Università Ca' Foscari Venezia

Part of this material has been reproduced (with permission) from the lecture notes of the course “Introduction to Geographic Information Systems” held by Prof. Claudio Silvestri

Computer Science Applications to Cultural Heritage

Geographic Information Systems

Filippo Bergamasco (filippo.bergamasco@unive.it)
http://www.dais.unive.it/~bergamasco
DAIS, Ca' Foscari University of Venice
Academic year 2018/2019

Università Ca' Foscari Venezia

GIS & Cultural Heritage

In the framework of cultural heritage, **Geographical Information Systems (GIS)**, can be considered today as a logical and **standard tool** that goes beyond a simple repository of structured data coming from excavations and relict analysis:

- Large number of information can be **correlated** to a place and inter-correlated between themselves in terms of geographic position
- Graphical User Interfaces allows interactive and powerful representation of geographical data through multi-layered maps
- Common interface to 3rd party data like road network, buildings, points of interests, etc

Università
Ca' Foscari
Venezia

QGIS

Suggested software:

Suggested reading:

"A Gentle Introduction to GIS"

https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/

Università
Ca' Foscari
Venezia

What is GIS?

A Geographical Information System (GIS) is **a collection of tools** for:

- Collecting
- Storing
- Querying
- Transforming
- Visualizing

Spatial data [Burrough, McDonnell 1998]

The goal of these tools is to help people transforming geographic **data** into geographic **information.**

What is GIS?

GIS is usually more than a simple geographical database, it consists of:

- **Digital data:** geographical data to view and analyze
- **Computer hardware:** used for storing data, displaying graphics and processing data
- **Computer software:** computer programs that run on the computer hardware and allow you to work with digital data. A software program that forms part of the GIS is called a GIS Application

Data (geographic or not), can be stored in files directly by GIS Application or managed by a Database Management System (DBMS).

GIS and Databases

We have seen that a DBMS is an efficient way to store, organize, update and query data.

Usually GIS uses a **Spatial DBMS**: a special kind of database that can also natively manage geographical data (positions, shapes, etc)

A Spatial DBMS must:

- Manage the digital representation of geographical data
- Extend the query language to deal with geographical data

Università Ca' Foscari Venezia

What can I do with GIS?

Imagine you are a health worker and you make a note of the date and place of residence of every patient you treat.

Longitude	Latitude	Disease	Date
26.870436	-31.909519	Mumps	13/12/2008
26.868682	-31.909259	Mumps	24/12/2008
26.867707	-31.910494	Mumps	22/01/2009
26.854908	-31.920759	Measles	11/01/2009
26.855817	-31.921929	Measles	26/01/2009
26.852764	-31.921929	Measles	10/02/2009
26.854778	-31.925112	Measles	22/02/2009
26.869072	-31.911988	Mumps	02/02/2009
26.863354	-31.916406	Chicken Pox	26/02/2009

Example from "A Gentle Introduction to GIS"

Università Ca' Foscari Venezia

What can I do with GIS?

Some considerations about the data:

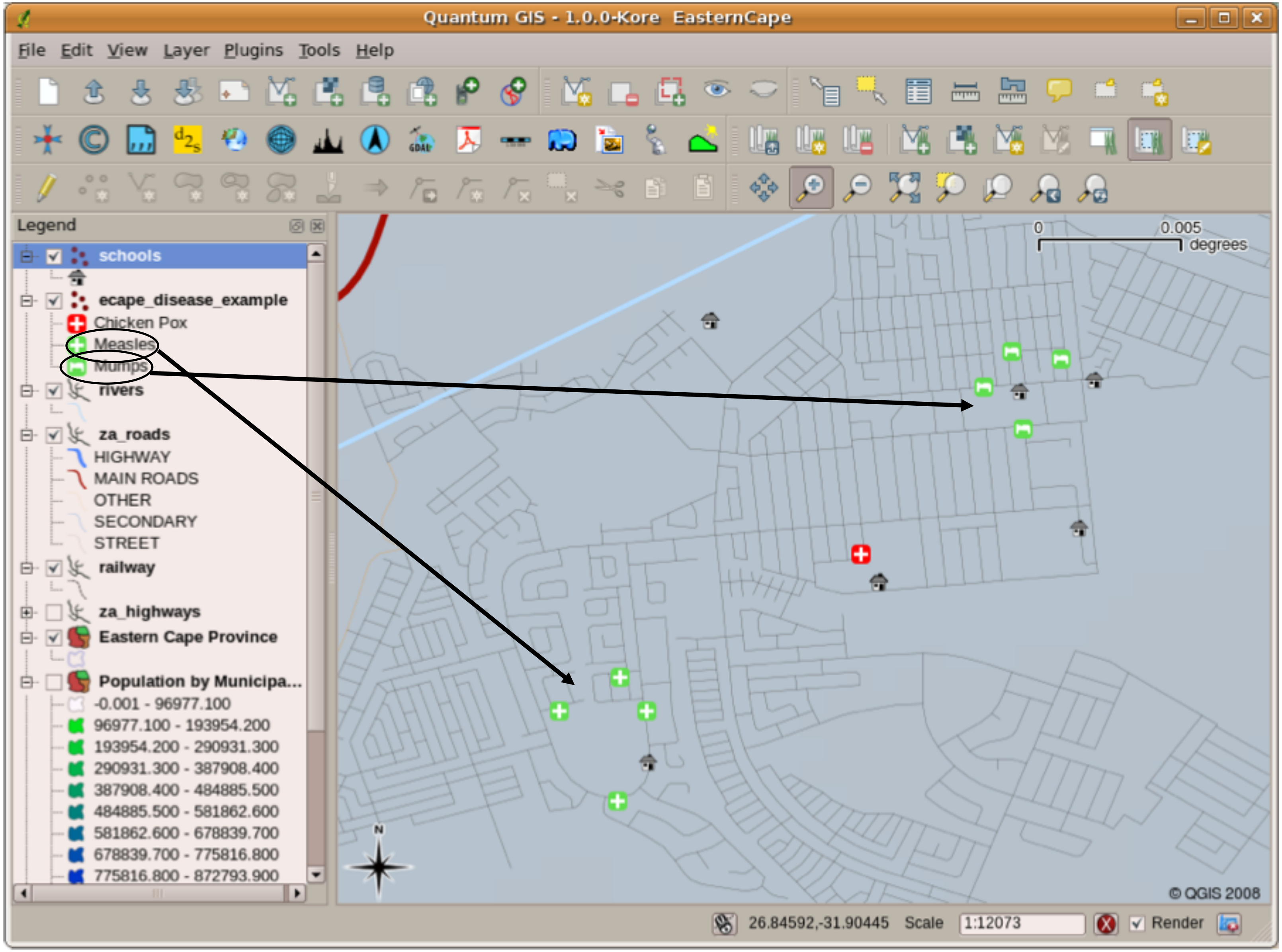
- Data is structured in a **table** like we usually do in a relational database
- Some attributes (like "latitude" and "longitude") represent geographical concepts (positions on the globe). Some others (like "Disease" and "Date") are additional (non geographical) data associated to the geographical concept

Without GIS, we can already observe that there were a lot of measles cases in January and February

Using GIS, we can discover interesting geographical patterns...

Example from "A Gentle Introduction to GIS"

What can I do with GIS?



Example from “A Gentle Introduction to GIS”

Università
Ca’ Foscari
Venezia

Università Ca' Foscari Venezia

Properties of geographic data

The table we have just seen shows a typical example of geographic data:

- Represents spatial locations and non-spatial attributes measured at given time (date)
- Geographic space is continuous (latitude and longitude are real numbers)
- Geographic space is nearly spherical: we need a mapping function to render points on a plane
- Geographic data tend to be spatially dependent

"Everything is related to everything else, but nearby things are more related than distant things."

[Tobler's first law of geography]

Typical geographic questions

Questions about space

- Where is the entity located?
- What is its extent?

Questions about attributes

- What are the attributes of the entities located there?
- Do its attributes match one or more criteria?

Questions about time

- When were the entity's location, extent or attributes measured?
- Has the entity's location, extent, or attributes changed over time?

Università Ca' Foscari Venezia

The power of GIS

A common feature of GIS is that they allow you to associate information (non-geographical data) with places (geographical data)

In a GIS Application we have a way to easily **change the appearance of the maps** we created **based on the non-geographical data** associated with places.

A common function of GIS Applications is to display **map layers**.

- Each map layer will represent something in the real world (roads, rivers, etc)
- In a map view, the layers are overlaid on top of each other.

Università Ca' Foscari Venezia

Representing geographical data

Geographical data (ie real-world features) can be represented in a GIS in two ways:

<u>Vector data representation:</u>

Object is described by:

- a geometric element
- a graphical representation
- one or more attributes.

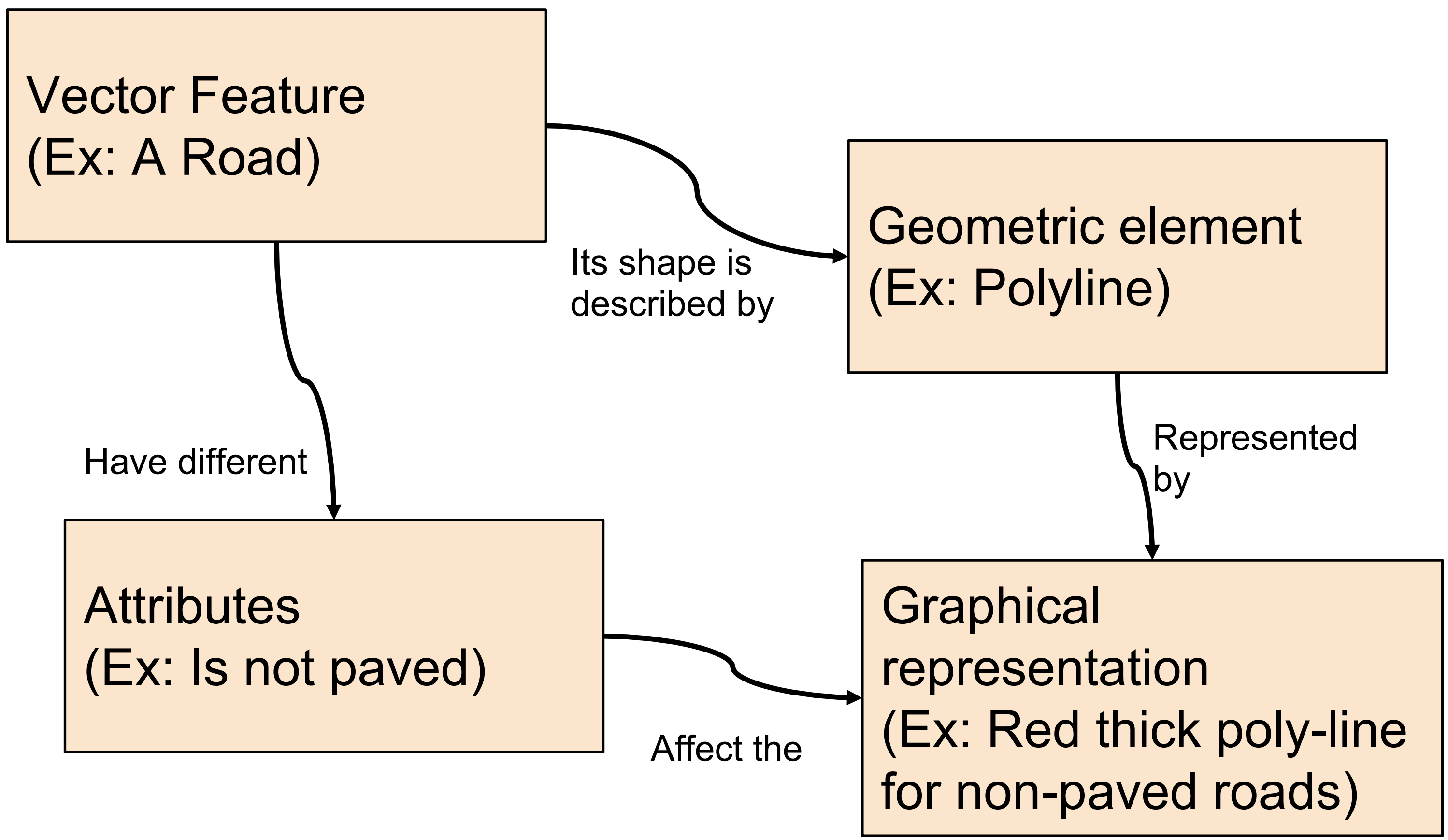
<u>Raster data representation</u>

Object is described as a continuous discrete function over an area (ex: the elevation map of a mountain)

Università
Ca' Foscari
Venezia

Vector based representation

Vector data provide a powerful way to represent real world features within the GIS environment.



Geometry

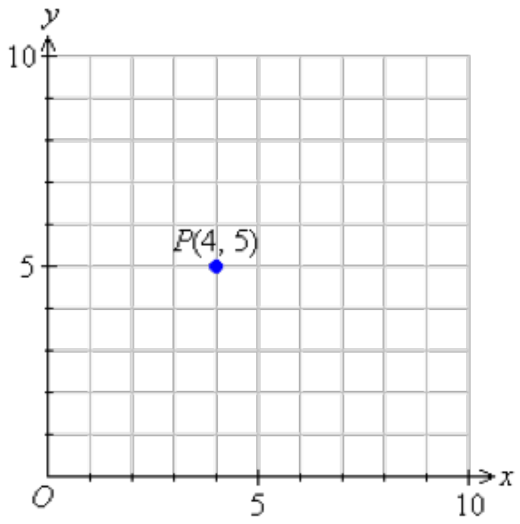
The geometry is made up of one or more interconnected vertices. A vertex describes a position in space using an X, Y and (optionally) z axis.

Basic types of geometries in GIS:

1. **Points**: composed by a single vertex
2. **Polylines**: composed two or more vertices, and the first and last vertex are not equal
3. **Polygons**: composed by three or more vertices, and the last vertex is equal to the first

Points

The point models objects with **no physical extension** (e.g., the peak of a mountain) or whose position is relevant but exact shape is not (e.g., a train station)



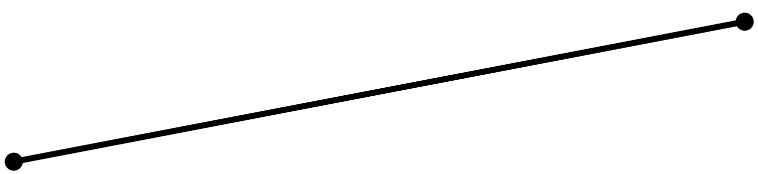
The X and Y values will depend on the Coordinate Reference System (CRS) being used. One of the most common reference systems is **Longitude** and **Latitude**.

Università Ca' Foscari Venezia

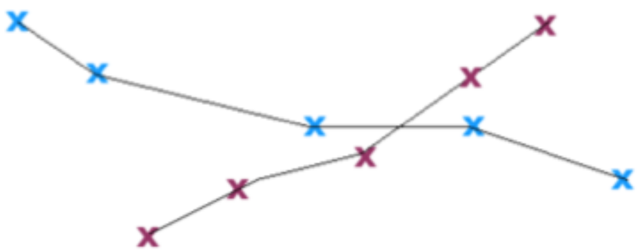
Lines

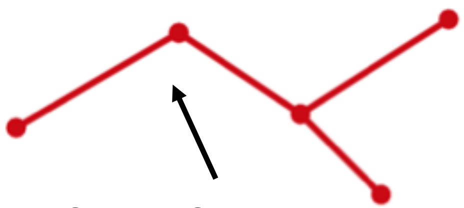
A line models spatial objects whose only relevant dimension is their **linear extension**.

