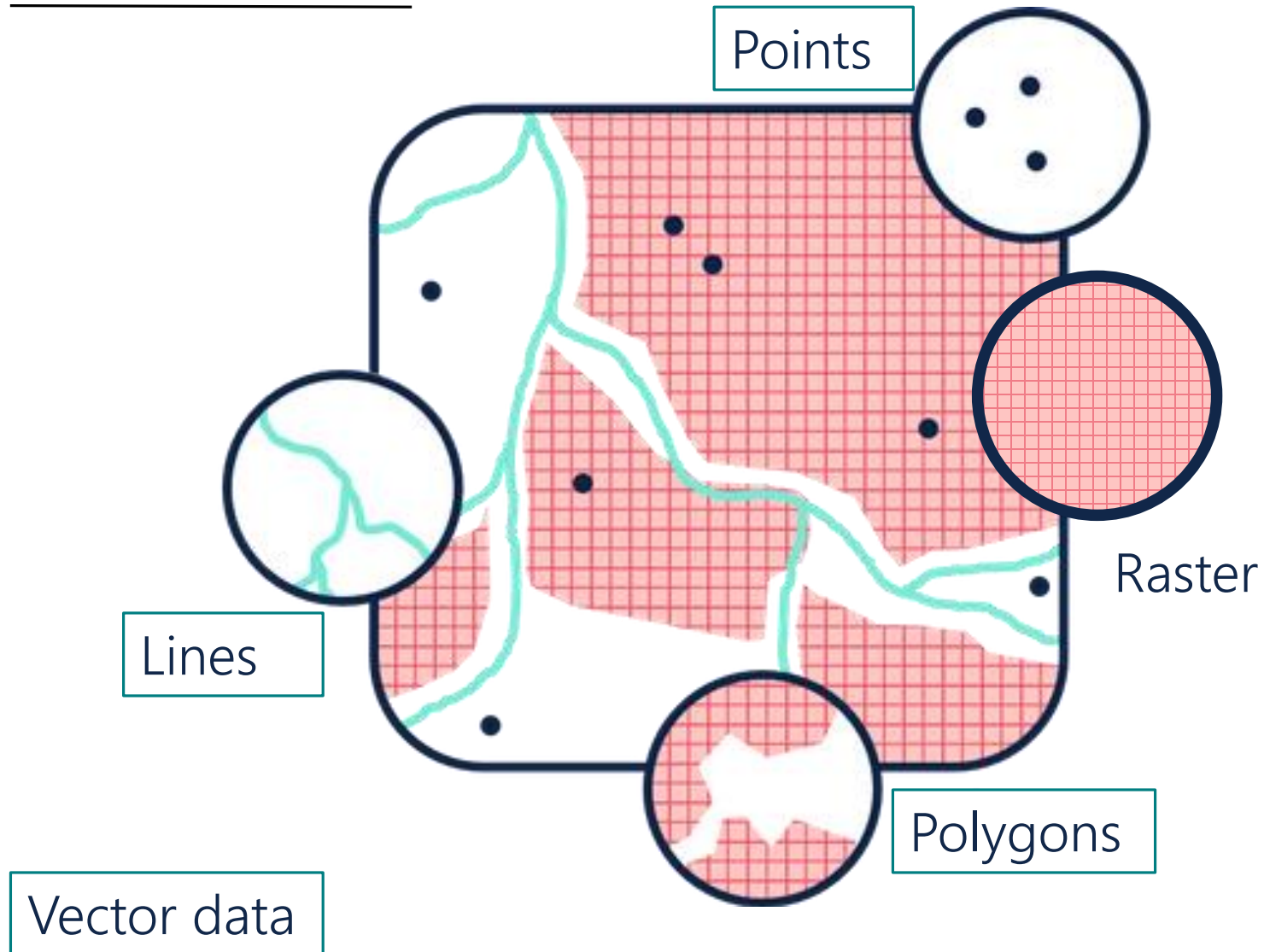


<https://geodacenter.github.io/>



Spatial data - Layers

# Spatial data types



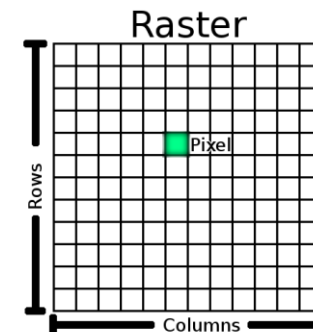
# Spatial data types

## **Vector:**

- Consists of individual points, stored as pairs of X (longitude) and Y (latitude) coordinates.
- These points can be joined to create lines, or enclosed shapes
- The data model behind vector data essentially consists of a long list of XY coordinates along with rules which determine how (and if) these points are joined

## **Raster:**

- Made up of pixels and each pixel has an associated value
- The associated value can correspond to land type, an elevation or rainfall.



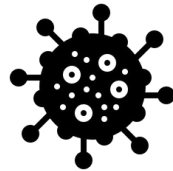
# Points

---

- XY Coordinates: longitude (X) latitude (Y)
- Indicates a location or a position of a point of interest
- Has no area or dimension



Health facility



Case



Settlements

## File format

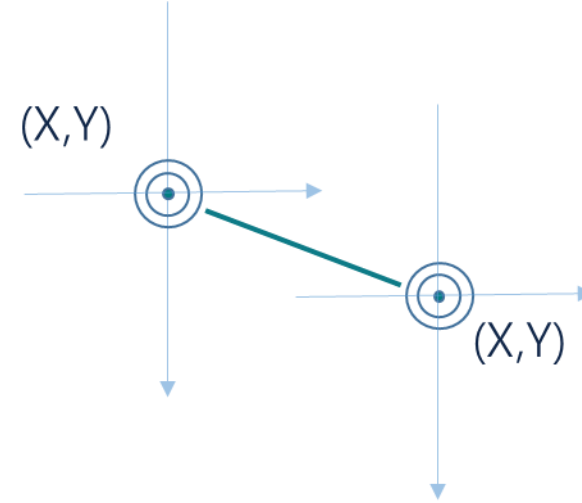
GPS coordinates can be stored in a text file **\*.CSV** or in a spatial data format such as a shapefile **\*.shp**



# Lines

---

- Forms a connection between two points
- Has no width, only length
- Visualizes networks: roads, electricity etc.



- Loading lines in QGIS



## File format

Line type data are stored in a spatial data format such as a shapefile \*.shp





# Polygons

- Lines connected into an enclosed shape
- The start and end point have the same coordinates
- Has an area/surface
- Represents features which cover a surface



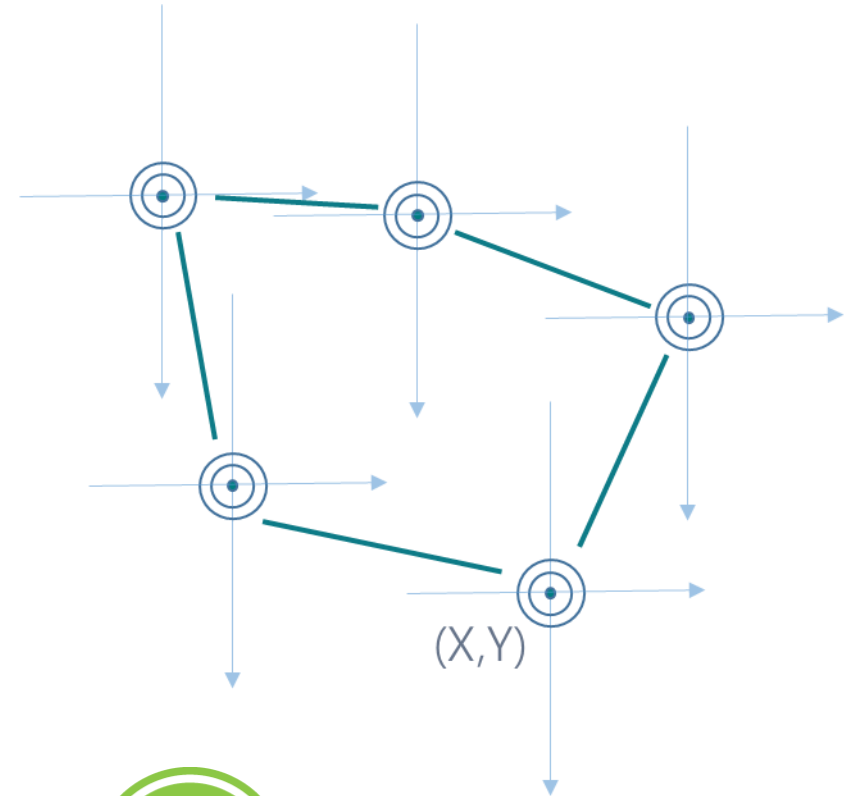
(sub) national boundaries



Water bodies



Land use



## File format

Polygons are stored in a spatial data format such as a shapefile \*.shp



# Lines, points, polygons

You can visualize lines, points and polygons together to get an idea of how certain services are distributed throughout within an area.



# Raster

- Divides the surface into equally sized 'pixels'
- Every pixel contains a value
- This value can be numerical
  - Number of people
  - Elevation
- Or categorical:
  - Land use

1	9	7	9	13	12	1	3	1	4	1	10
13	5	2	5	6	12	12	5	1	8	8	4
6	13	12	8	8	13	5	6	14	2	6	15
3	4	1	0	12	14	4	7	13	12	9	5
5	14	4	3	14	11	6	6	8	2	8	12
5	9	6	7	5	4	9	13	15	4	7	8
2	4	5	11	11	6	5	2	1	4	1	15
2	3	13	4	2	9	12	0	11	6	13	1
14	13	5	6	5	12	13	1	14	0	1	13
0	3	9	9	2	3	5	0	12	0	12	2
9	11	13	12	5	11	10	15	6	13	12	3
11	10	2	13	1	14	9	3	4	8	1	12
5	4	12	5	1	12	11	3	0	4	1	8
8	15	2	12	5	2	2	5	11	8	1	5
12	9	4	7	2	6	8	2	2	15	15	10

6 individuals live  
in the area  
covered by this  
pixel

## File format

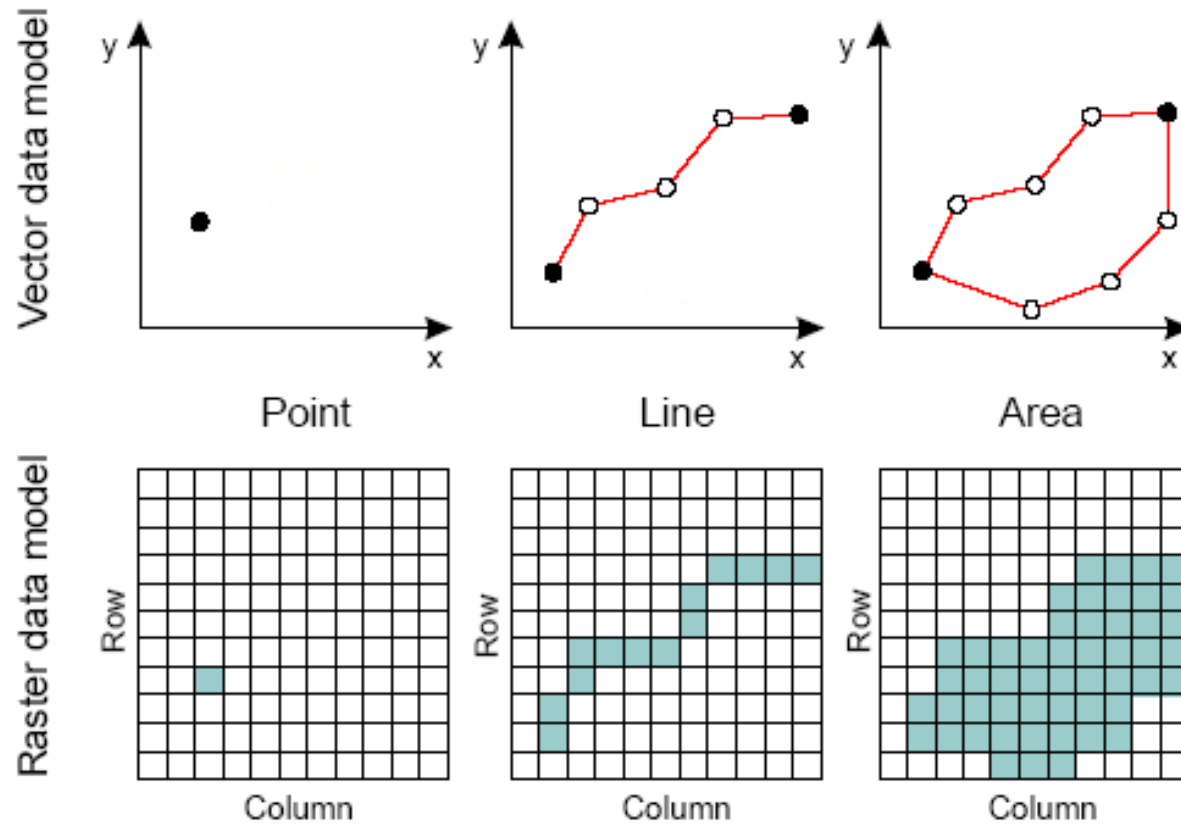
Raster data are often stored in a 'Tagged Image File Format' \*.tiff