A **segment** is defined by a pair of points, called endpoints



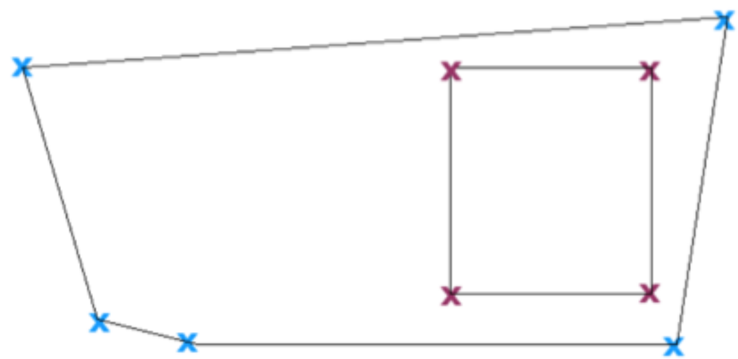
A **polyline** consists of a sequence of consecutive segments

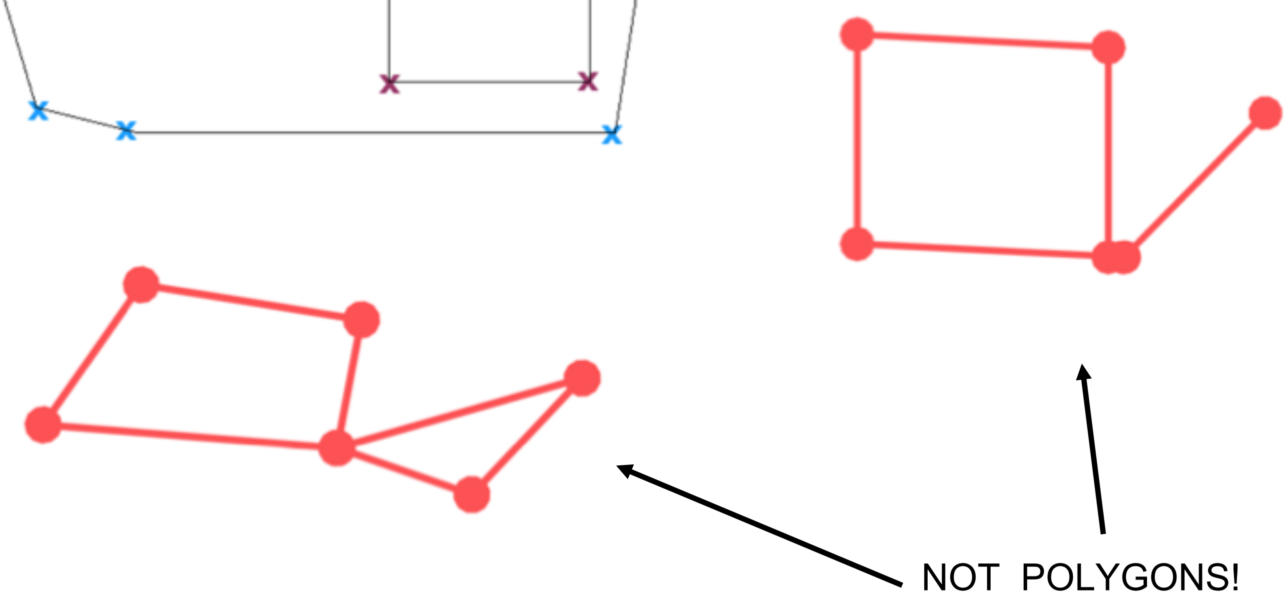




Polygons

A polygon, like a polyline, is a sequence of vertices. However in a polygon, the first and last vertices are always at the same position.

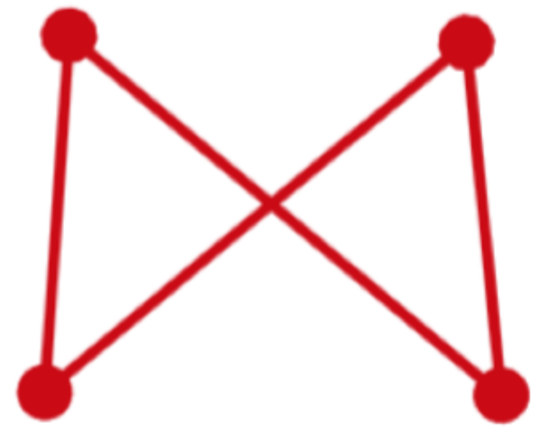




Polygons

A polygon is simple when its segment does not intersect

Simple



Not Simple

A polygon may contain holes

E.g.: the Italian peninsula without San Marino territory

Polylines and Polygons

A polyline is the typical representation adopted for transport networks: road, railways, cables

A polygon is used to describe an object having both linear and areal characterization

Example

- Physical objects: lakes
- Administrative entities: municipality
- Abstract elements: low population density zones

Attributes

Vector features may have associated **attributes**. Some attribute can be computed from the geometry, such as the area of a region

Examples:

- Road polyline may have attributes that describe whether it is surfaced with gravel or tar, how many lanes it has, whether it is a one way street, and so on
- A dam polygon may have attributes for depth and water quality.

Università Ca' Foscari Venezia

Graphical representation

The graphical representation describe **the appearance of an entity**

It is important to distinguish graphical representation and geometry:

- The geometry describe the object position and shape
- the graphical representation how it should be presented

Graphical representation

Sometime determining the geometry of an object is not trivial

Ex: Borders are not always well defined!

Which is the correct border of the wood?

Graphical representation

Sometimes determining the geometry of an object is not trivial

Ex: the more suitable geometry depends on the intended use of the data

Roads can be lines or polygons depending on our interest for the road surface

Università Ca' Foscari Venezia

Common problems with vector data

Inaccurate vector data may create artifacts in the visualization of the maps.

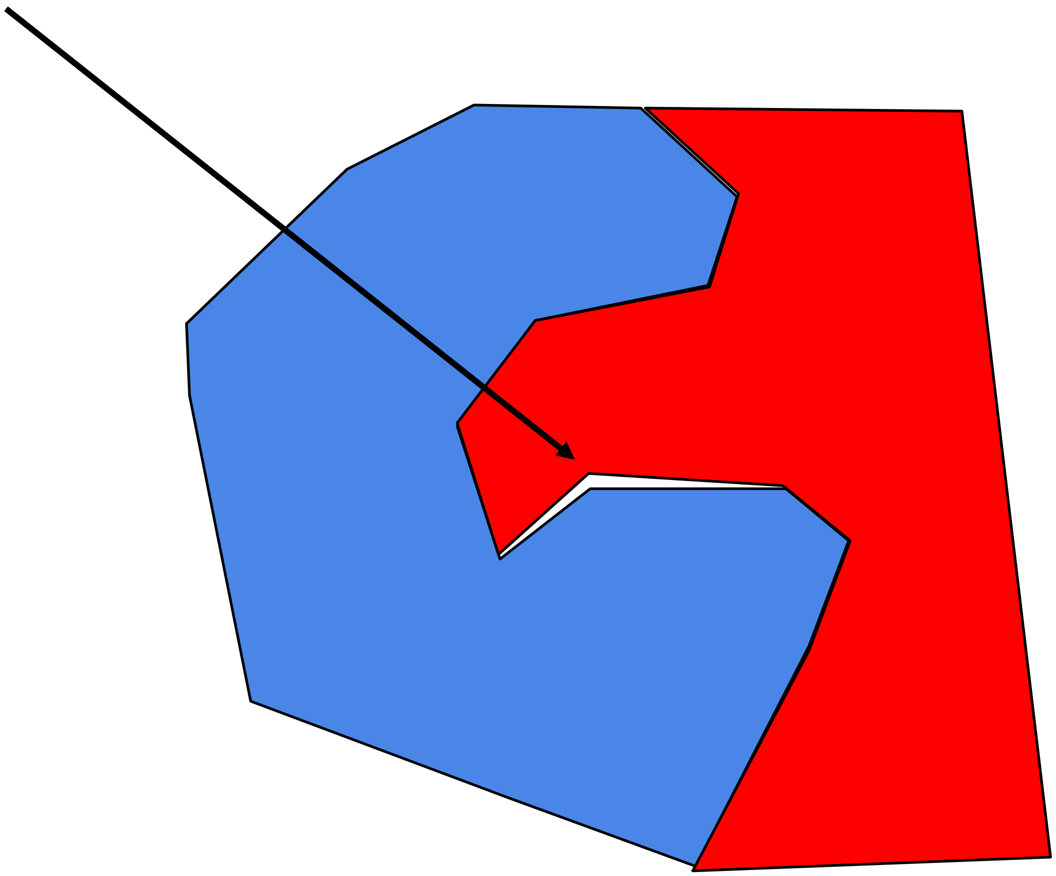
Typical causes:

- Instruments used to capture the data are not properly set up
- People capturing the data aren't being careful
- Time or money don't allow for enough detail in the collection process
- Etc

Because of these types of errors, it is very important to digitise data carefully and accurately.

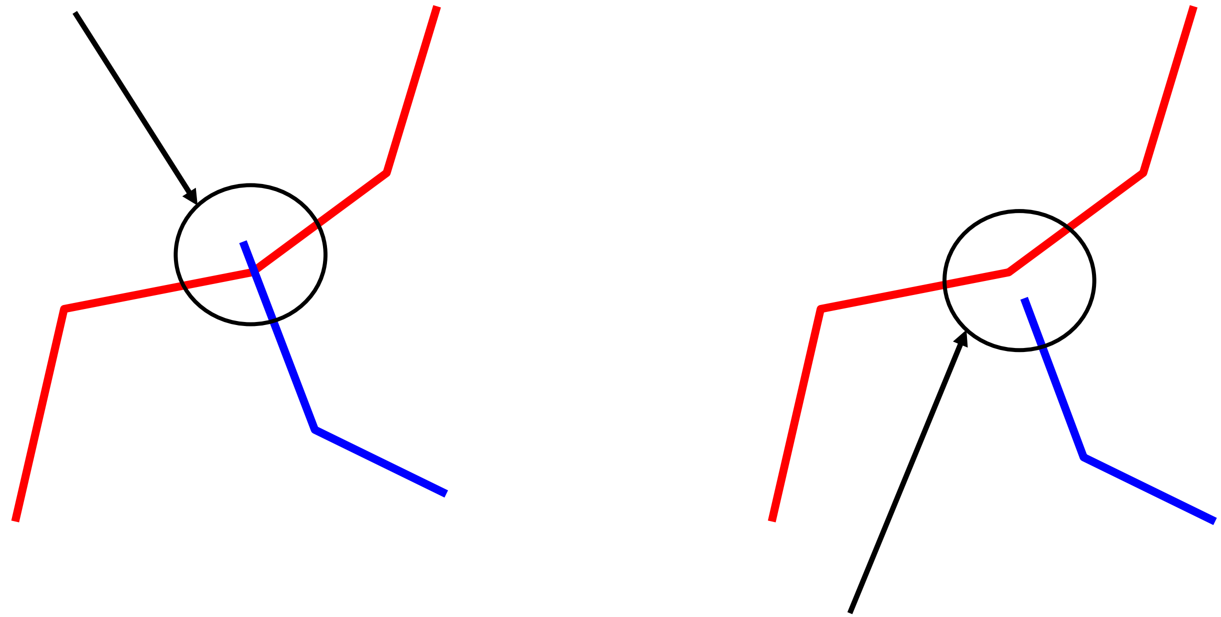
Common problems with vector data

Slivers can occur when the edges of two polygon areas don't meet properly



Common problems with vector data

Overshoots can occur when a line feature such as a road does not meet another road exactly at an intersection.



Undershoots can occur when a line feature (e.g. a river) does not exactly meet another feature to which it should be connected.

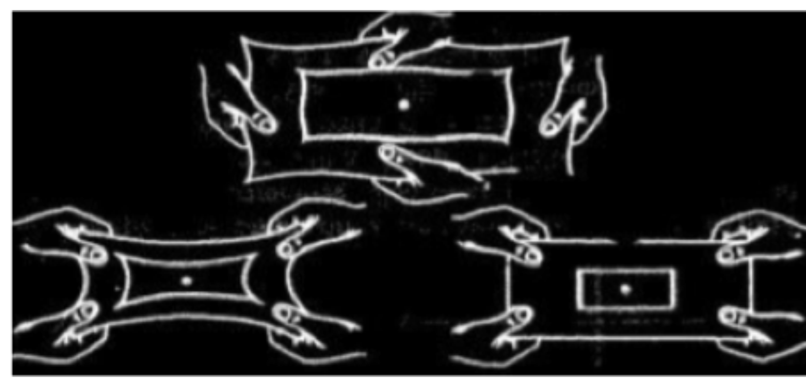
Topology

Università Ca' Foscari Venezia

Slivers, overshoots and undershoots are typical kinds of **topological errors** that we may have when working with geometrical data

"In mathematics, topology is concerned with the properties of space that are preserved under continuous deformations, such as stretching, crumpling and bending"

Source: Wikipedia



Example: the topological property of a point to be enclosed in a polygon is preserved under stretching

Topology

In a GIS, topology expresses the spatial relationships between connecting or adjacent vector features (points, polylines and polygons)

Topological or topology-based data are useful for detecting and correcting digitising errors (e.g. two lines in a roads vector layer that do not meet perfectly at an intersection).

Topology is necessary for carrying out some types of spatial analysis, such as network analysis.

Università
Ca' Foscari
Venezia

Topology

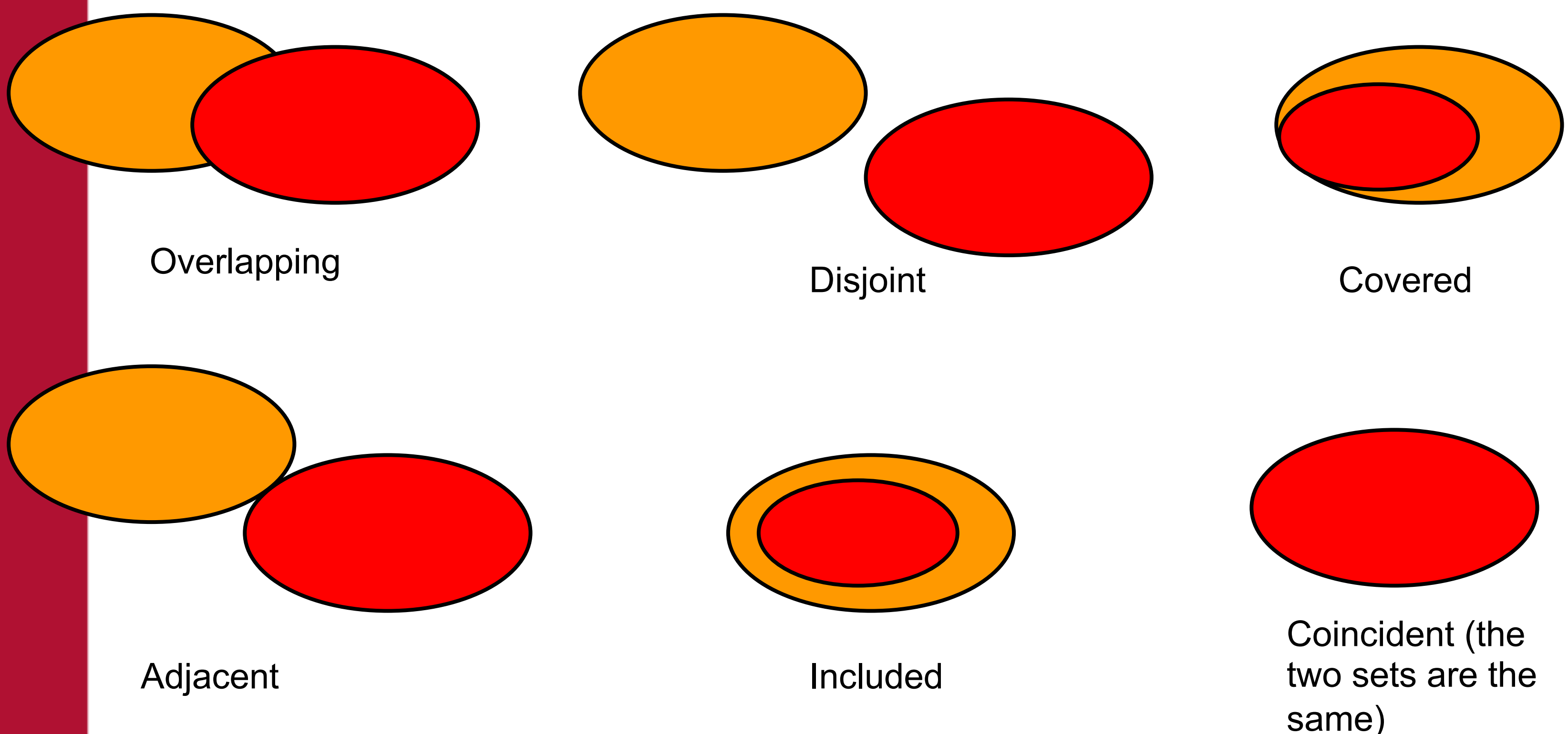
Example: The underground map of london

Circles are used to mark intersecting lines so that we can change train. The “intersection” is a topological information

Topology rules

Topological relations between two geometric elements are finite and enumerable.