# Raster

---

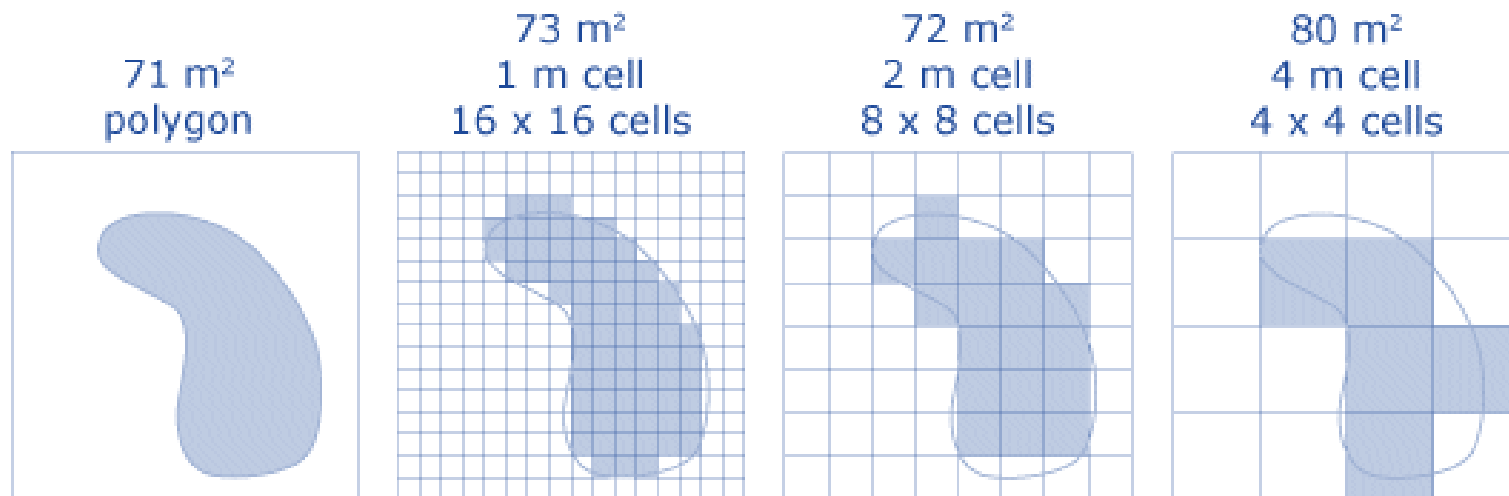
- Raster data can display points, lines and areas



# Raster

---

- The size of the pixel determines the resolution.
- The smaller the pixels, the higher the resolution

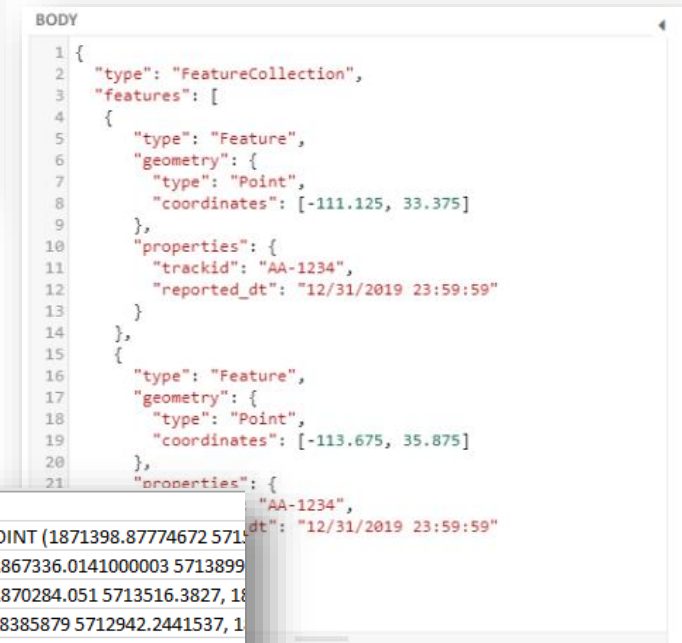
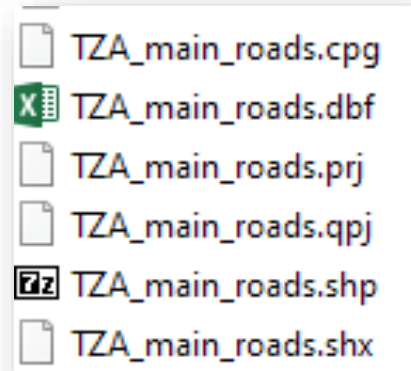


Data Type  $\neq$  File Type

Data formats

# File types - data formats

- Shapefile
  - Consists of multiple files each holding specific information
  - Native to ESRI, used across applications.
- geoJSON
  - Hierarchically structure text file
  - Commonly used in web-applications
- WKT
  - Spatial data stored in a tabular format
  - Used in spatial databases
  - One line per feature, one column holding the spatial data (“geometry”)
- And many more...



	SpatialLocation
	POINT (1871427.98982823 5715934.85121039), POINT (1871398.87774672 571
	POLYGON ((1867375.0352999996 5713908.8962, 1867336.0141000003 5713899
	POLYGON ((1870308.8731000004 5713538.1421, 1870284.051 5713516.3827, 1
	POINT (1869943 5712928), POLYGON ((1869942.58385879 5712942.2441537, 1
	POLYGON ((1838476.84789884 5687993.76778652, 1838469.14305032 5687961
	POLYGON ((1845296.0232999995 5743338.0379, 1845265.1827999996 5743295
	POINT (1885584.04019775 5725586.80404679), POLYGON ((1885668.32589827
	POLYGON ((1839594.57062493 5682788.03998412, 1839583.98204807 5682797
	POLYGON ((1837020.29948424 5685696.94562652, 1837033.9053094 5685663.
	POLYGON ((1869833.36958009 5713280.6899985, 1869809.27094626 5713249.
	POINT (1871185.29368715 5716530.21168081), POLYGON ((1871270.5658 571
	POLYGON ((1867201.54716832 5713978.94730966, 1867207.71753994 5713954
	POINT (1870097.50248512 5712954.59715977), POLYGON ((1870046.70215557
	POINT (1870111.58999994 5713302), POLYGON ((1870101.3407238 5713308.9

# Coordinate Reference Systems

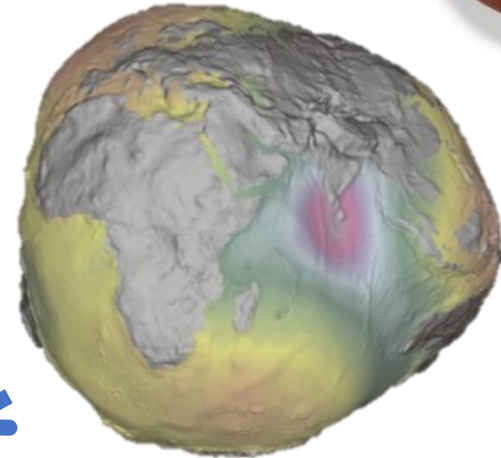


# Flat Earth or Potato



# The Geoid

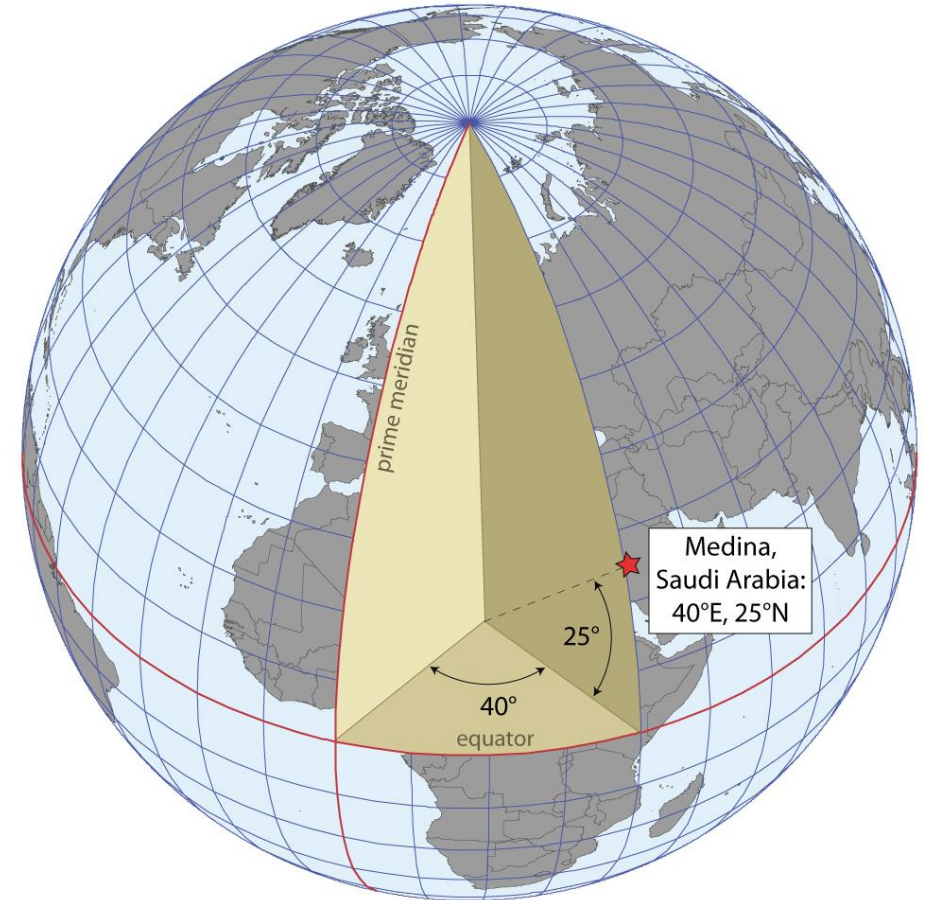
- Earth is an oblate spheroid
- Impossible to map in 2D without distortion
- Different maps for different purposes
- Properties:
  - Area
  - Shape
  - Direction
  - Bearing
  - Distance
  - Scale