Example: 2 polygons



Preventing topological errors

Many common errors that can occur when digitising vector features can be prevented by topology rules that are implemented in many GIS applications.

Topology is not enforced by default but is defined as relationship rules and let the user choose the rules, if any, to be implemented in a vector layer.

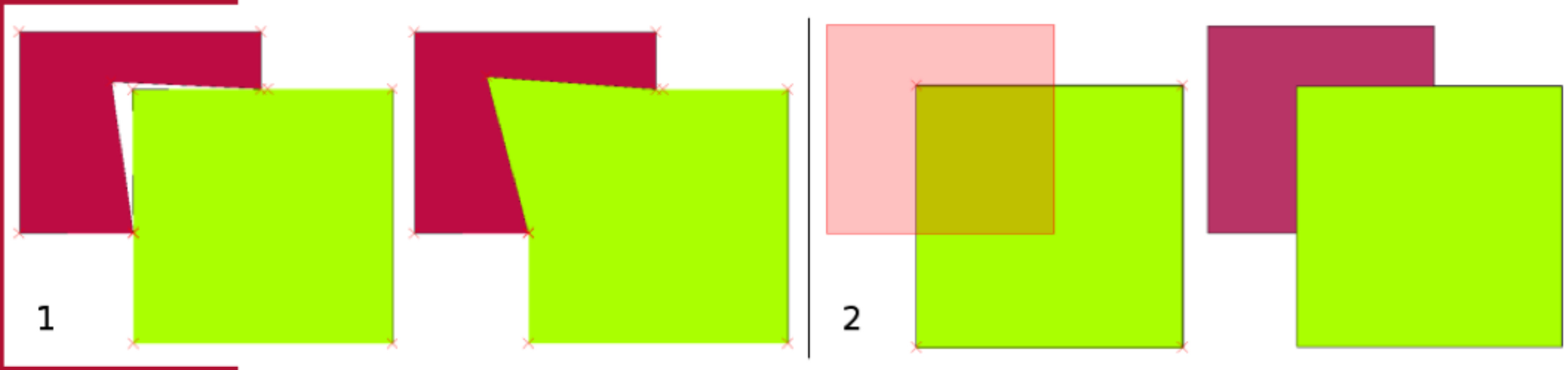
Examples:

- Area edges of a municipality map must not overlap
- Area edges of a municipality map must not have gaps (slivers).
- Polygons showing property boundaries must be closed.

Topological editing

Many GIS tools provide a functionality called "topological editing" assisting the user when defining (drawing) geometric primitives so that topological errors are prevented

Ex: topological editing in QGIS



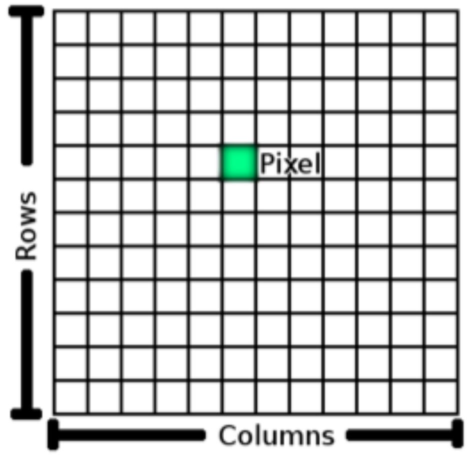
(1) Topological editing to detect shared boundaries, when moving vertices. When moving a vertex, all features that share that vertex are updated. (2) To avoid polygon overlaps, when a new polygon is digitised (shown in red) it is clipped to avoid overlapping neighbouring areas.

Raster data representation

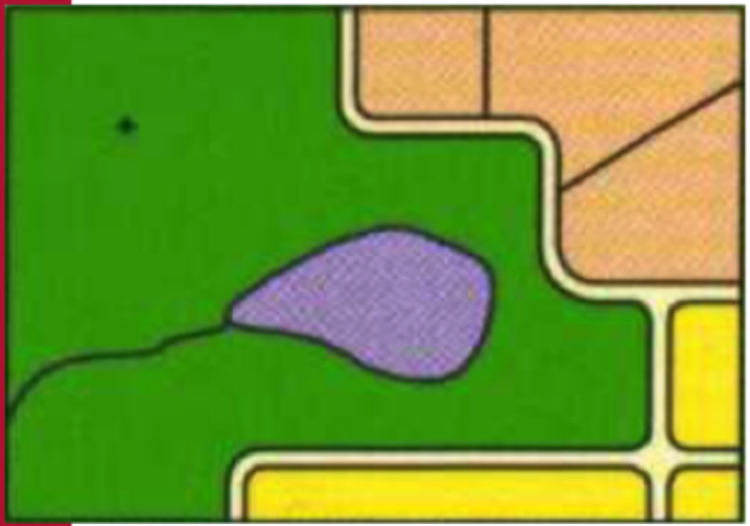
Vector features use geometry (points, polylines and polygons) to represent the real world features

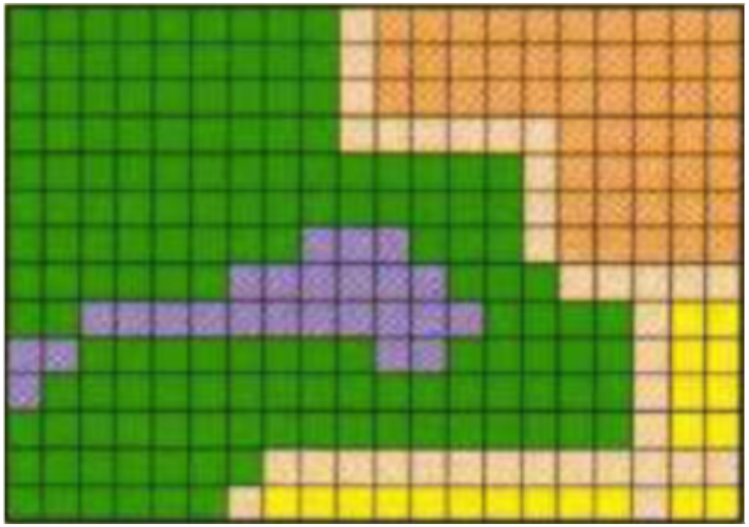
Raster data: completely different approach!

Features are represented as matrix of pixels (also called cells), each containing a value that represents the conditions for the area covered by that cell



Vector vs. Raster



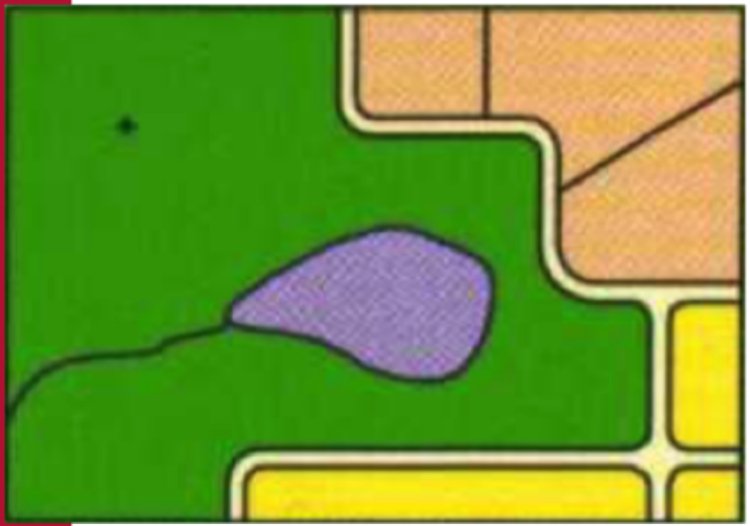


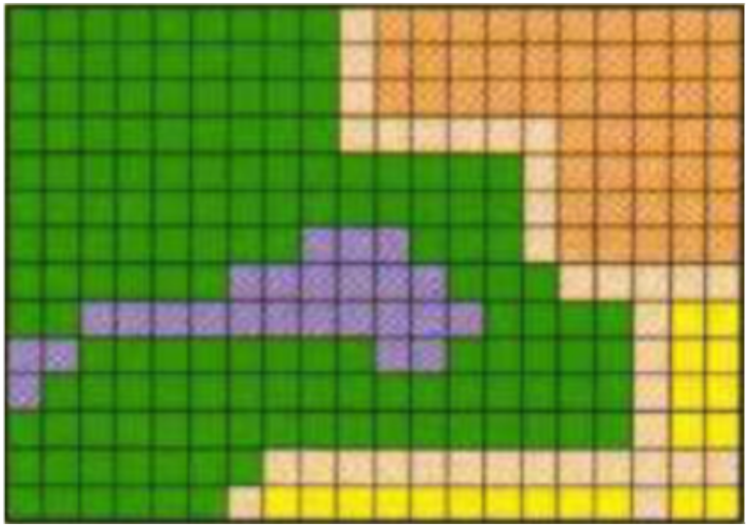
Space is defined by geometric primitives

Each pixel is associated with a position in space. Thus, its value can be associated to one or more objects

Università Ca' Foscari Venezia

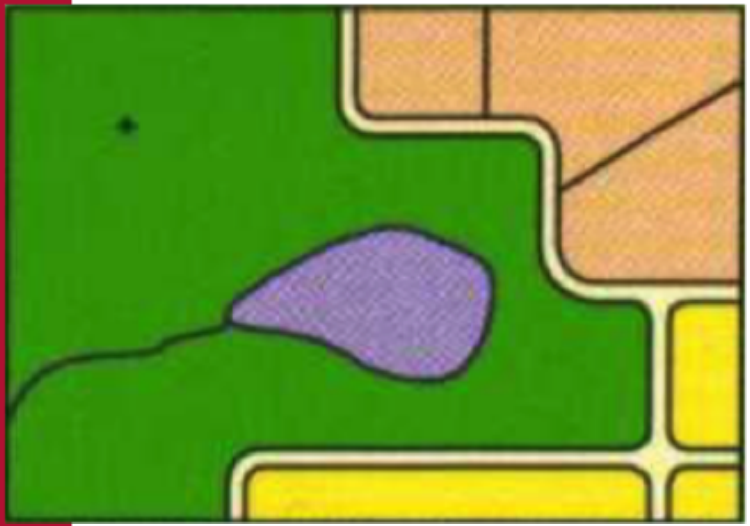
Vector vs. Raster

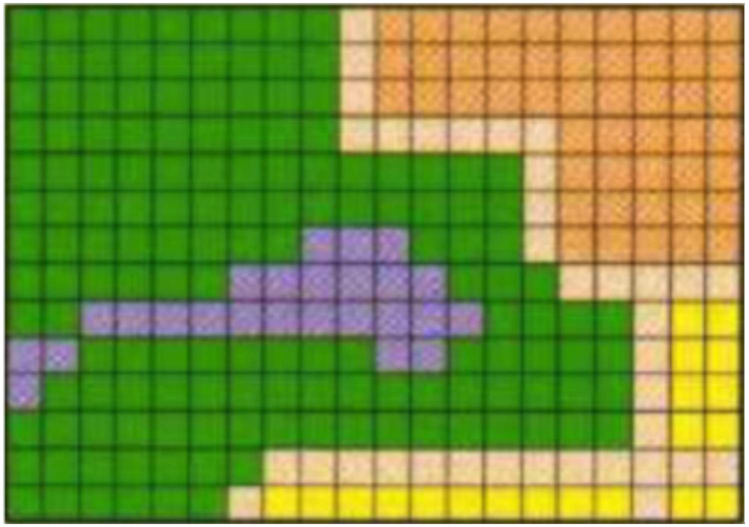




Raster data is used in a GIS application when we want to display information that is **continuous across an area** and cannot easily be divided into vector features.

Vector vs. Raster





Well defined borders

Borders are not well defined but represent phenomena that changes smoothly in space

Università Ca' Foscari Venezia

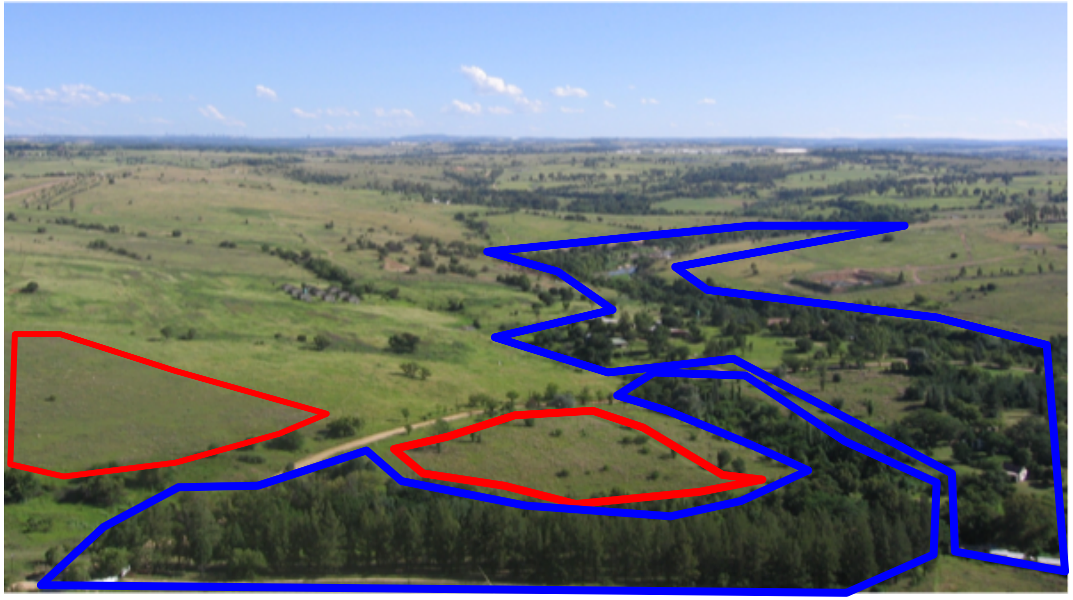
Vector vs. Raster

Example:

The grasslands shown below have many variations in colour and density of cover

Vector vs. Raster

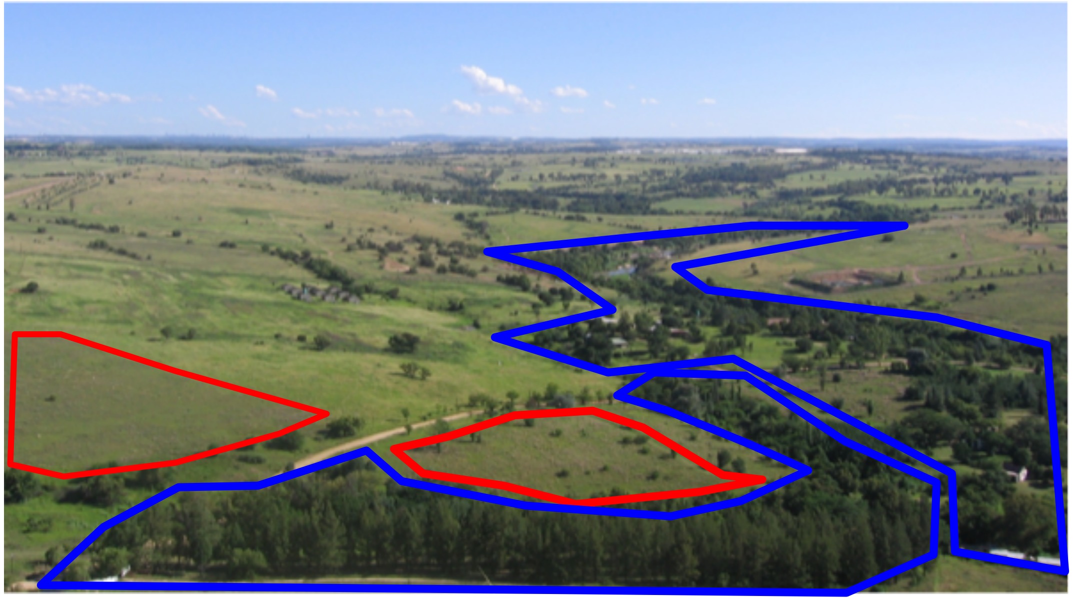
Vector approach: make a single polygon around each grassland area



PROBLEM: a lot of the information about the grassland would be lost in the process of simplifying the features to single polygons

Vector vs. Raster

PROBLEM: a lot of the information about the grassland would be lost in the process of simplifying the features to single polygons

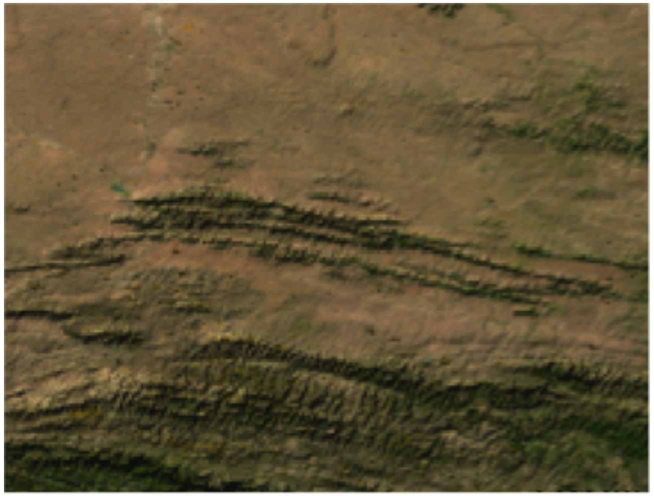


When you give a vector feature attribute values, **they apply to the whole feature**

Vectors aren't good at representing features that are not homogeneous (entirely the same) in space.

Vector vs. Raster

The raster approach solves this problem.

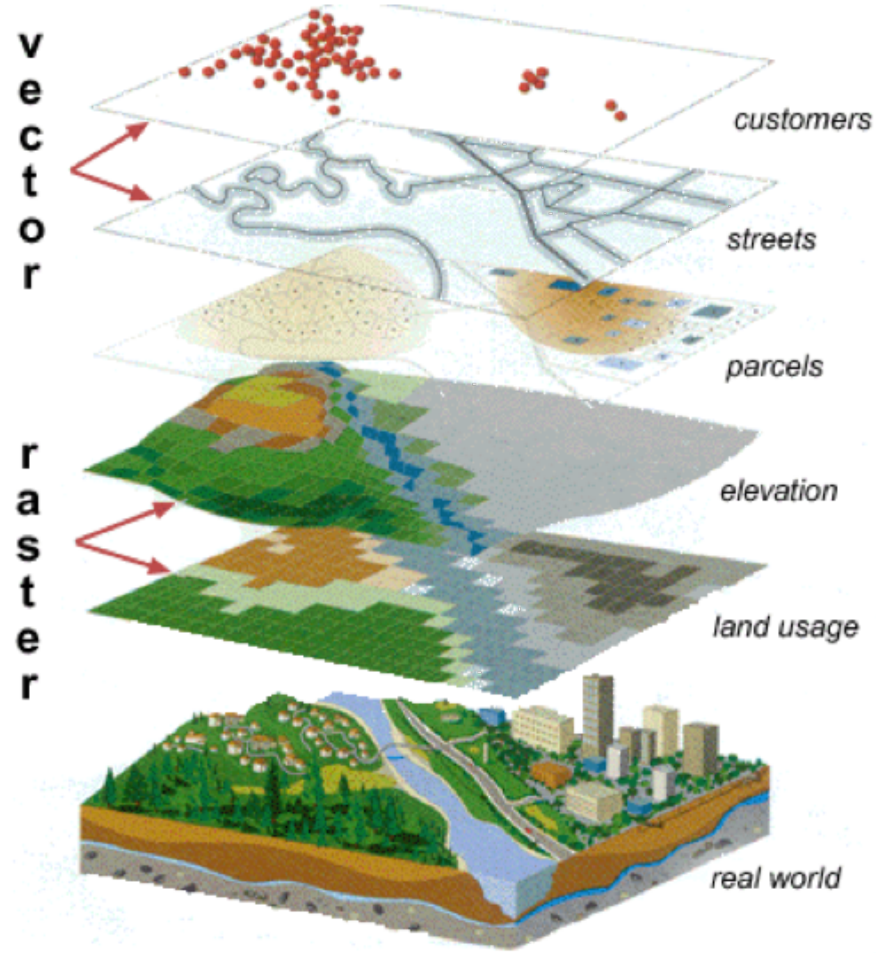


Usually, raster data is used as a backdrop behind vector **layers** in order to augment the vector information.

Layers

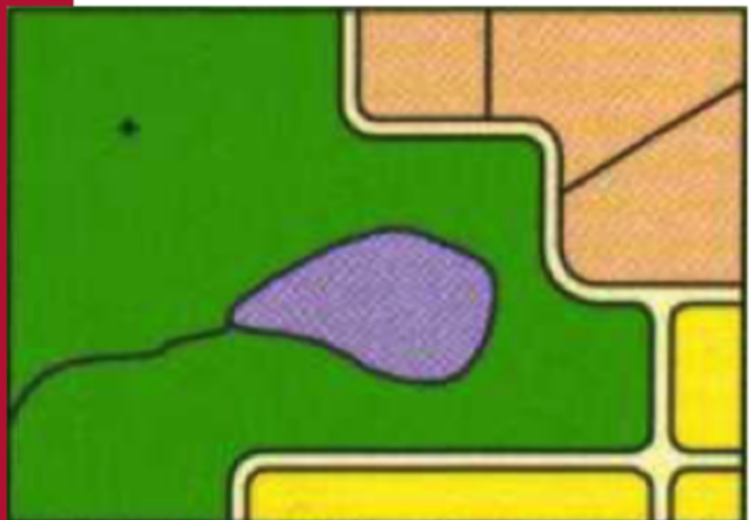
GIS data are organized in homogeneous groups (layers)
E.g.: road layer, river layer, border layer

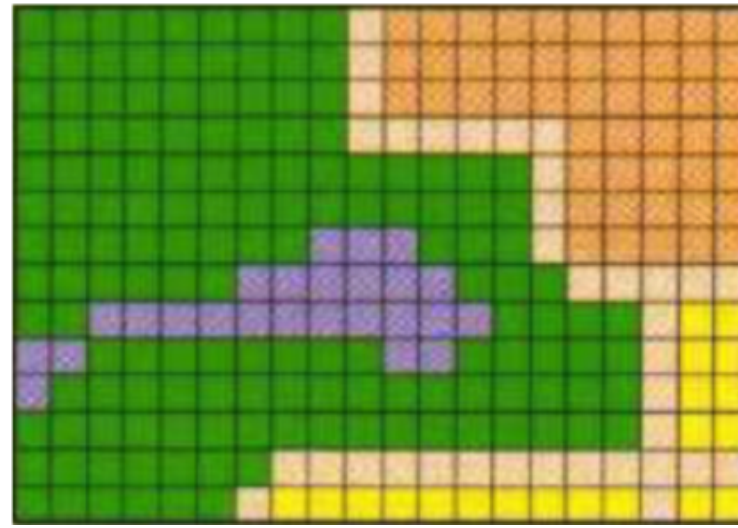
The geographic information of each layer allows for composition, correlation and joint analysis



Università
Ca' Foscari
Venezia

Vector or Raster layer?





Good choice:

Elements having an exact position and shape

Bad choice:

Continuous phenomena or elements having vague/uncertain borders.

Good choice:

Image and continuous phenomena representation

Bad choice:

Point and line elements.

Regular areas with well defined borders

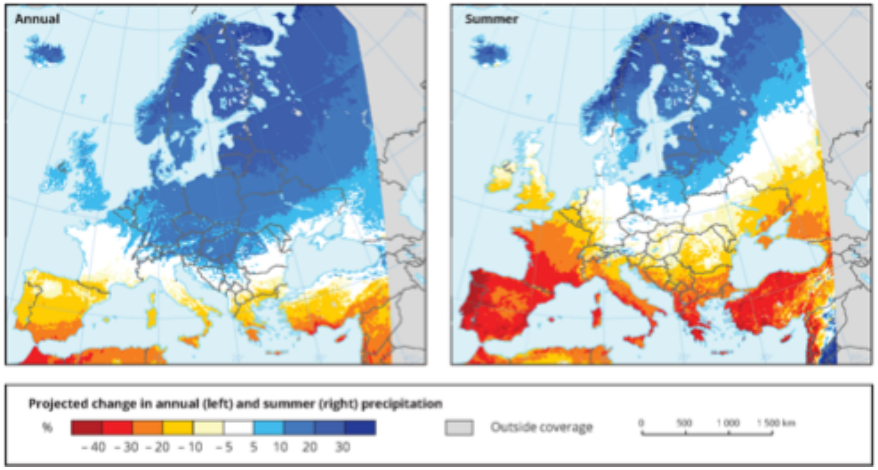
Università
Ca' Foscari
Venezia

Working with raster layers

Usually, raster layer is used to show:

- images that depict the real world surface (e.g. satellite images and aerial photographs)

- Images that represent continuous spatial distribution of measures (ex. rainfall trends over an area)



Working with raster layers

An image can be handled as a geographic data only when it is possible to **associate image pixels with geographic space regions** (one-to-one correspondence).

In other words, mapping between pixels and lat/lon coordinates

Georeferencing is the process of defining exactly where on the earth's surface an image or raster dataset was created

The georeferencing information for a raster is often provided in a small text file accompanying the raster

Università Ca' Foscari Venezia

Georeferencing

To georeference a raster usually we need to define:

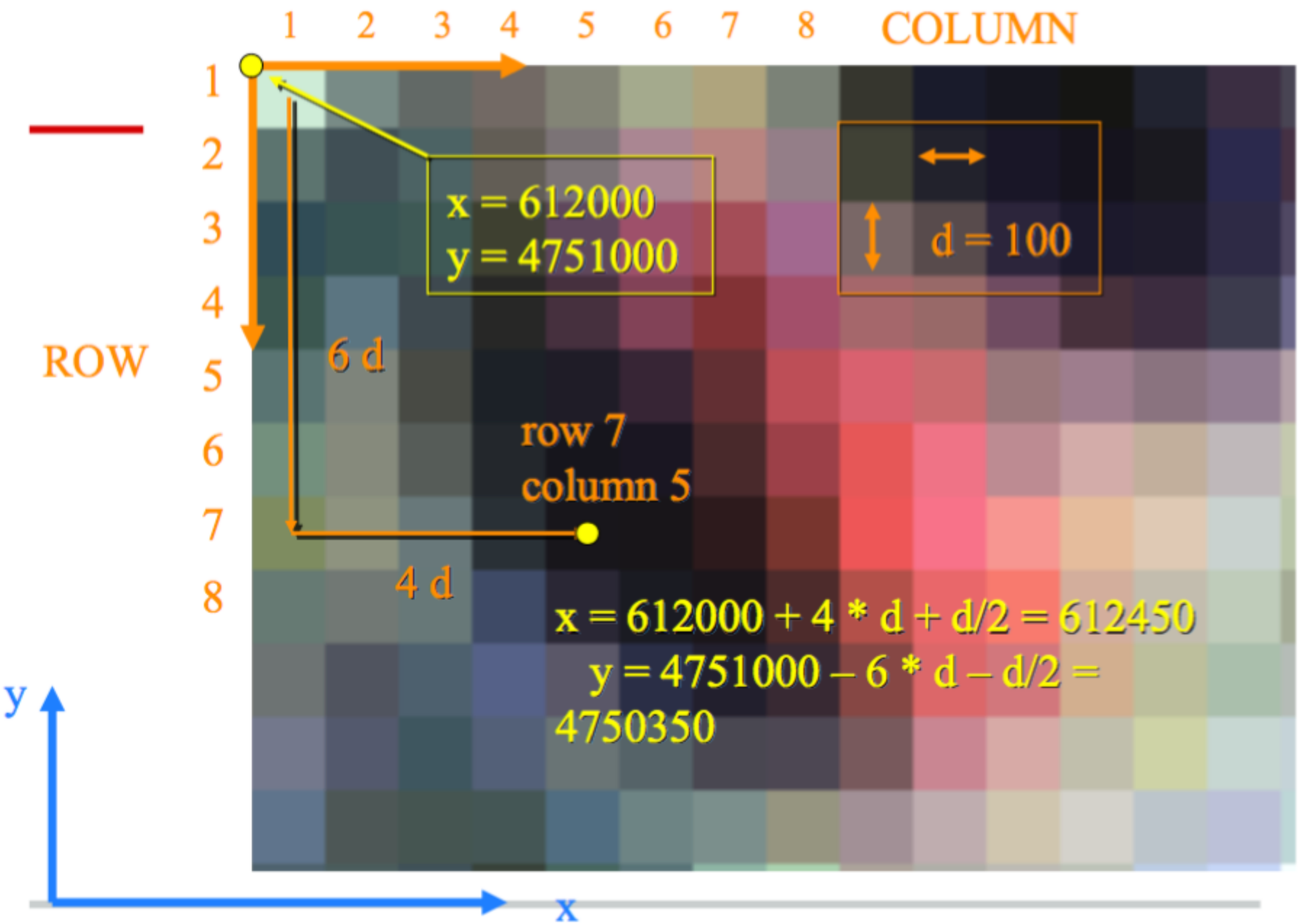
- coordinate for the top left pixel in the image
- the size of each pixel in the X direction
- the size of each pixel in the Y direction
- the amount (if any) by which the image is rotated

In the simplest case, we just use:

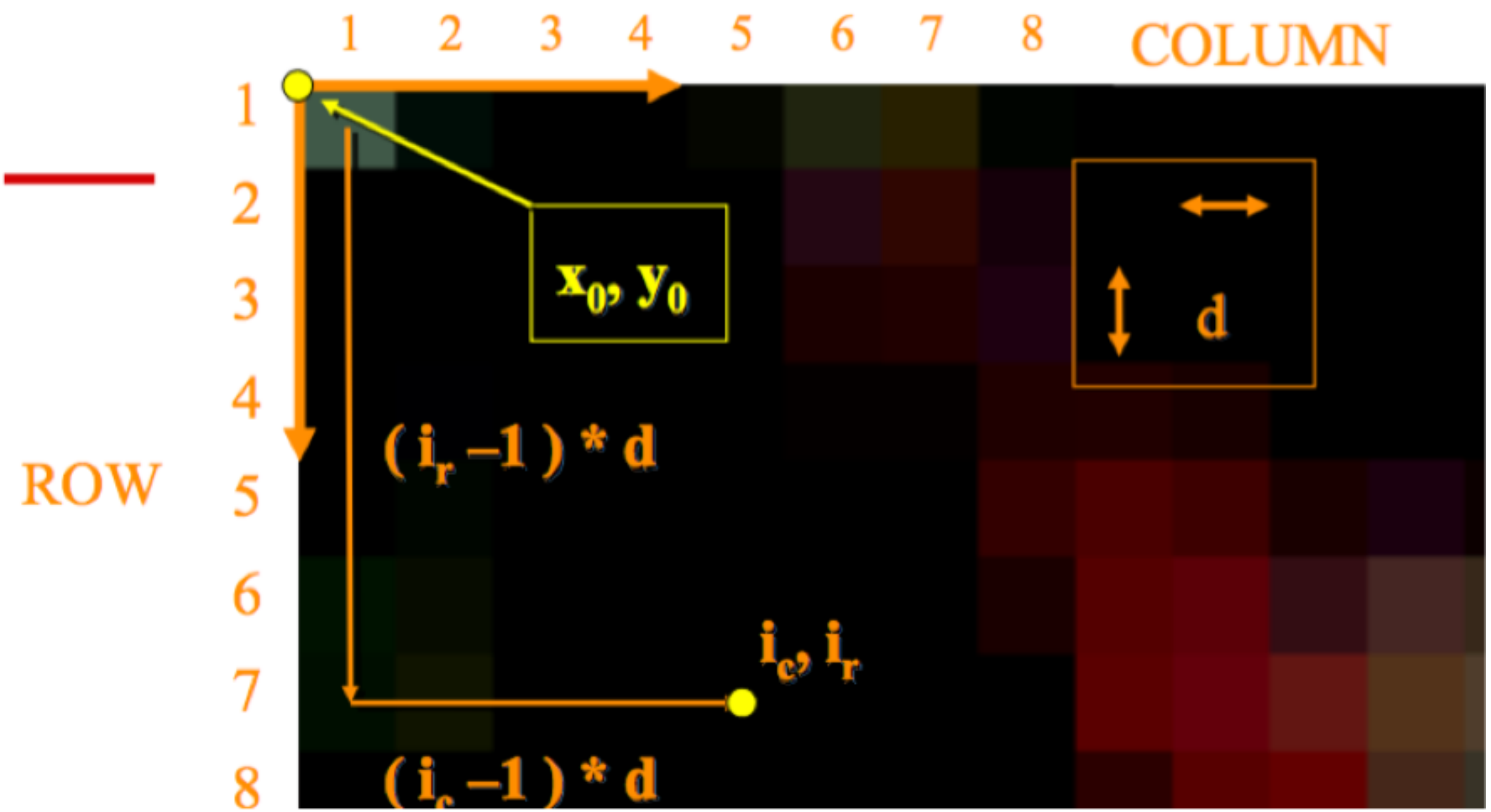
(x0,y0): coordinates of the top-left pixel