# Geographical Coordinates System (GCS)

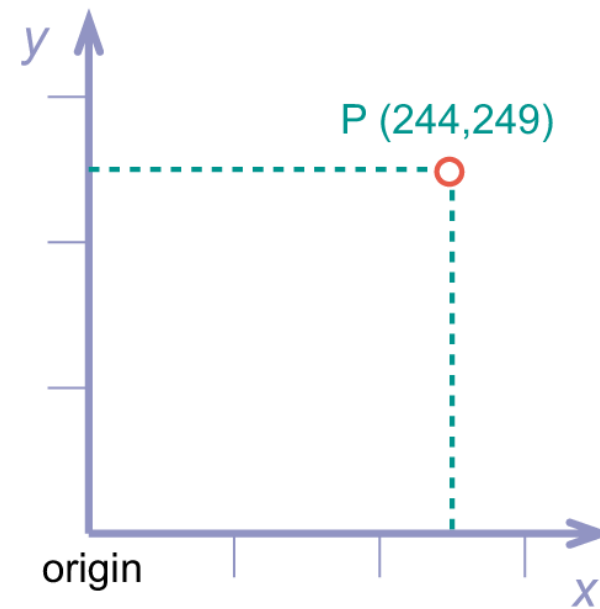
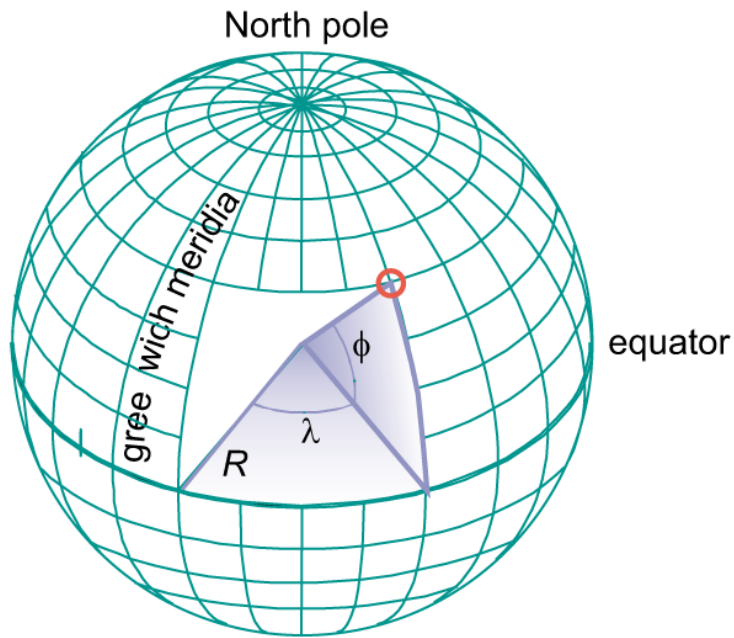
Made up of a **datum**; a **spheroid**, **angular units of measure**, and a **prime meridian**

- **Spheroid**: the object of your attention. The earth in its different representations (geoid, ellipsoid, round, somewhere in between)
- **Datum**: A frame of reference for measuring locations on the surface of your spheroid (i.e. latitude & longitude)
- **Angular units of measure**: degrees, but also Hours, Minutes, Seconds, to measure distance from the...
- **Prime Meridian**: your base point (e.g. Greenwich Meridian Time)



# From Geographical Coordinates to Projections

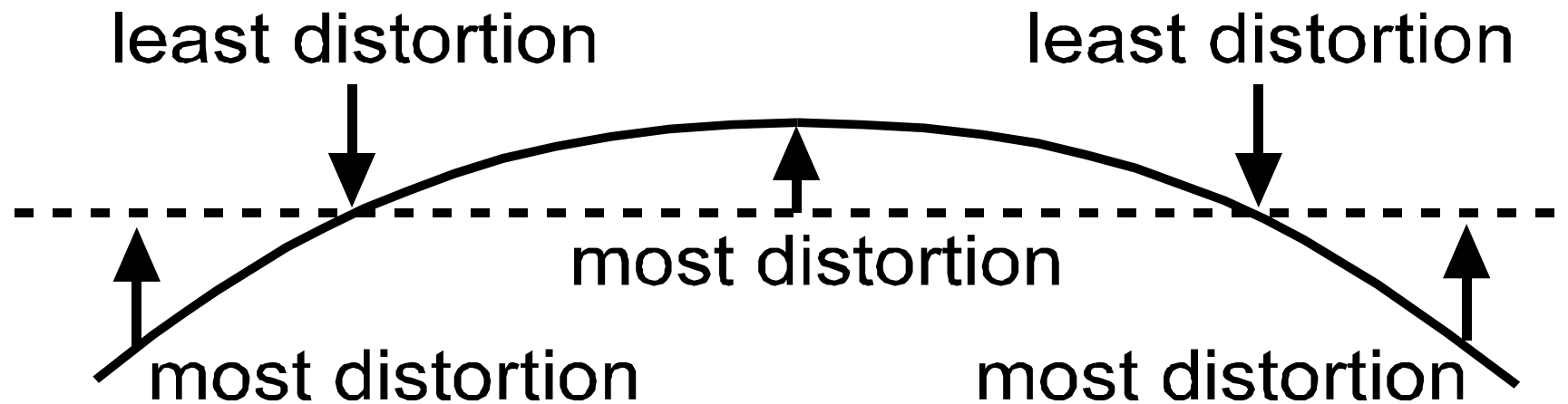
To represent the surface of the earth on a flat paper map the curved surface must be mapped onto a flat plane



# From Geographical Coordinates to Projections

- Geographical coordinate system is **NOT** your Projection.
- **Projection** = a series of transformation from your original point coordinates (from the Geographical Coordinate System, in 3D) to the Projected Coordinate System (flat surface, 2D).
- **GCS= (Spheroid + Datum + Angular Unit + Prime Meridian)**
- **Projection = GCS + Transformation**
  - Shape
  - Angle

# Projection planes distortions



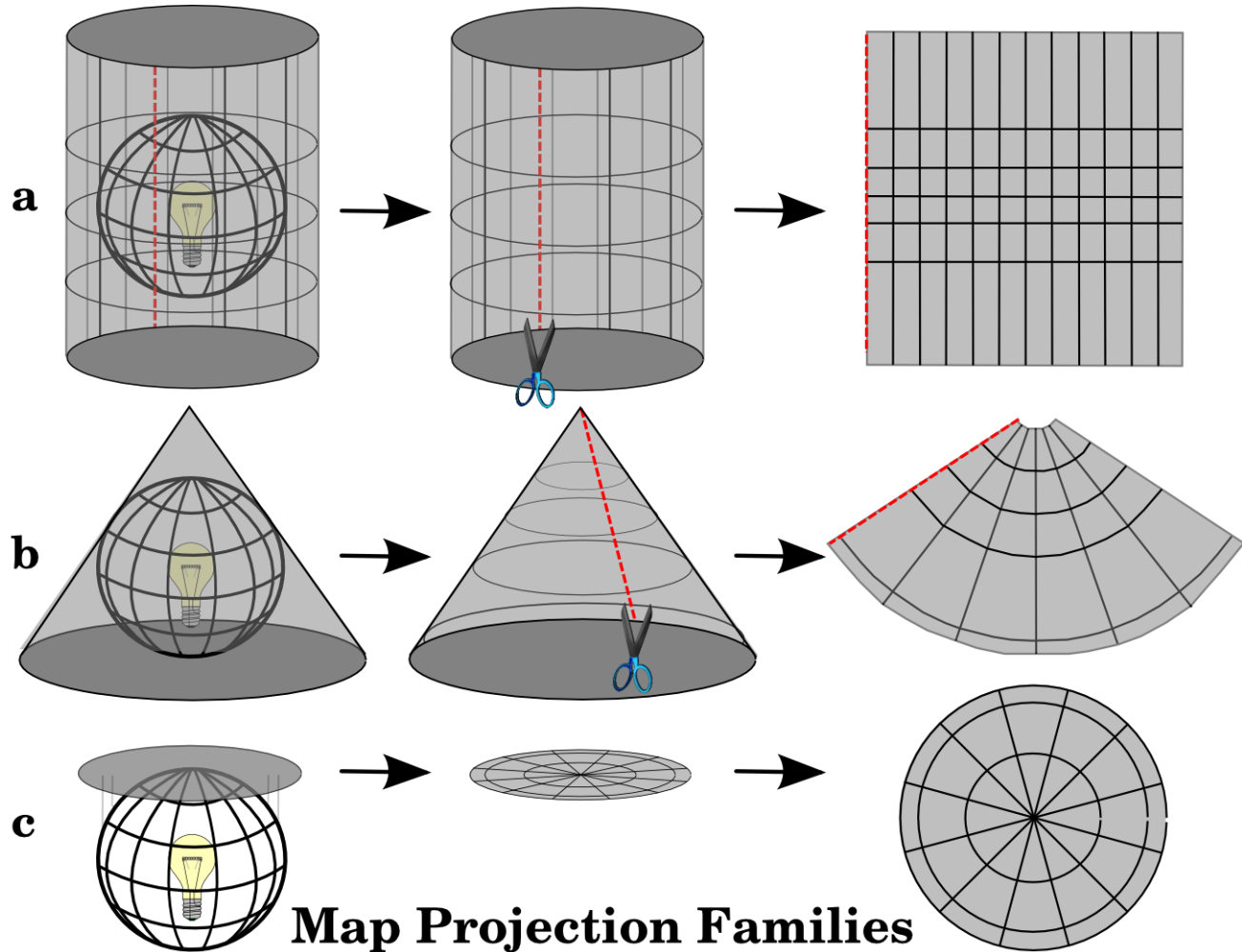
It is not possible to display a round object on a flat surface without introducing **distortions**.

What can be distorted by the projection:

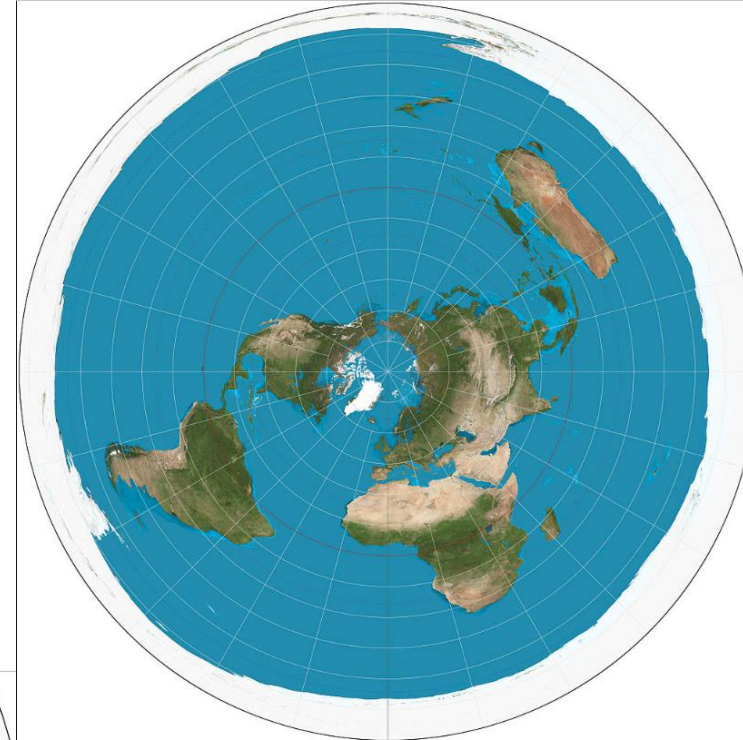
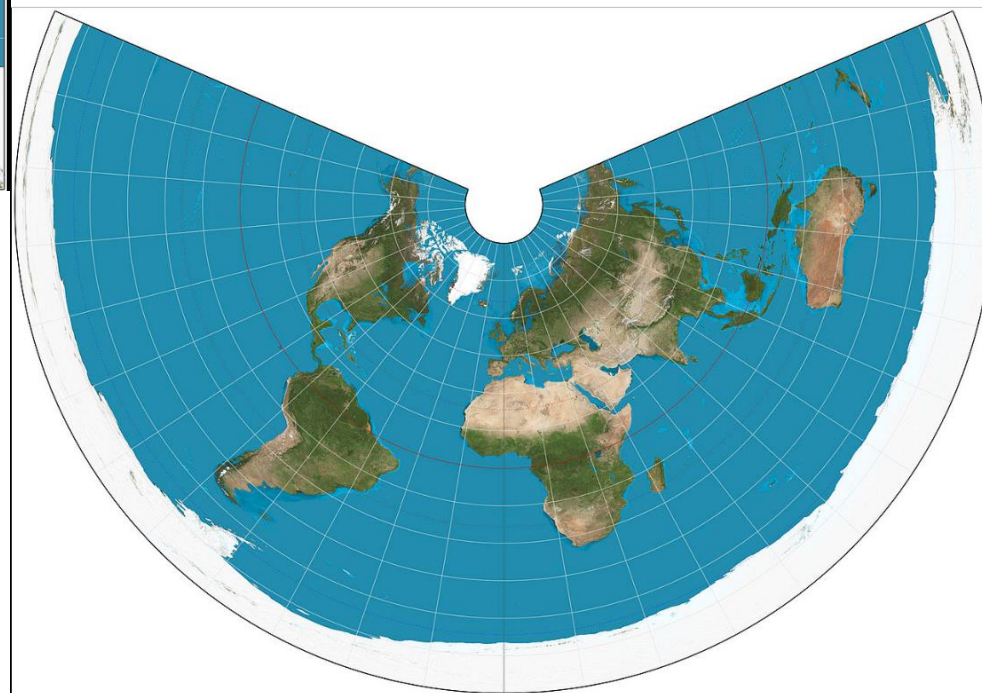
- Distance
- Area
- Scale
- Direction

# The Type of Projections

- A = Cylindrical (i.e. mercator)
- B = conical (tangent)
- C = Azimuthal/Planar
- There are a lot more.









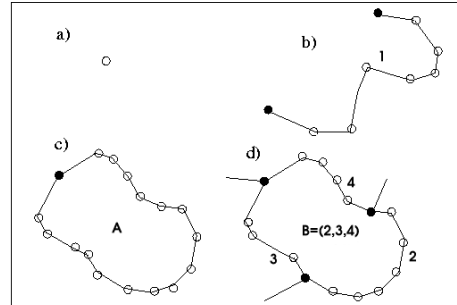
# SPATIAL DATA LAYER



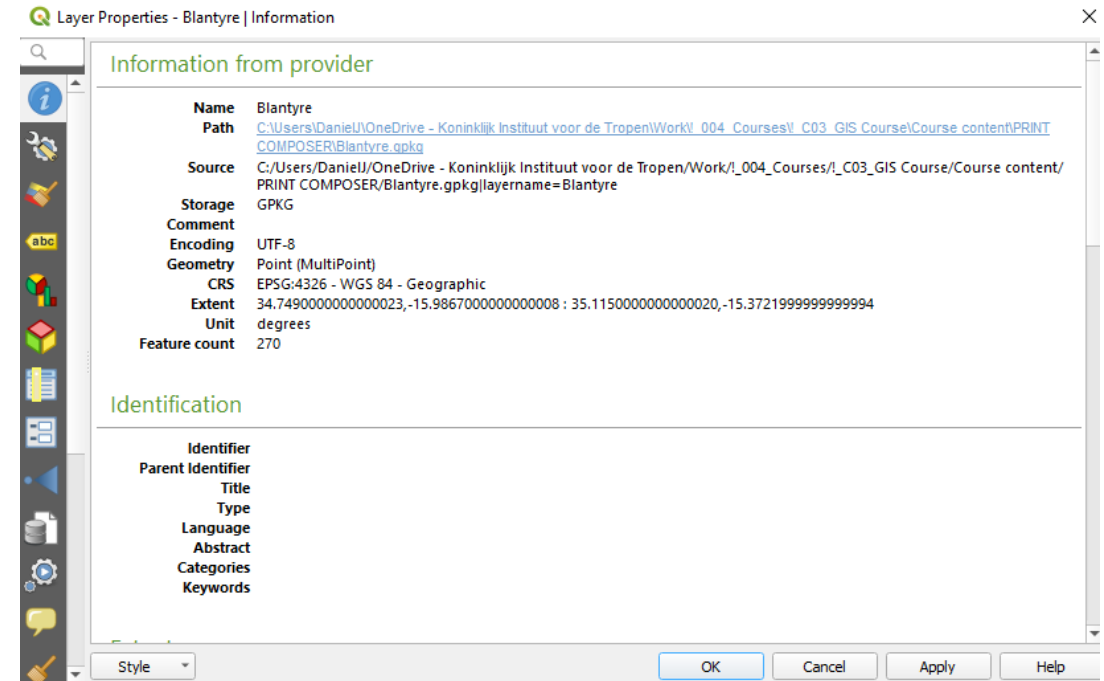
Geographic coordinate  
System Projection file  
**.prj**

Geometry (e.g.  
geographic  
coordinates)  
**.shp**

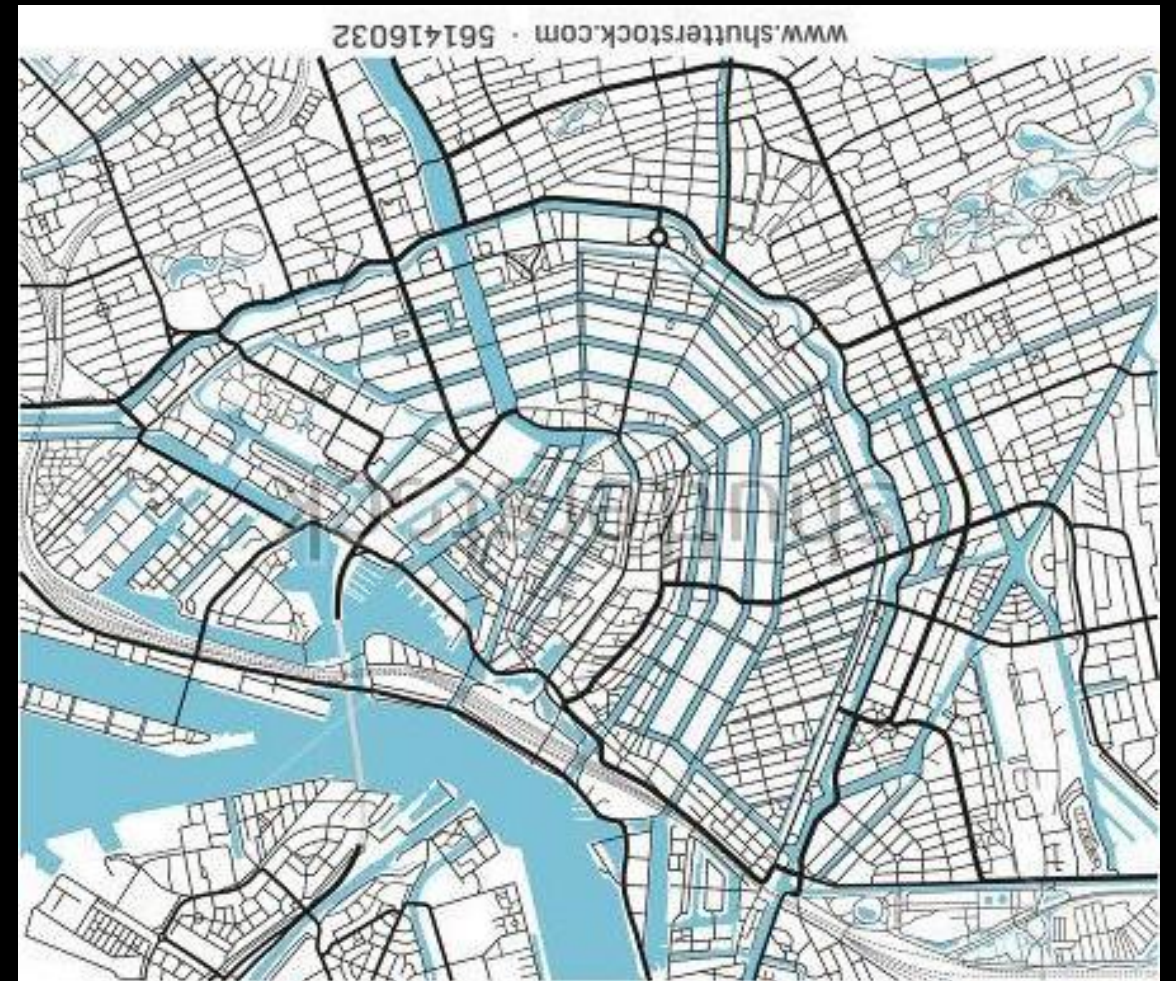
```
PROJCS["NAD83(NSRS2007) / Maine  
East",GEOGCS["NAD83(NSRS2007)",DATUM["D_",SPHEROID["GRS_198  
0",6378137,298.257222101]],PRIMEM["Greenwich",0],UNIT["Degree",  
0.017453292519943295]],PROJECTION["Transverse_Mercator"],PARA  
METER["latitude_of_origin",43.66666666666666],PARAMETER["central  
_meridian",-  
68.5],PARAMETER["scale_factor",0.9999],PARAMETER["false_easting",  
300000],PARAMETER["false_northing",0],UNIT["Meter",1]]
```



# LAYER CRS PROPERTIES







Amsterdam 1650 vs 2017

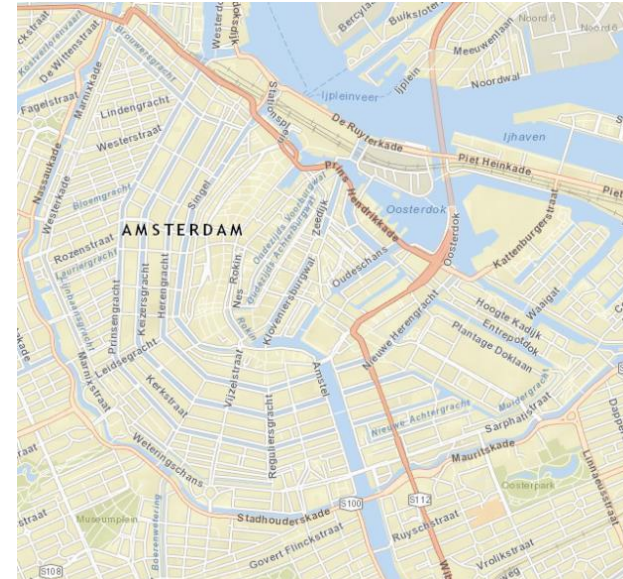
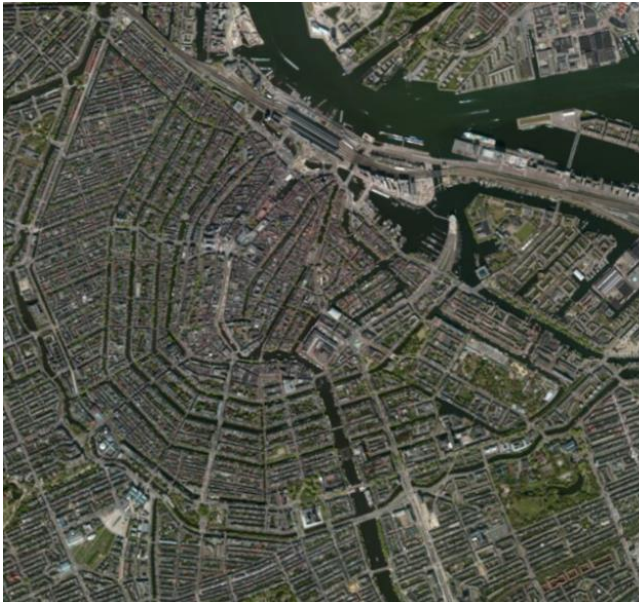
# THE ART OF MAP MAKING



# Maps offer an abstraction of reality

By mapping reality is:

- simplified (generalised, leaving out certain details)
- classified (assign thematic definitions to generalisations)
- symbolised (color, shape and texture)
- additional information added (street names)

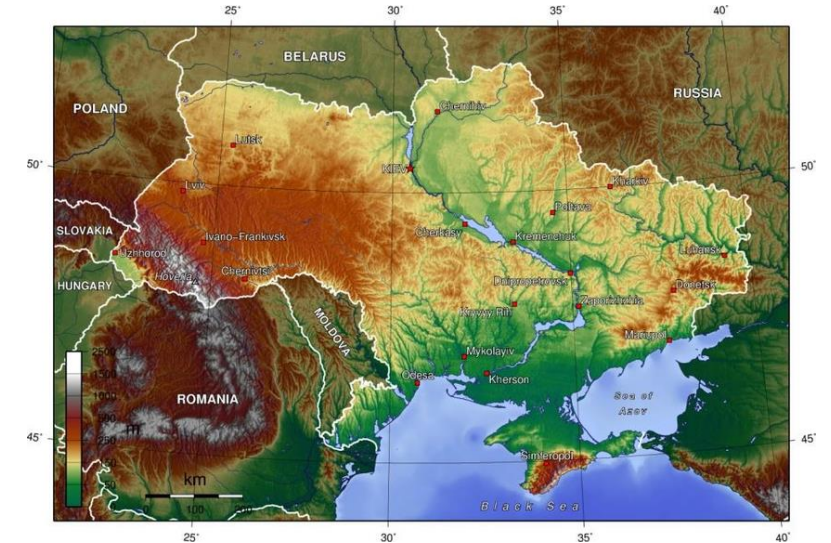


# Topographic vs. Thematic maps

*Traditional distinction in maps*

## 1. topographic maps

accurate representation of the earth's topography



## 2. thematic maps

one or more particular themes are emphasized  
used to visualize single phenomena like election outcomes

*Distinction not so relevant in a digital environment*



# Data types in mapping

## NON SPATIAL

### NOMINAL DATA

- Discrete classes without ordering
- No quantitative nature
- Qualitative

### ORDINAL DATA

- Ordered classes
- Discrete Scores with to quantitative relation between them.

### CONTINUOUS DATA

- Ratio vs interval
- Continuous vs discrete

## SPATIAL

### Point pattern

- Individual cases or events
- Have no magnitude!

### Lattice data (areal)

- Counts of cases
- Rates of disease

### Geostatistical data

- Point measurements of a continuous surface
- Measure of continuous data at point location
- Used for geostatistical interpolation (grids)

Seminal work by Cressie (1993)

# Bertin's visual variables

## Size

- use for **absolute counts or rates**

## Value (lightness)

- **use for rates**, **not** for discrete data!

## Texture

- use for **discrete/nominal**

## Hue (color)

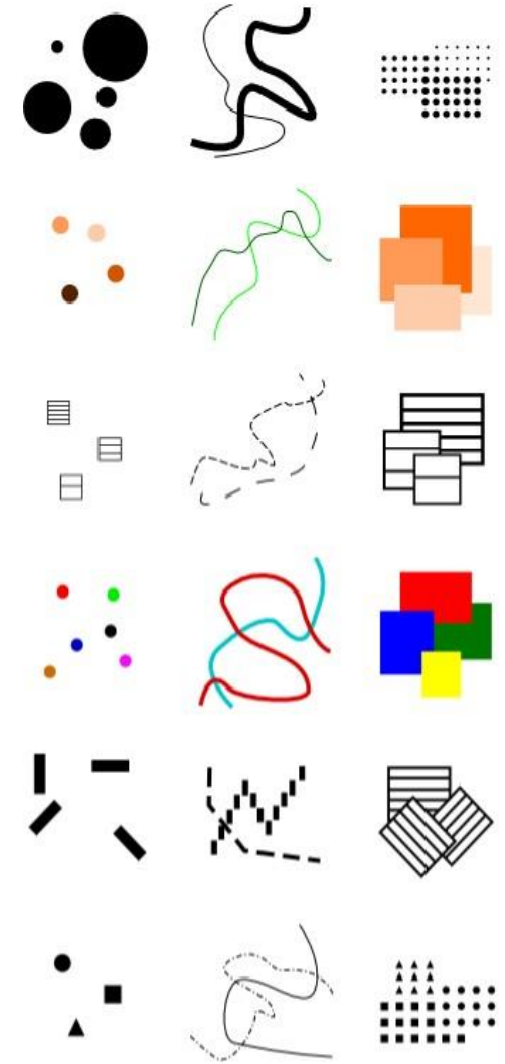
- use for **discrete/nominal**

## Orientation

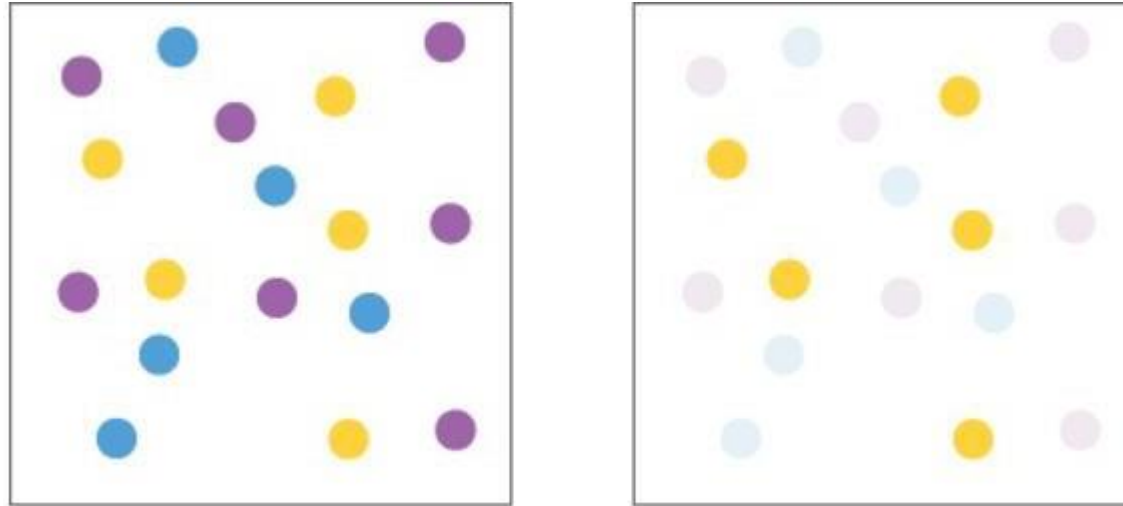
- Use to indicate angular direction

## Shape

- use for **discrete/nominal**



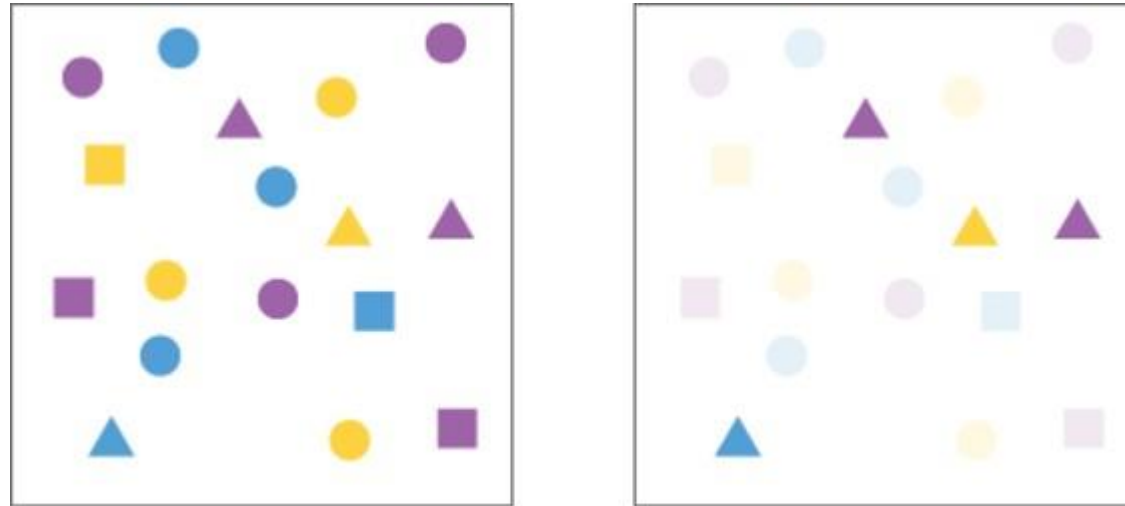
# MAPPING NOMINAL DATA I (groupings)



We can quickly and easily perceive a group of symbols based on color hue, e.g., the yellow symbols appear as a group. Therefore hue is *selective*.

**Color** differences (hue) allow you to **perceive groups** of similar features

# MAPPING NOMINAL DATA II (distinct)



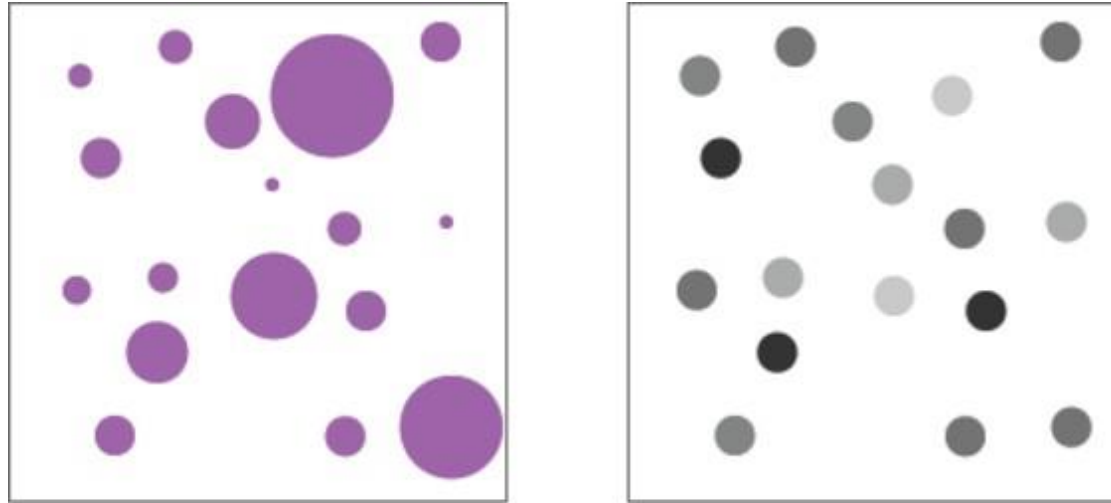
It is not easy to isolate symbols based on shape. Even with all the triangles pulled out at right, they still don't look like a group. Shape is not selective.

**Shape** differences do **NOT** allow you to perceive groups of similar features and should not be used to visualize patterns of similar features.