d: pixel size

Georeferencing



Georeferencing

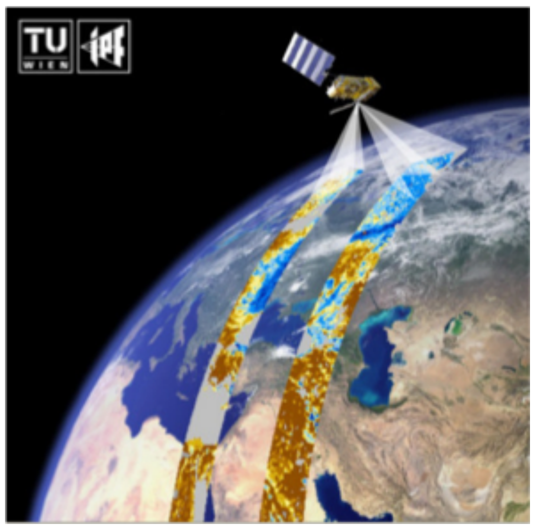


$$x = x_0 + (i_c - 1)d + \frac{d}{2}$$

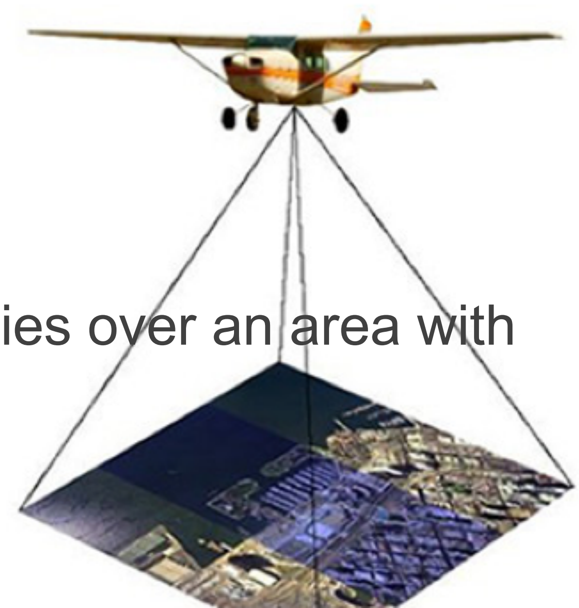
$$y = y_0 + (i_r - 1)d + \frac{d}{2}$$

Source of raster data: remote sensing

Satellite imagery is created when satellites orbiting the earth point special digital cameras towards the earth and then take an image of the area on earth they are passing over.

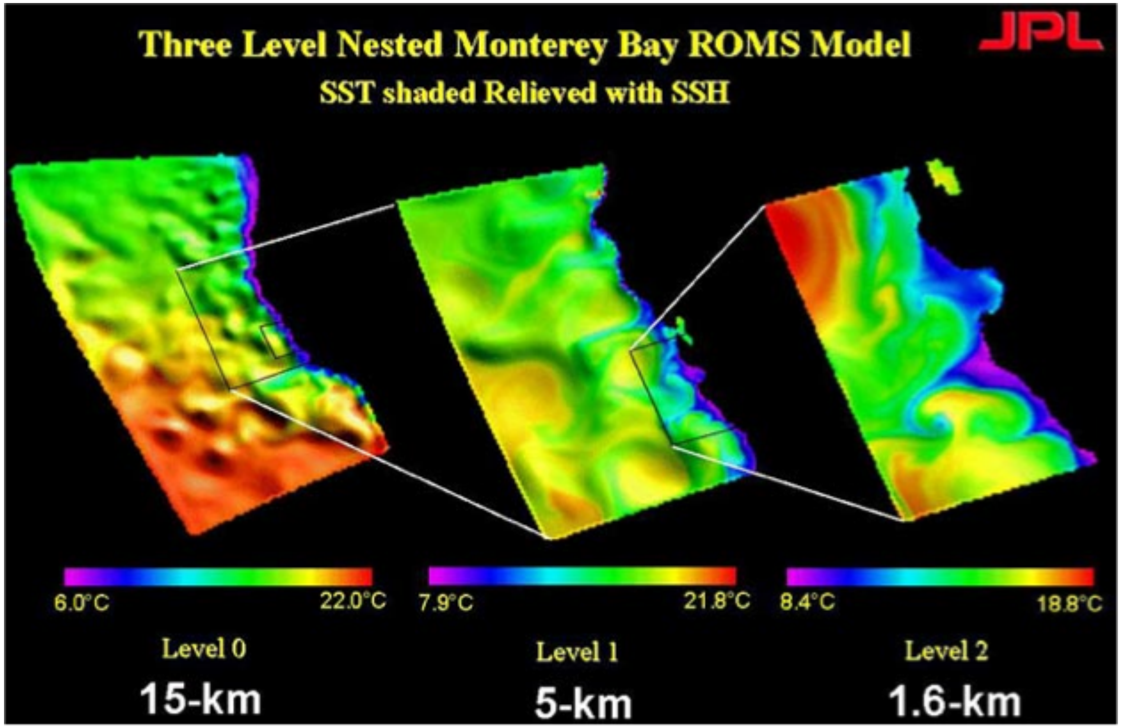


In **aerial photography**, an aeroplane flies over an area with a camera mounted underneath it



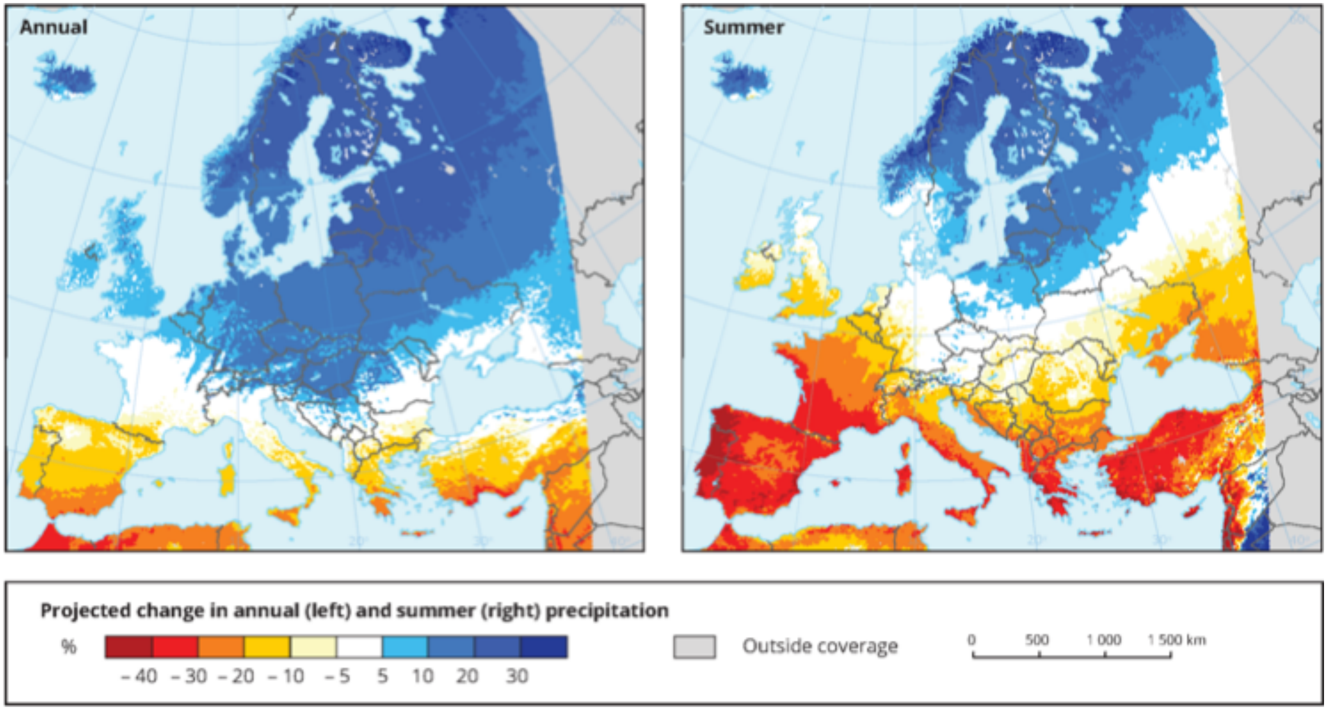
Source of raster data: computed raster data

From forecasting/physical models:



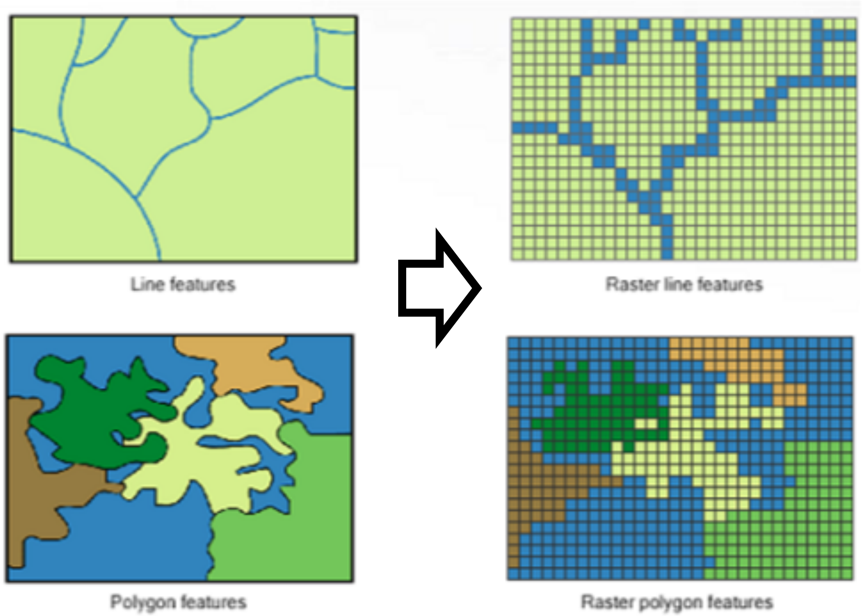
Source of raster data: computed raster data

From scattered data measurements that are interpolated over an area of interest



Source of raster data: computed raster data

From vector data because the data owners want to share the data in an easy to use format

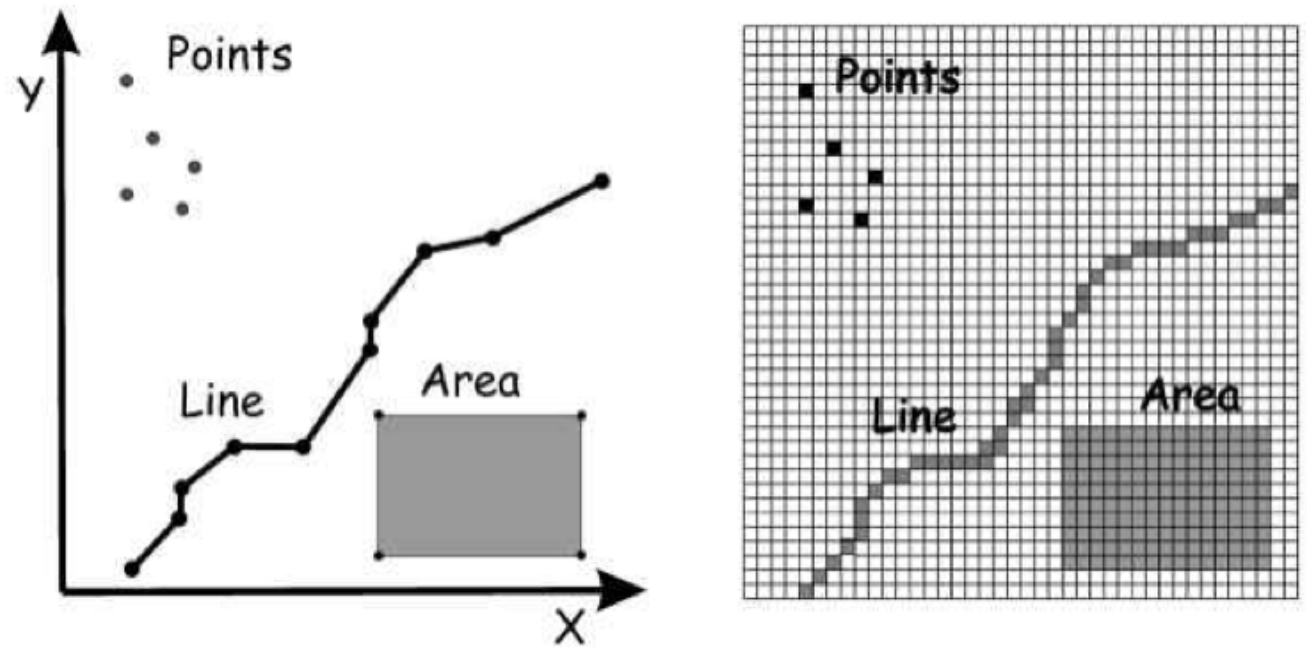


only useful if the attributes, that users need to be aware of, can be represented on the map with **labels** or **symbology.**

> Raster data cannot have attributes directly associated

Vector to raster conversion

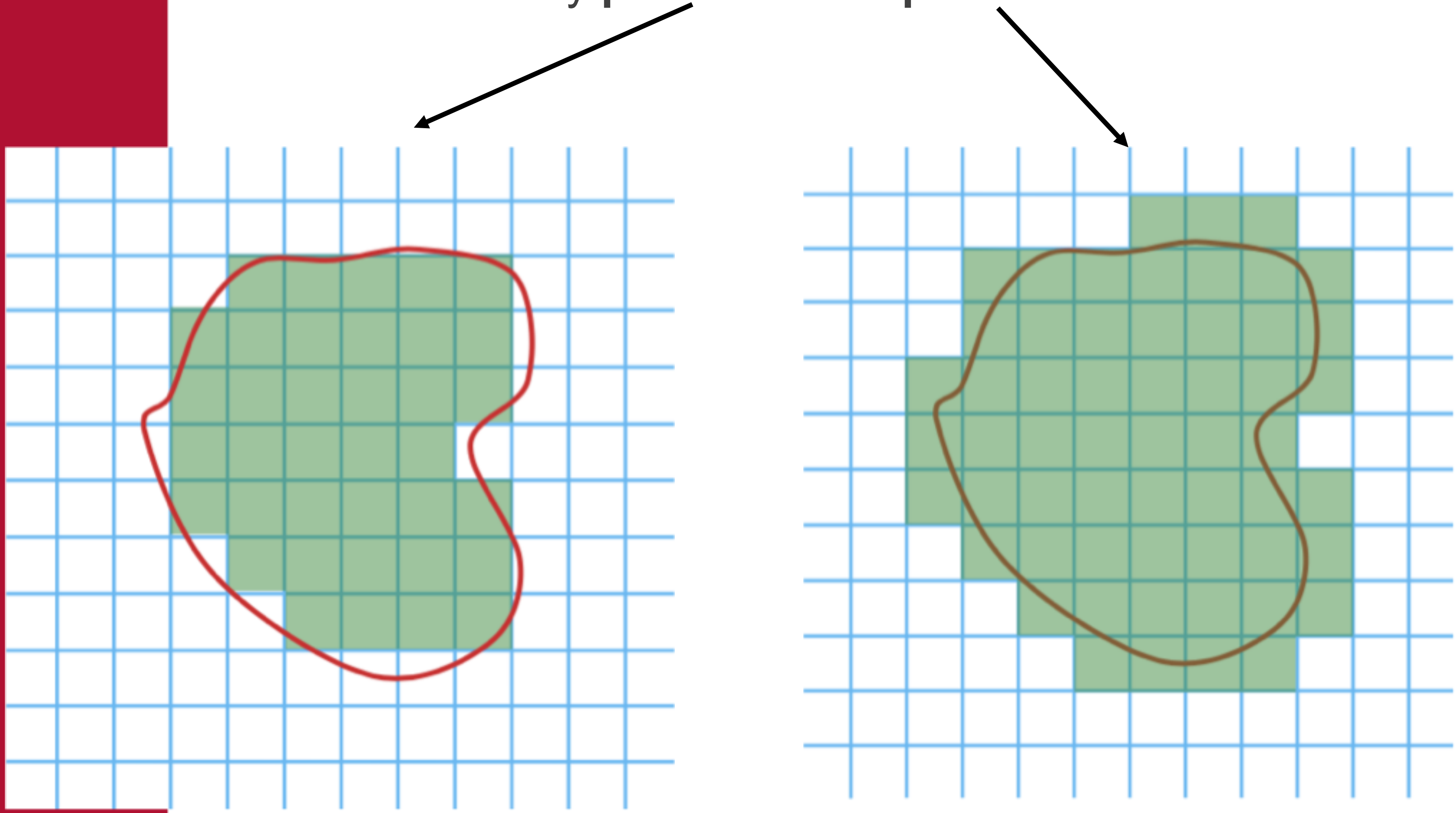
It is intrinsically imperfect due to the non alignment of geometry to the pixel grid