**Shape** variation can be used to distinguish points without emphasizing groupings



# MAPPING ORDERED DATA (ORDINAL)

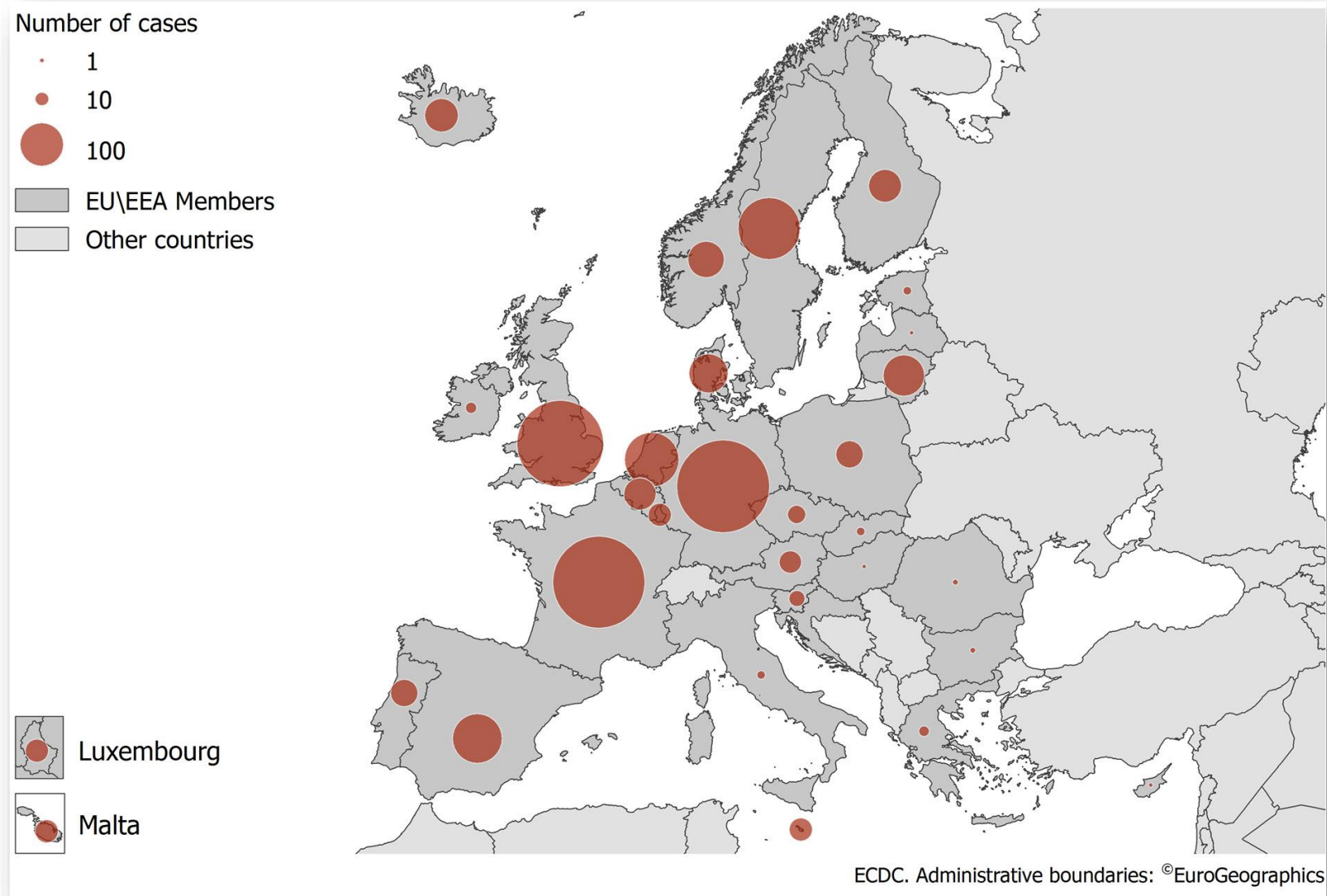


With both size and value, it is immediately obvious that there is some sequence to the symbols (small to large, light to dark). Size and value or *ordered*.

**Size** and **value** differences allow you to perceive **ordering** in your data

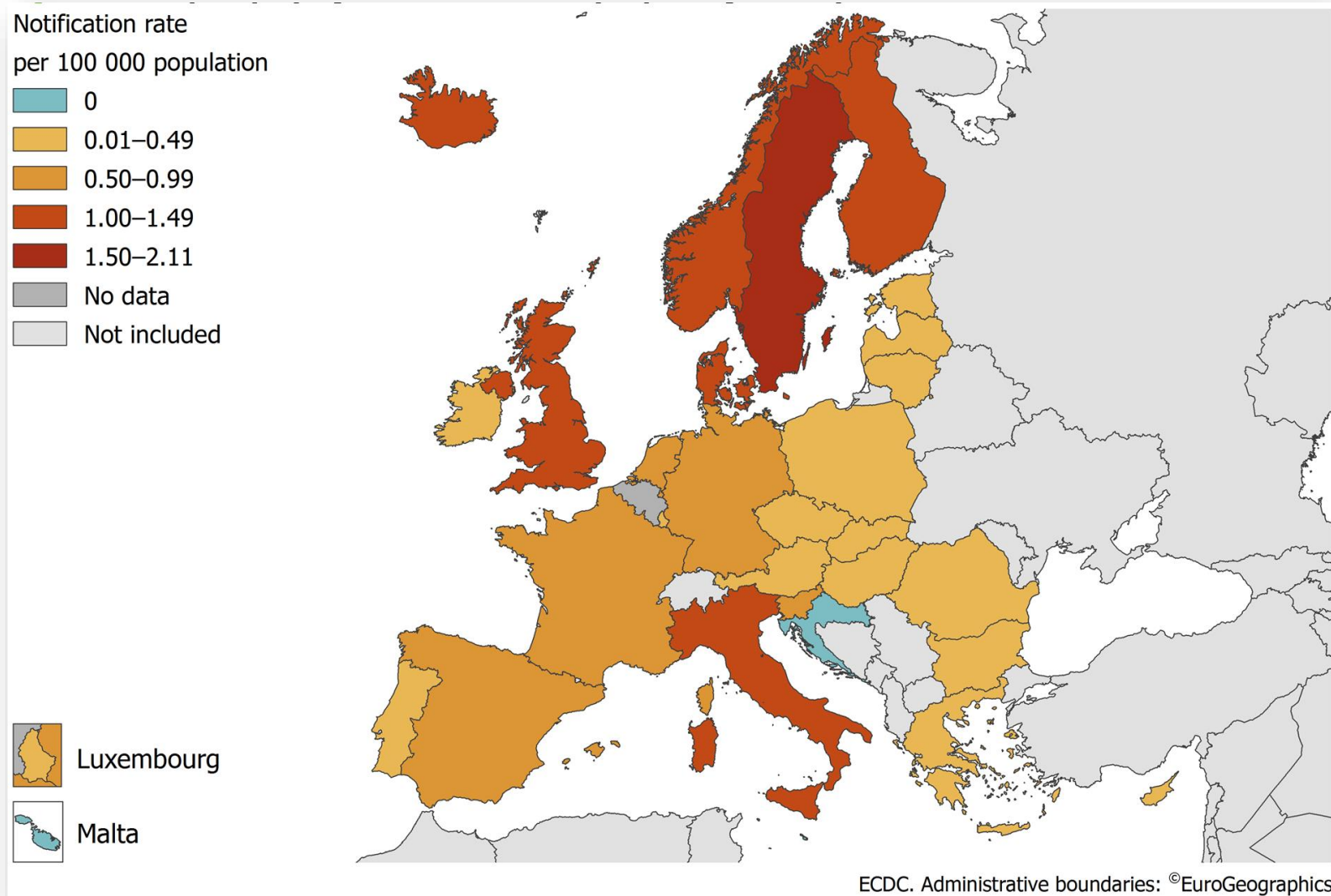
Examples of thematic maps

# Proportional symbol map



# Choropleth map

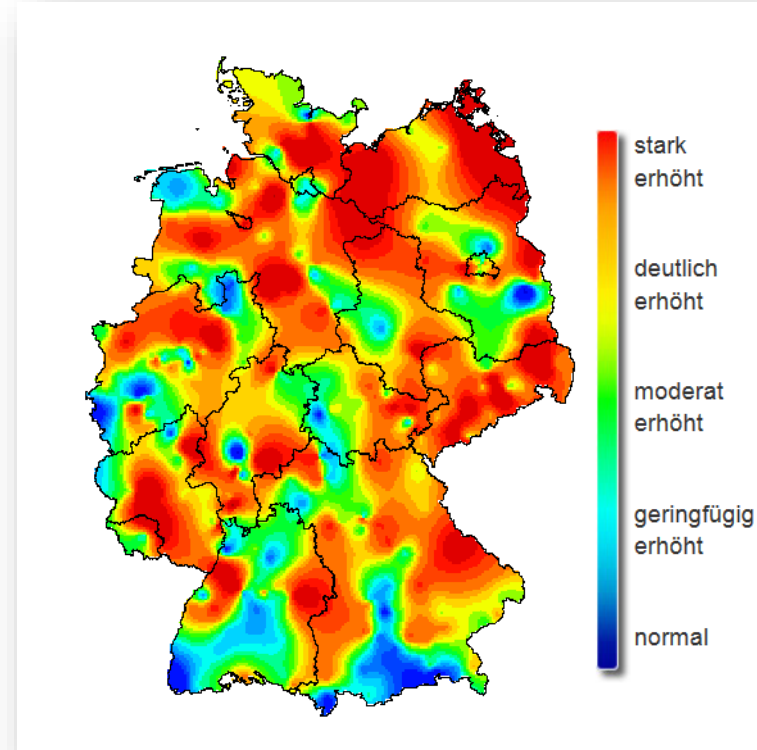
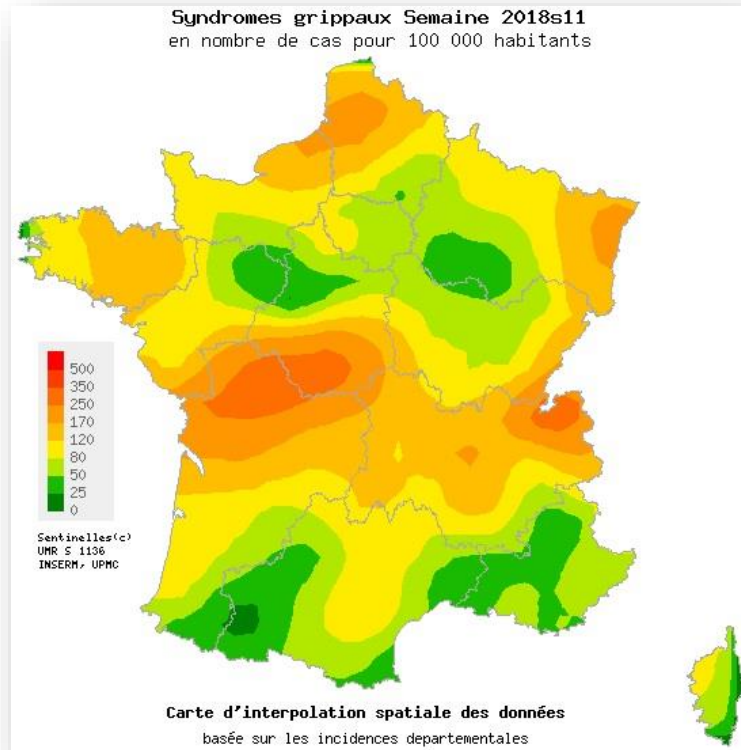
## Manual data classification



Source: <https://ecdc.europa.eu/en/publications-data/guidelines-presentation-surveillance-data>

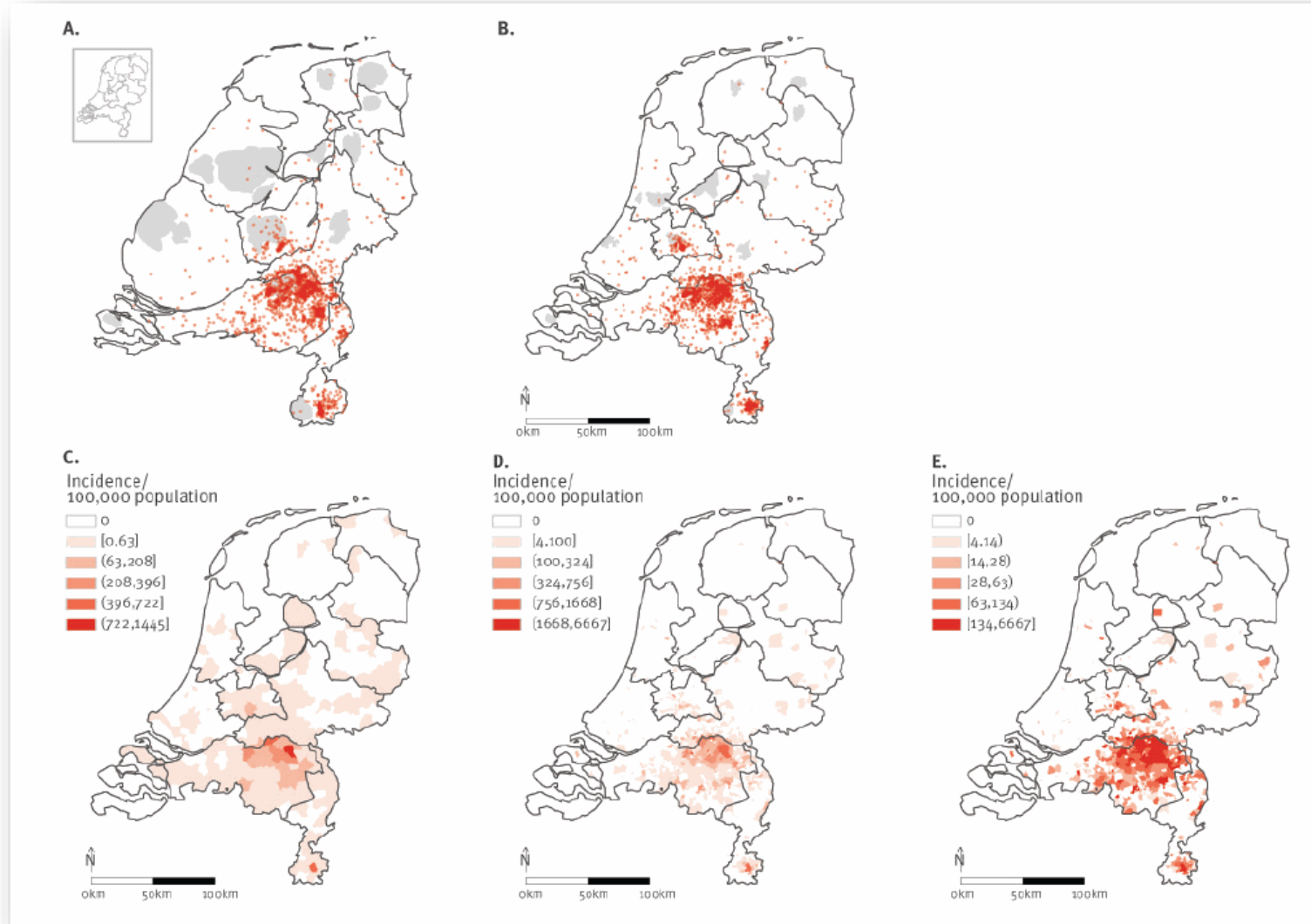
# Isopleth maps (heat maps /risk maps)

- To represent a gradient derived from sentinel locations
- Data smoothing (Kriging algorithm), not real data
- Equal intervals data classification of indicator values



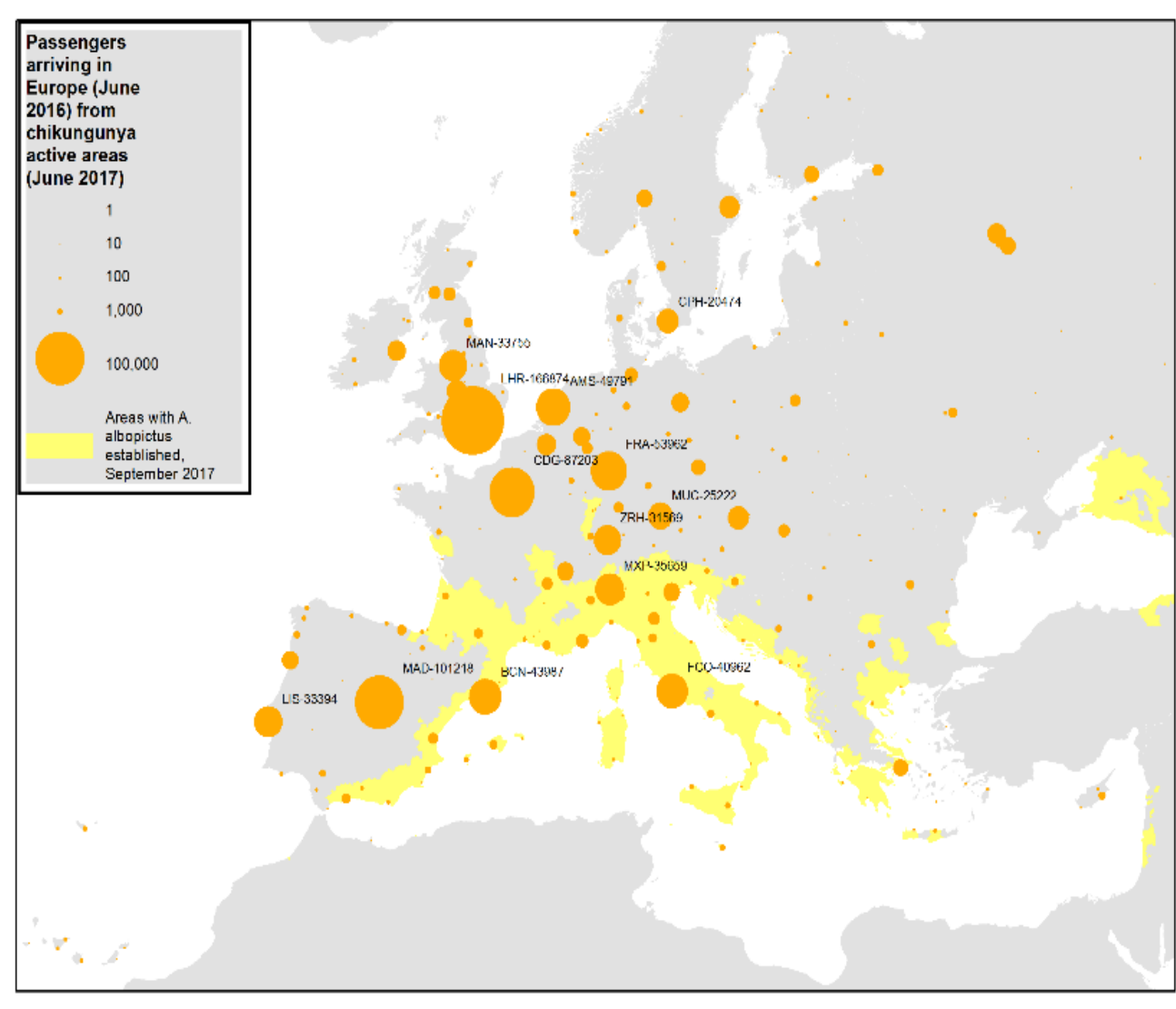
# Q fever outbreak, The Netherlands

Five types of maps, the same data



# Composite maps – multiple indicators

Single symbol & Proportional symbol



Disease determinants for establishment of a local transmission:

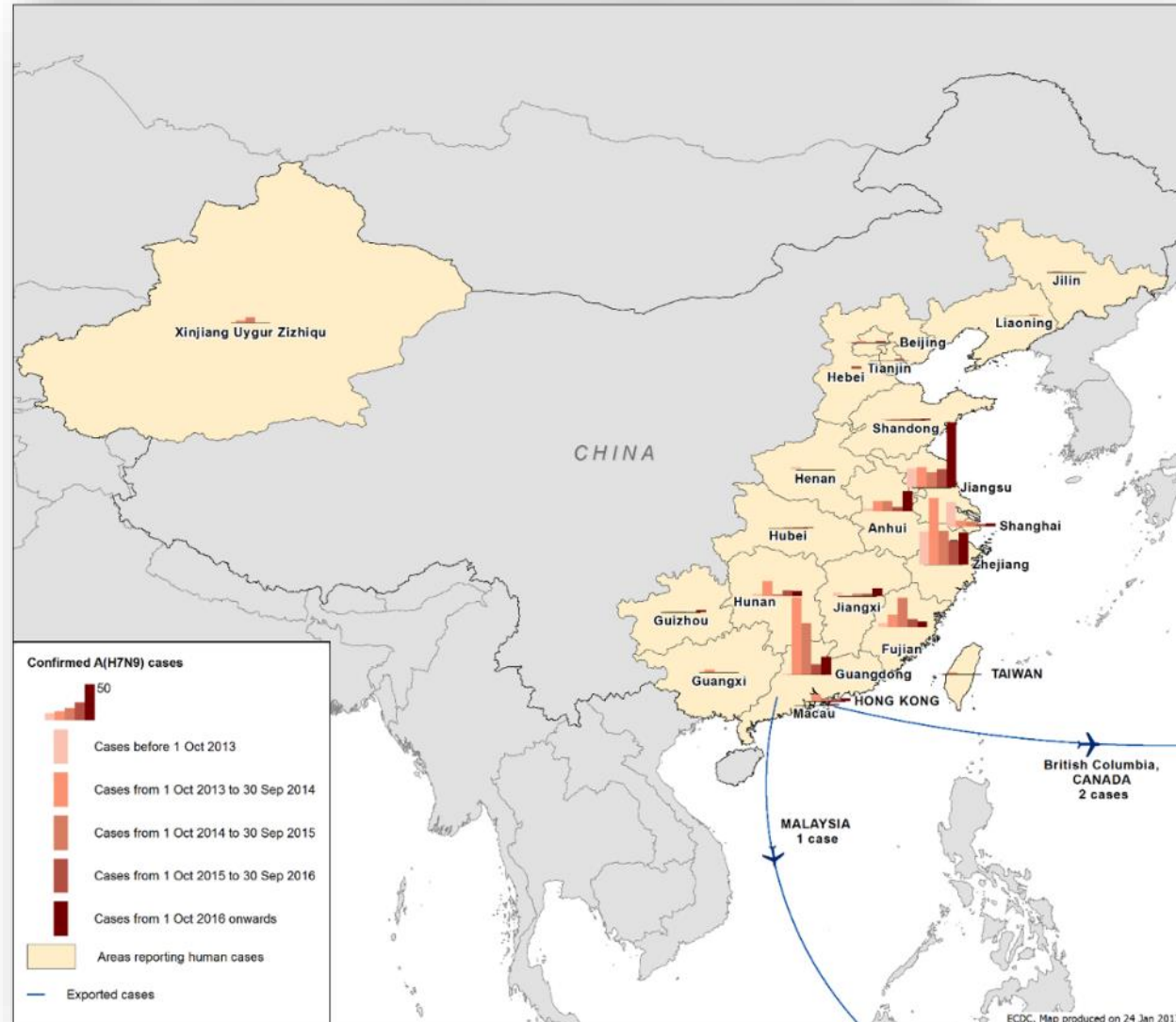
- Areas where *Aedes Albopictus* was established
- Number of air passengers arriving from countries experiencing chikungunya transmission