Università Ca' Foscari Venezia

Vector to raster conversion

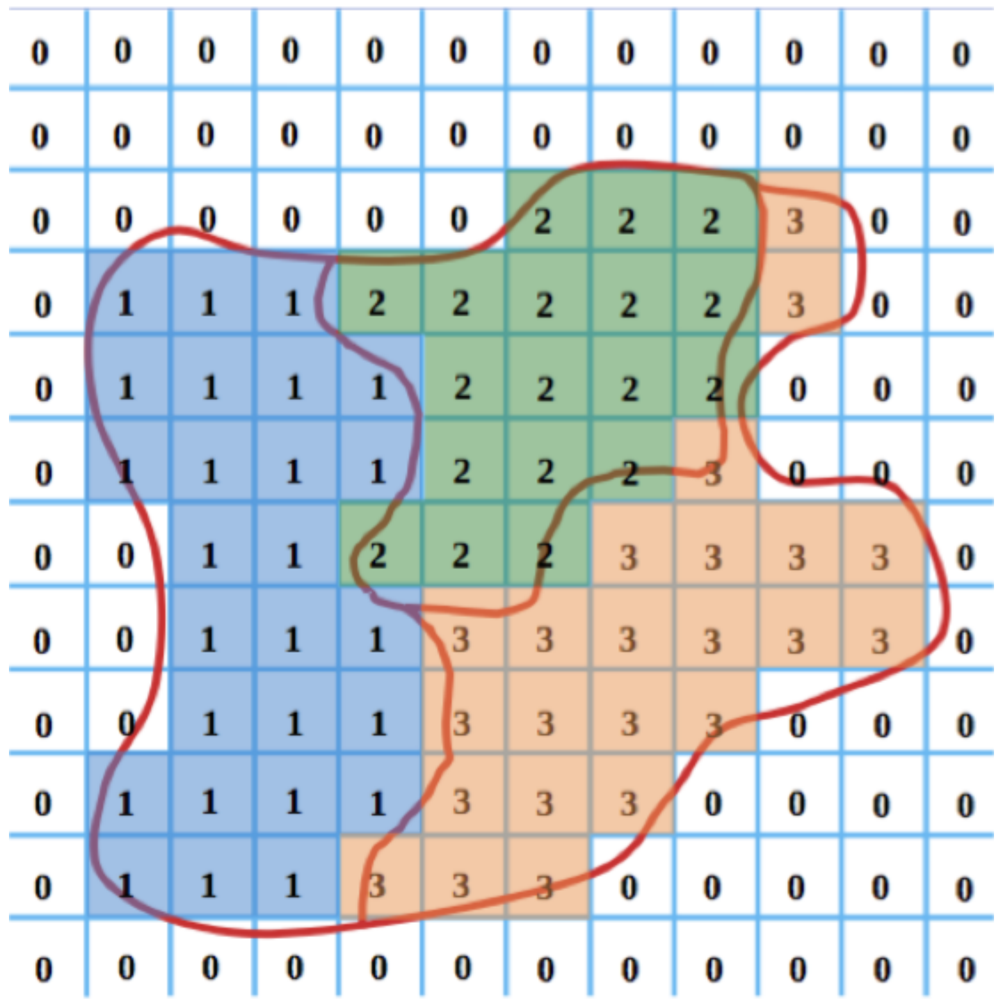
Areas converted by **prevalence** or **presence**



Università
Ca' Foscari
Venezia

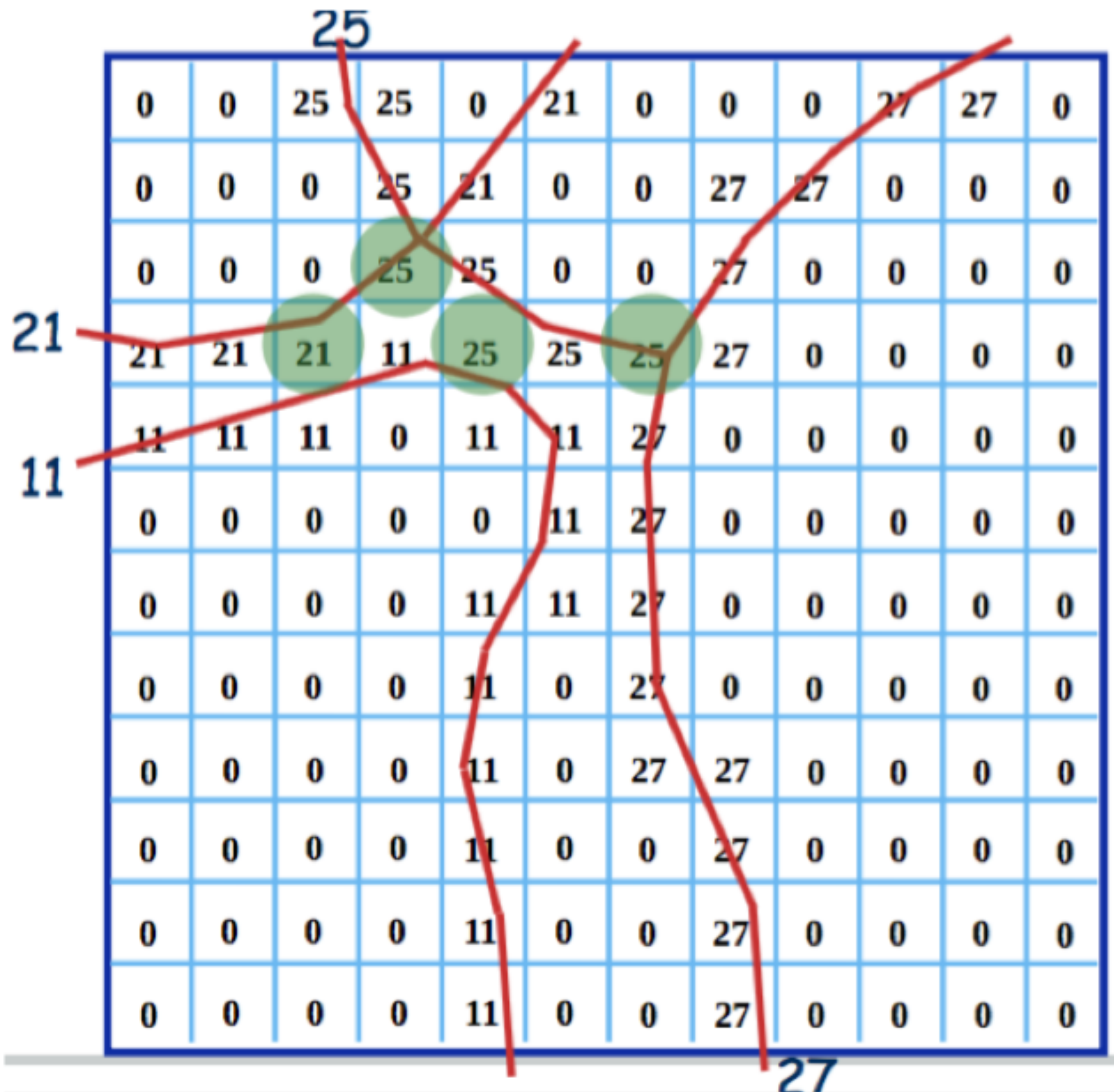
Vector to raster conversion

Adjacent areas by **prevalence**



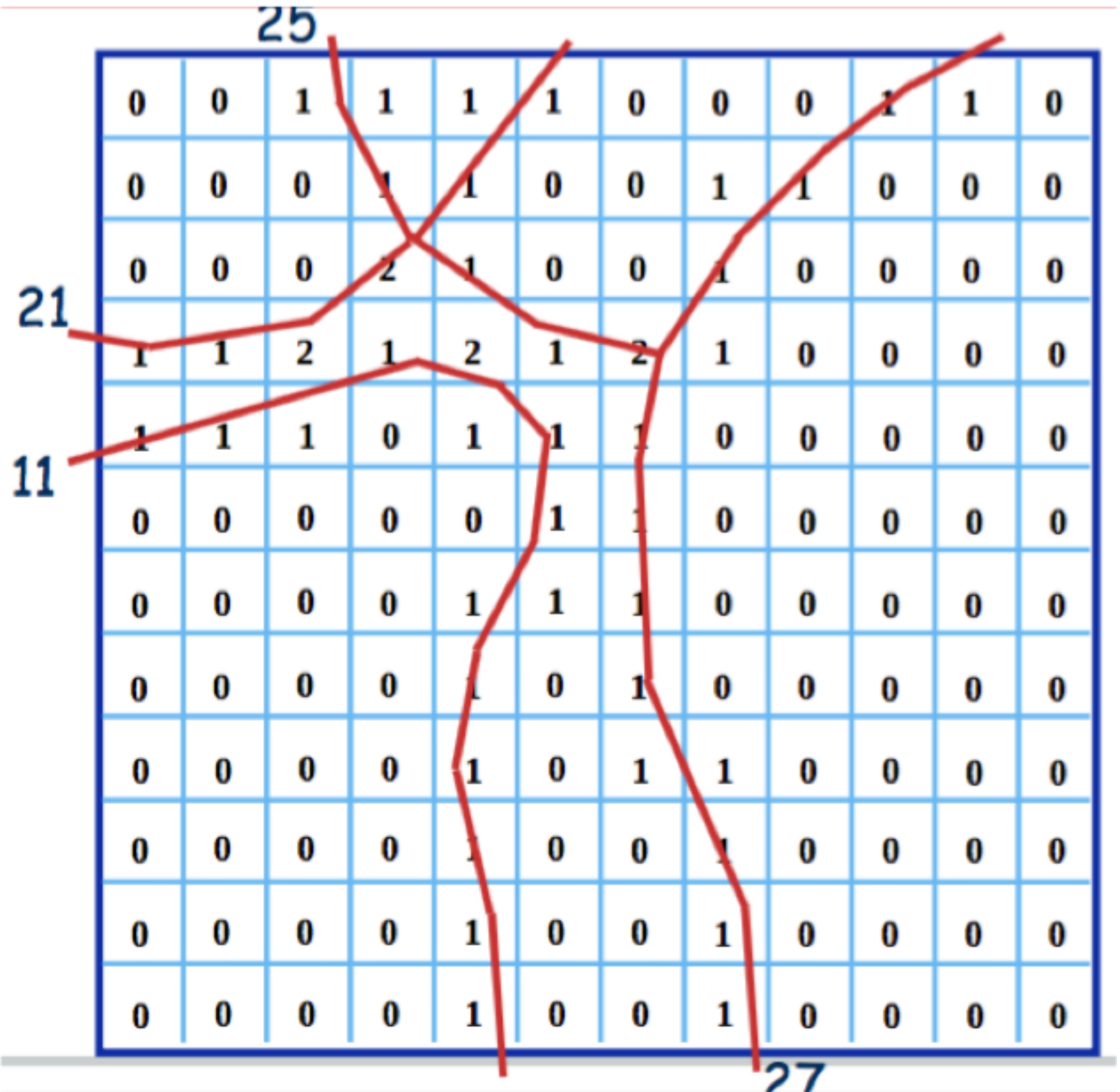
Vector to raster conversion

Attribute coding to raster lines



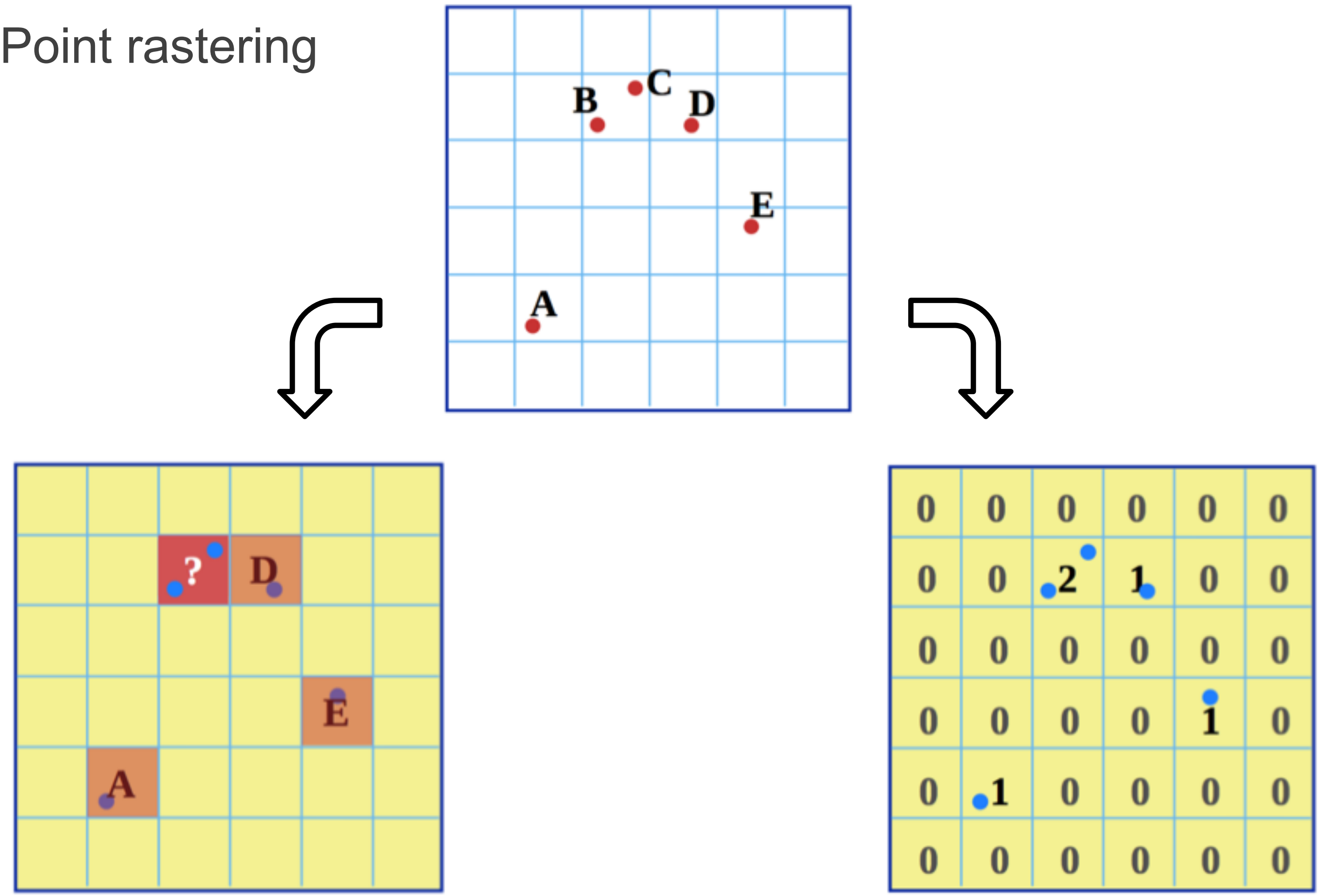
Vector to raster conversion

Presence coding to raster lines



Vector to raster conversion

Point rastering



Attribute coding

Presence coding

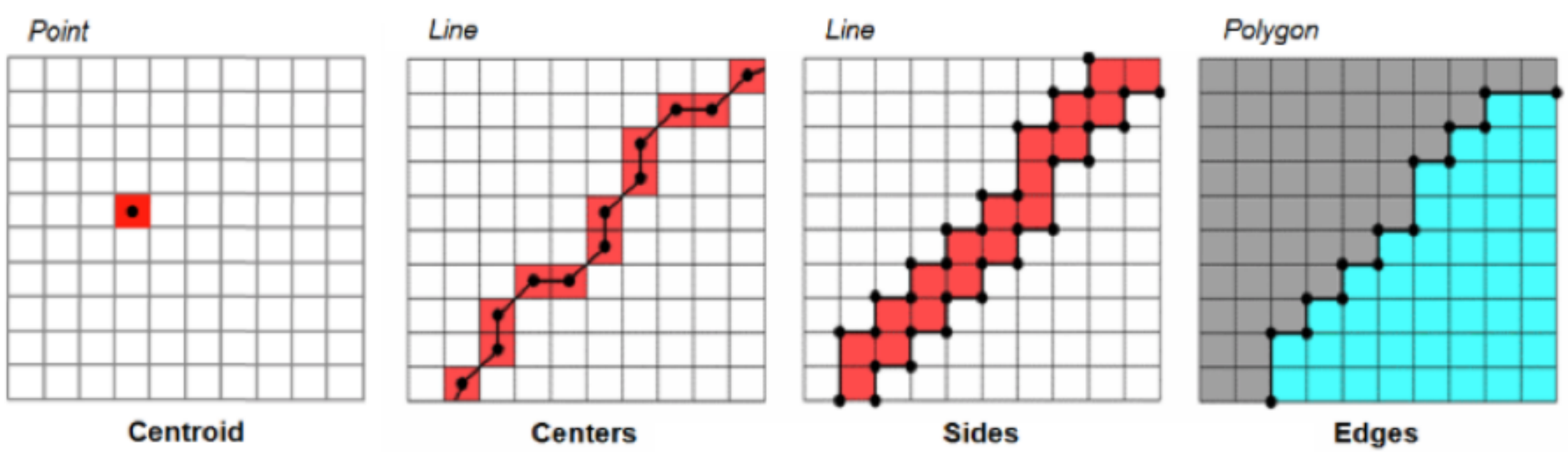
Raster to vector conversion

Why?

- The data are not in their “natural” format

  e.g. image of maps

- The current format is not suitable for efficient processing



Raster to vector conversion

Points and lines can be vectorized quite easily

General workflow:

1. Create a new vertex for each pixel
2. Connect neighbouring pixels
3. Simplify the geometry (semi-aligned nodes removal)

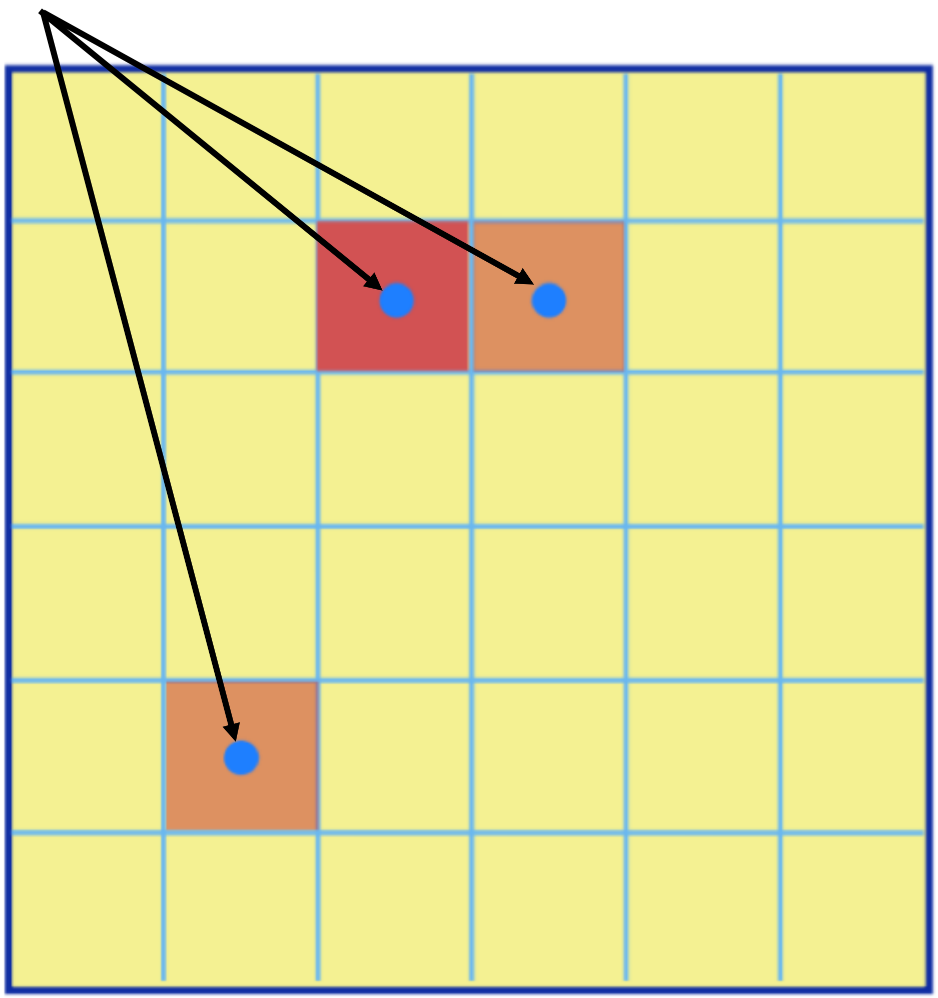
Area vectorization is usually complex and may require human intervention

General workflow:

1. Classification (based on pixel similarity)
2. Aggregation
3. Vectorization (boundary following)

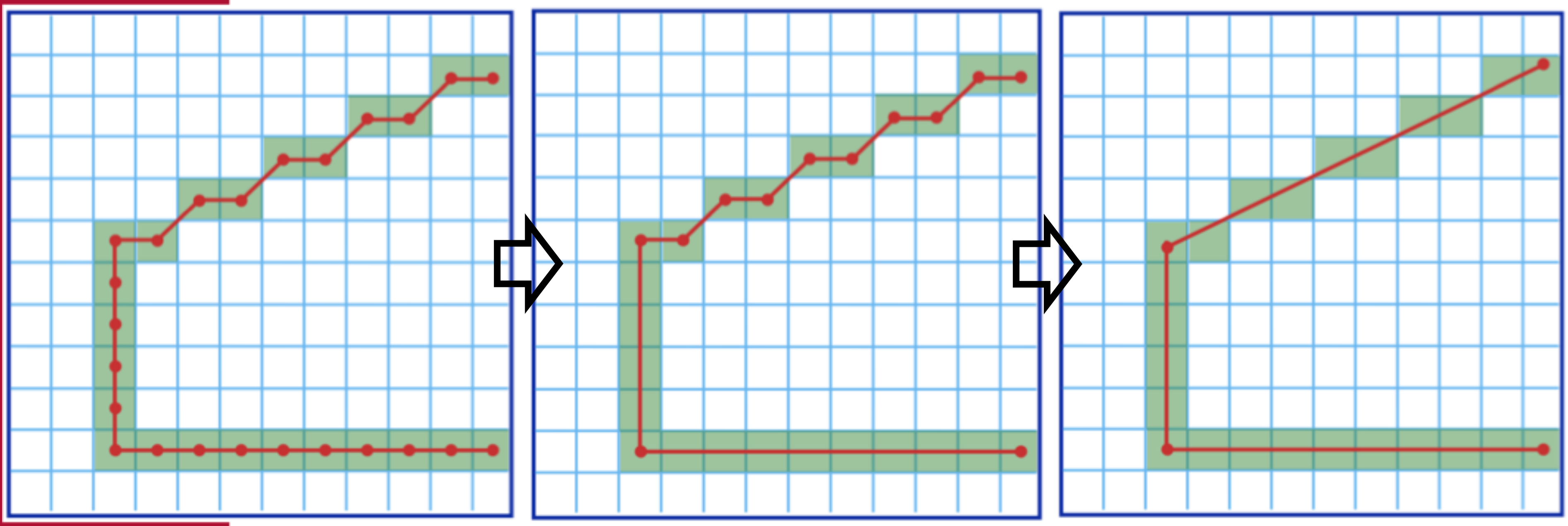
Raster to vector conversion

Point vectorization



Raster to vector conversion

Line vectorization



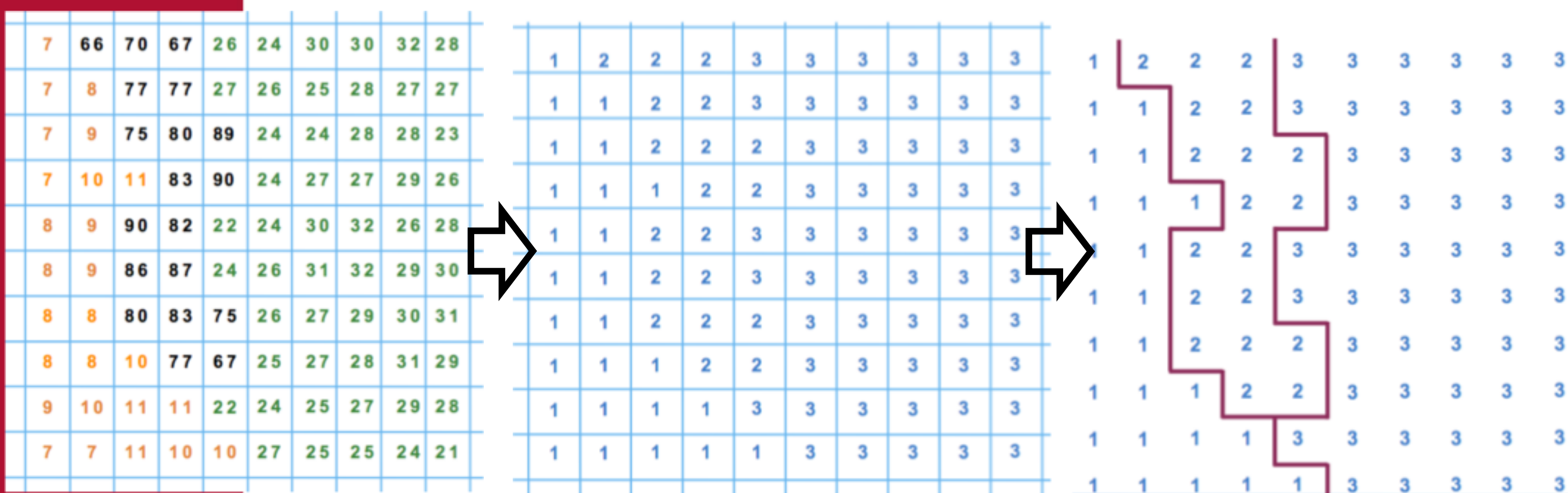
Aligned nodes removal

Semi-aligned nodes removal

Università
Ca' Foscari
Venezia

Raster to vector conversion

Area vectorization



Pixels are **classified** considering their similarity

Image is **partitioned** into homogeneous regions

Map scale

Map scale is the proportion between a distance on a map and a corresponding distance on the ground

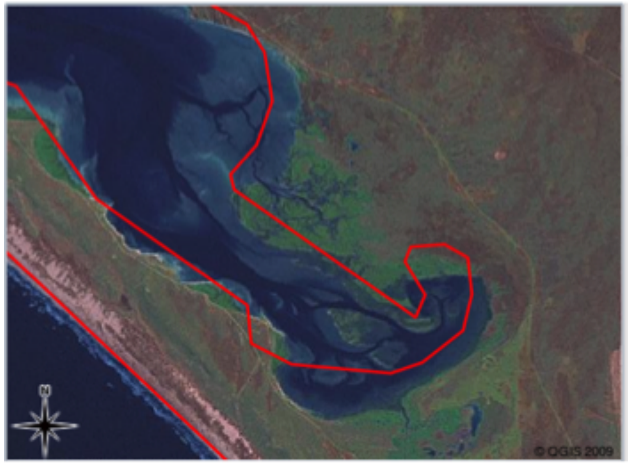
Conventionally expressed as a "representative fraction" like 1:10000

Scale is an important issue to consider when working with vector data in a GIS as it affects the level of detail

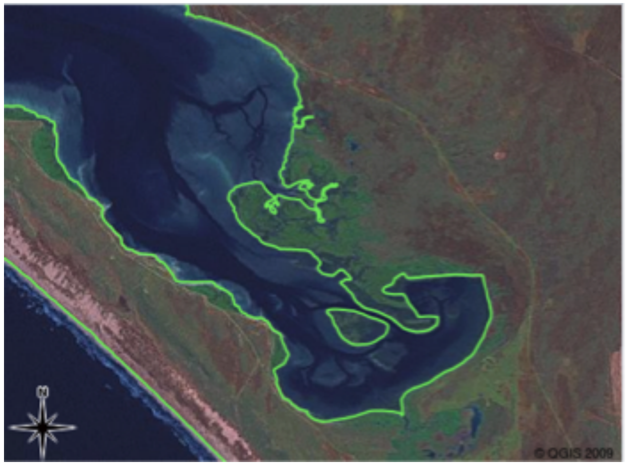
Map scale

Maps have different scales. When importing vector data from a map into a GIS environment (for example by digitising paper maps), the digital vector data will have the same scale issues as the original map

Issues can arise from making a poor choice of map scale



Vector data digitised from a small scale (1:1000 000) map



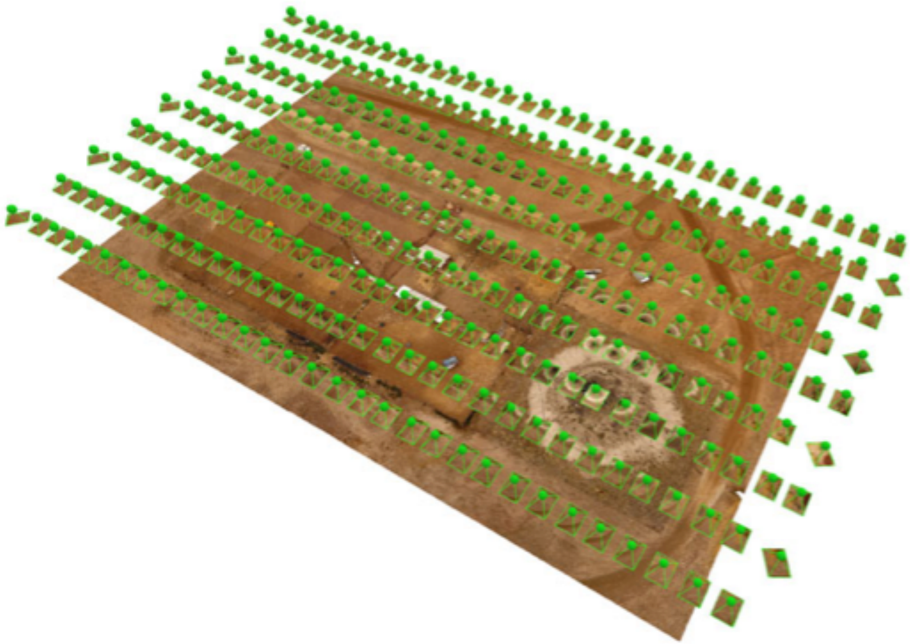
Vector data digitised from a large scale (1:50 000) map

Scale consistency

In topographic maps the scale is consistent across each map and within each map series. When working with GIS, this may not be true in general.