# Composite maps

## Single symbol & Bar chart (epicurve)

Distribution of confirmed cases of A(H7N9) by five periods(weeks 07/2013 to 4/2017)

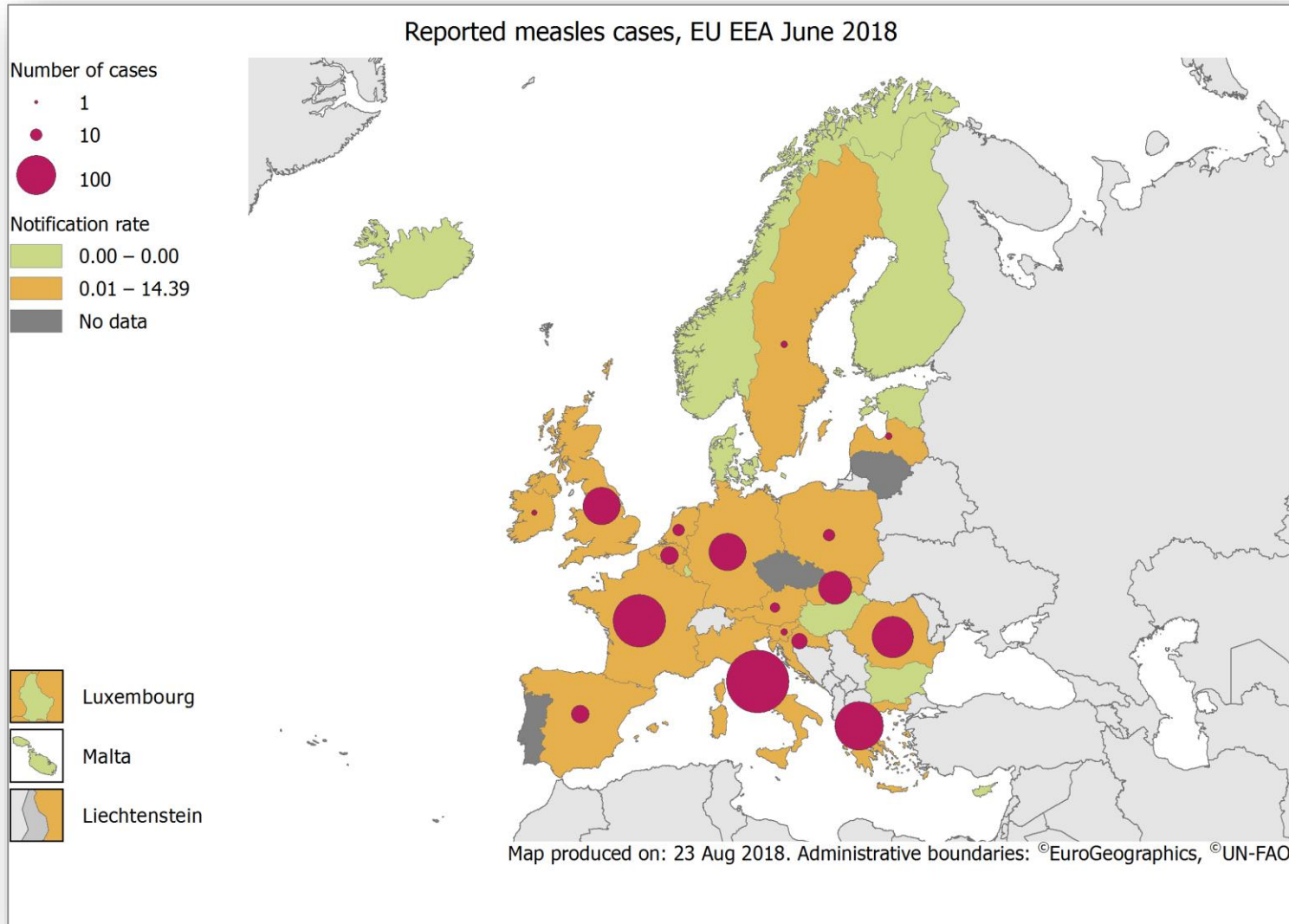


Burden of new emerging disease:

- Number of confirmed A(H7N9) cases per season
- Areas reporting human cases



# Map checklist



- **Self-explanatory**
- **Minimum digital/text info**
- Title
- Map
  - Base map
  - Symbol selection for data layer
  - Colour(s) selection
  - References
  - (Disclaimers)
- Legend
  - Names of layers
  - Layer data units
  - Number of classes
  - Break points / non-overlapping values
- Invisible areas
- Date of production



# What visualization to use?

## Depends on:

1. Aim of the map (think about it before you start!)
2. Message you want to communicate
3. Target audience
4. Data available
5. Resources you have to take action, based on the conclusions of your map

# Spatial epidemiology

# Spatial Epidemiology

## Objective

### Disease mapping

- Understanding distribution of disease and identifying inequities in health

### Inferential analysis

- Relating health outcomes to (socio-economic) environmental determinants

### Health service analysis

- Analysis of supply and demand
- Resource allocation and optimization

## Methods

### Geovisualisation:

- Visualising spatial trends spatial distribution: Are there areas with higher risk than others?

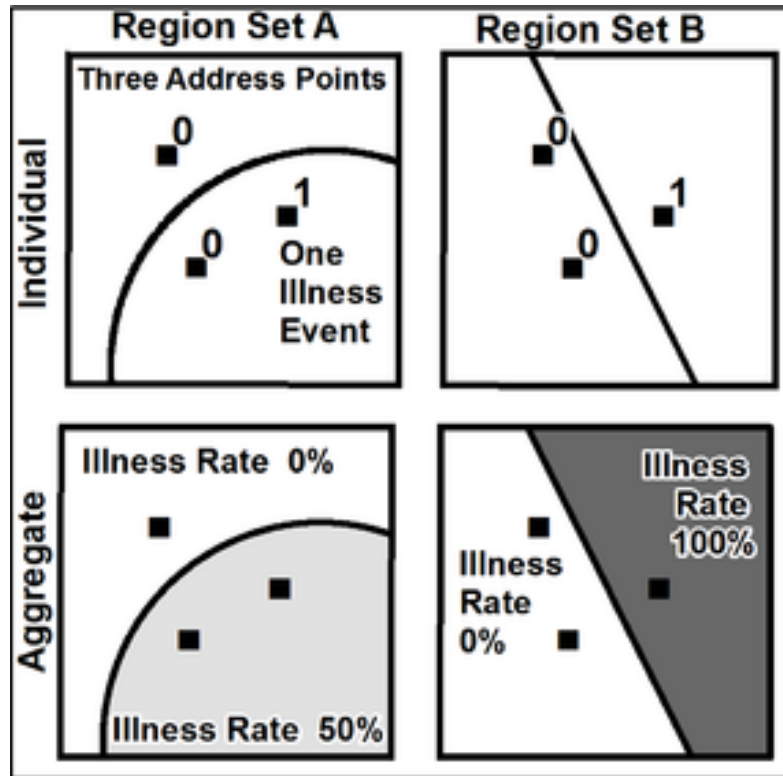
### Spatial analysis (statistical):

- Spatial trends: Are single events or, high risk areas geographically clustered or randomly spread
- Spatial association between determinants of health disease outcome

### Geoprocessing (non - statistical):

- Access and coverage analysis
- Network analysis and routing

# Disease rate calculation over aggregated data



Depends on regional delineation

Distortion more likely within smaller areas

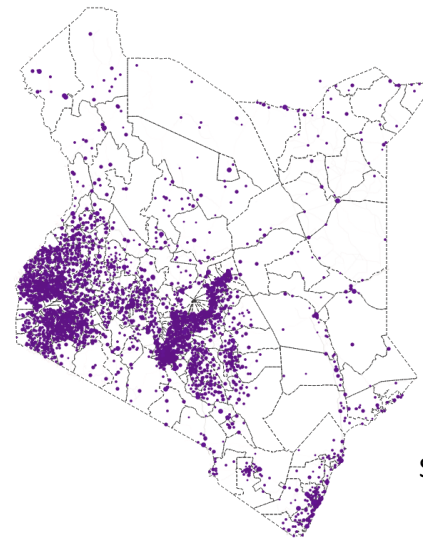
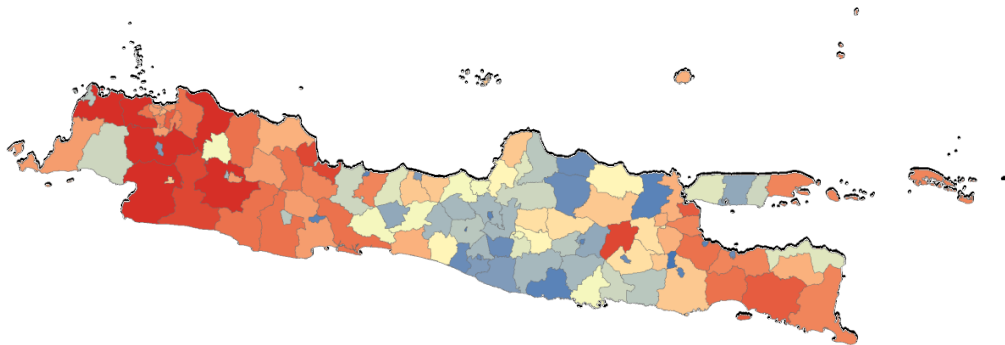
**Modifiable areal unit problem**

*As the area decreases the denominator population decreases and rates become binomial outcome*

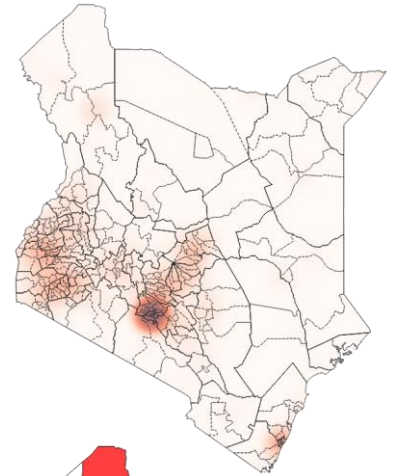
# Common issues in spatial analysis of (disease) rates

- ***Relative risk estimation notoriously unstable. Small expected count can lead to large diseases rates. Creates spatial outliers***

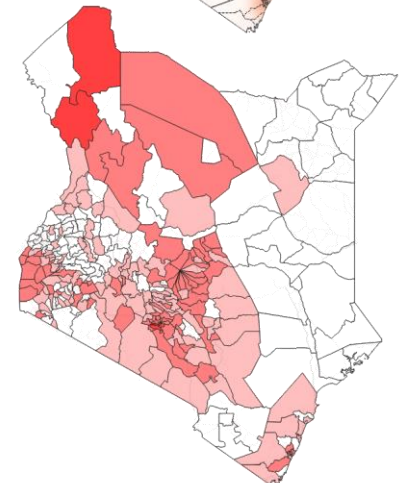
- ✓ Solution I: To avoid arbitrary political boundaries or aggregate up
- ✓ Solution II: Spatial smoothing: borrowing strength from neighboring areas



Kernel Density estimation



Smoothed average using Spatial weights



# Tobler's first law of geography

*Everything is related to everything else, but near things are more related than distant things.*

Tobler W., (1970): A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(2): 234-240

ThankYou