Example: In aerial imagery the scale is consistent only if the images have been ortorectified

In general, the scale of an aerial image varies from place to place as a function of the elevation of the terrain/aircraft shown in the scene

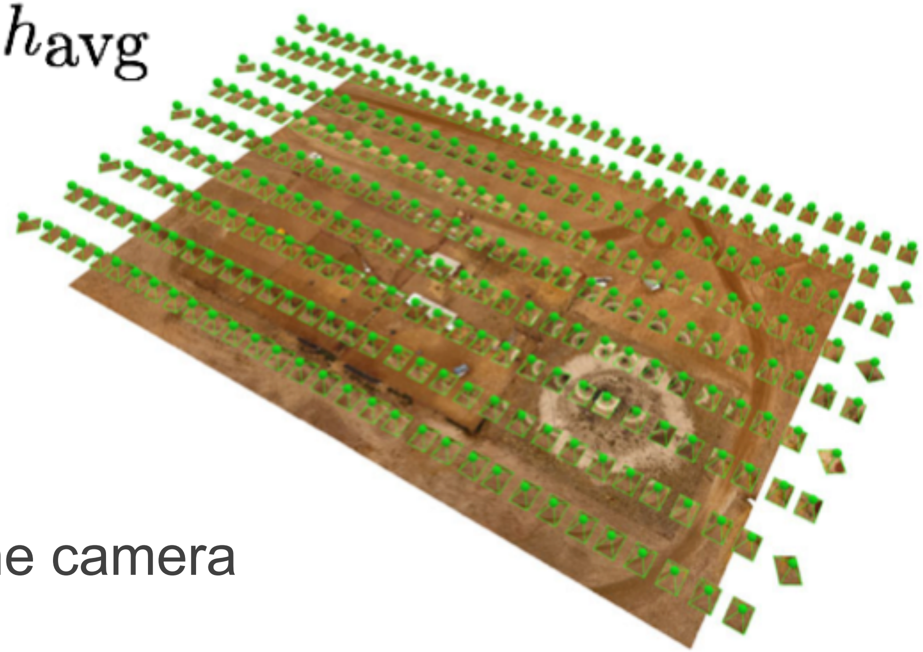


Università Ca' Foscari Venezia

Scale of aerial images

Average scale of an unrectified air photo:

$$S_{\text{avg}} = \frac{f}{H - h_{\text{avg}}}$$



f : Focal length of the camera

H : Aircraft altitude

h_{avg} : Average terrain elevation

Map projections

A traditional method of representing the earth's shape is the use of globes

Pro:

- Preserve the majority of the earth's shape and illustrate the spatial configuration of continent-sized features

Cons:

- Difficult to carry in one's pocket
- convenient to use at extremely small scales (e.g. 1:100 million)

Map projections

To overcome these limitations, cartographers have developed a set of techniques called **map projections**

GOAL: show, with reasonable accuracy, the spherical earth in two-dimensions.

Each map projection is a **representation** of the reality designed to not only represent features, but also their shape and spatial arrangement

Università Ca' Foscari Venezia

Map projections

Each kind of map projection has **advantages** and **disadvantages**:

- The projection to a 2D plane introduce distortions that are more or less significative depending on the specific projection and the scale
- At very small scales (like a map of a city), the distortion is negligible because the earth can be approximated as being locally flat
- The amount of distortion can vary in different locations of the map
- Some projections are good for mapping areas with a large East-West extent, and some are better for mapping areas with a large North-South extent

Università Ca' Foscari Venezia

Map projections

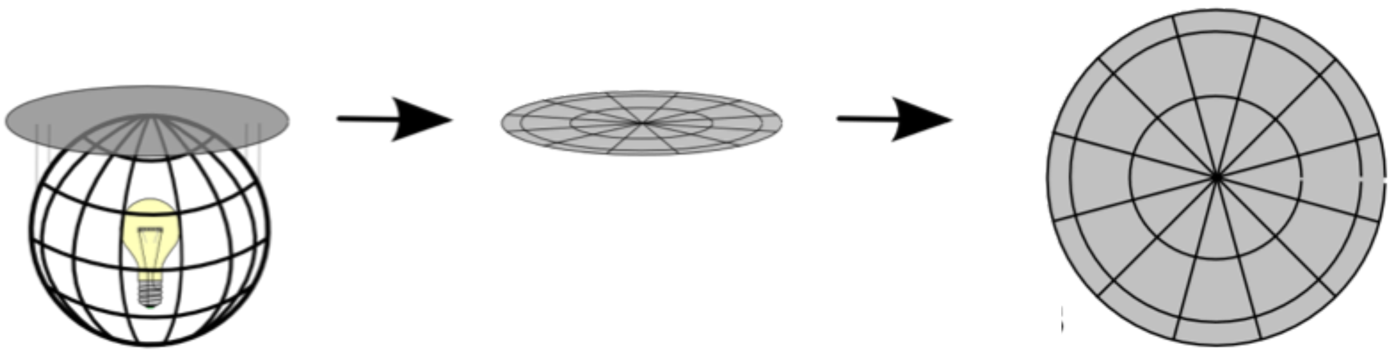
How do we create map projections?

Imagine to put a light source inside a transparent globe on which opaque earth features are placed

What shadow would be cast?

Map projections

How do we create map projections?

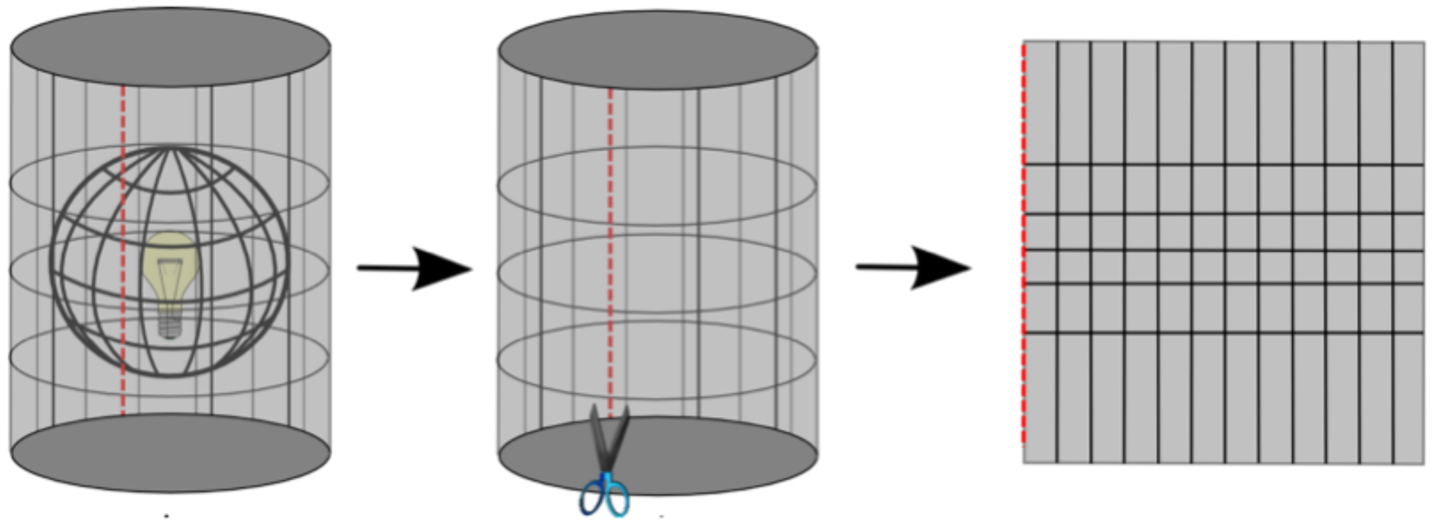


If we simply put a plane close to the the globe we obtain the class of **planar projections**

Università
Ca' Foscari
Venezia

Map projections

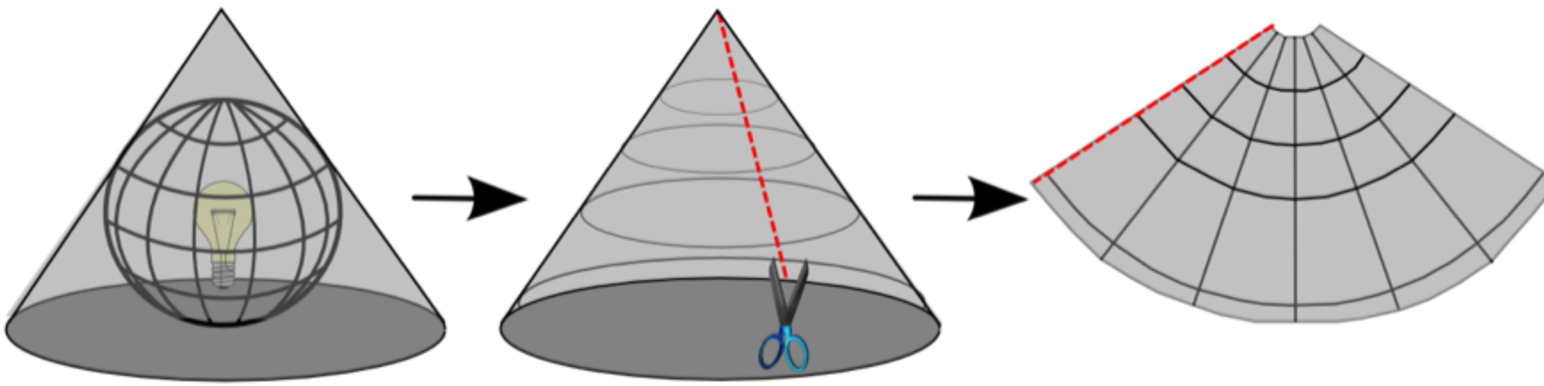
How do we create map projections?



If we wrap the globe with a cylinder, we draw the shadows and finally unwrap the cylinder to a flat surface we obtain the class of **cylindrical projections**

Map projections

How do we create map projections?



If we wrap the globe with a cone, we draw the shadows and finally unwrap the cone to a flat surface we obtain the class of **conical projections**

Università Ca' Foscari Venezia

Map projections

A map projection can introduce distortions with respect to the angles, the distances or the areas.

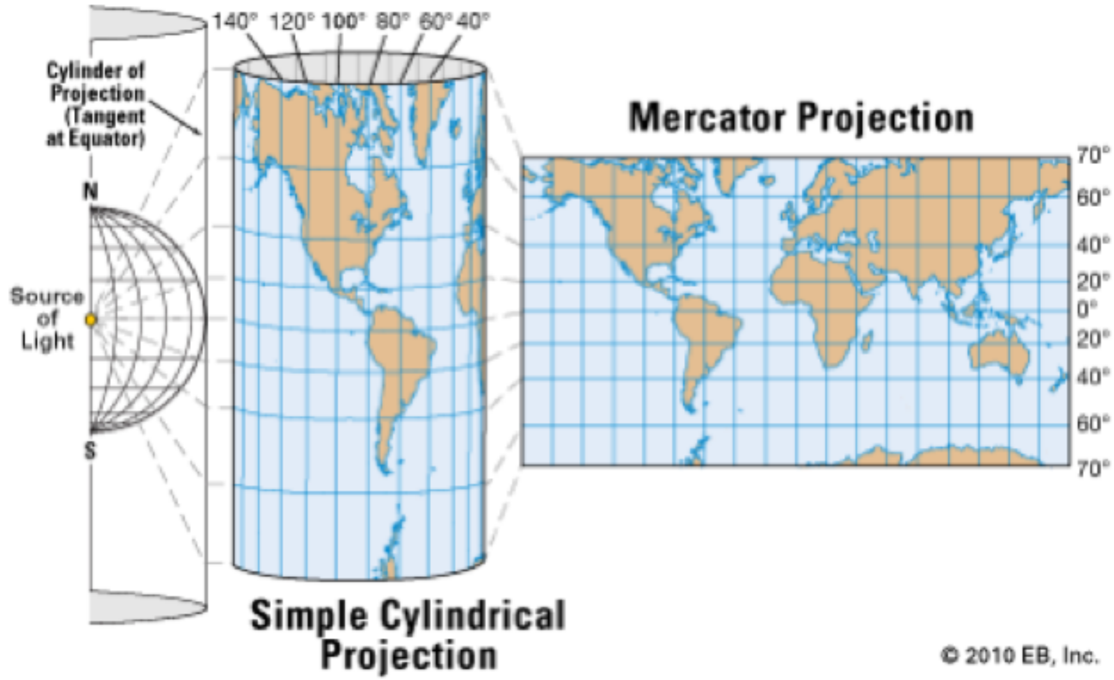
Conformal map: preserve angles locally

Equidistance map: maintain accurate distances from the centre of the projection or along given lines

Equal-areas map: preserve areas

Some projection distort may be a compromise that distorts area, distance and angular conformity but within some acceptable limit

Conformal map projections



Mercator projection

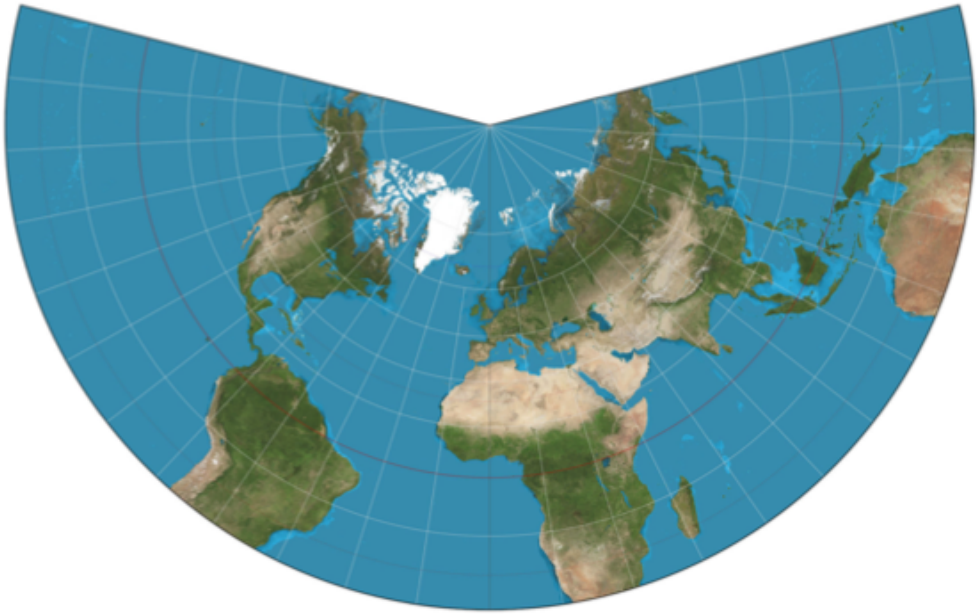
Type: Cylindrical

Used for: Navigation/meteo

Characteristics: Lines of constant bearing (rhumb lines) are straight, aiding navigation. Areas inflate with latitude, becoming so extreme that the map **cannot show the poles**

Università Ca' Foscari Venezia

Conformal map projections



Lambert conformal conic (LCC)

Type: Conic

Used for: Aeronautical charts

Characteristics: Pilots use aeronautical charts based on LCC because a straight line drawn on a Lambert conformal conic projection approximates a great-circle route between endpoints for typical flight distances.

Università Ca' Foscari Venezia

Equidistant map projections

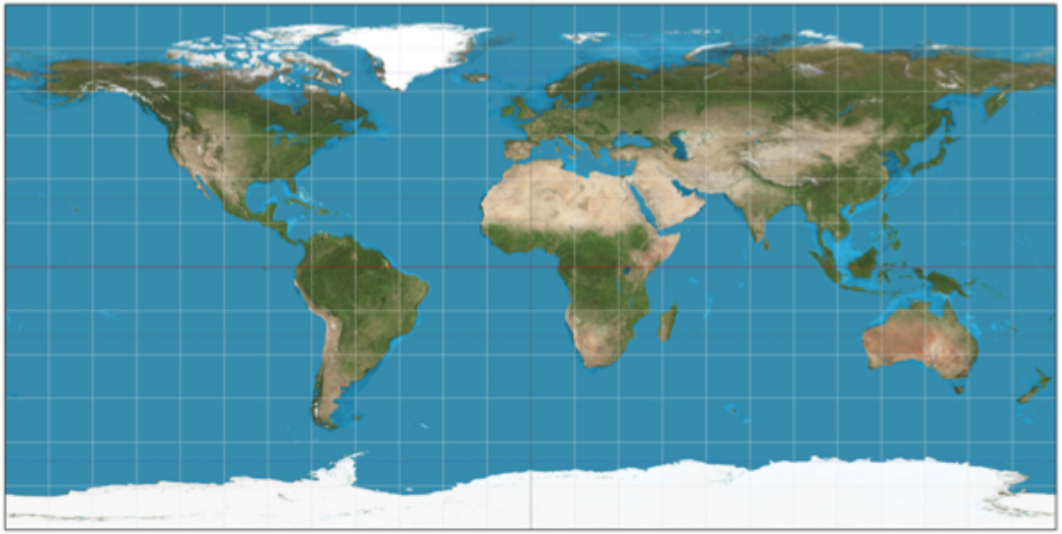


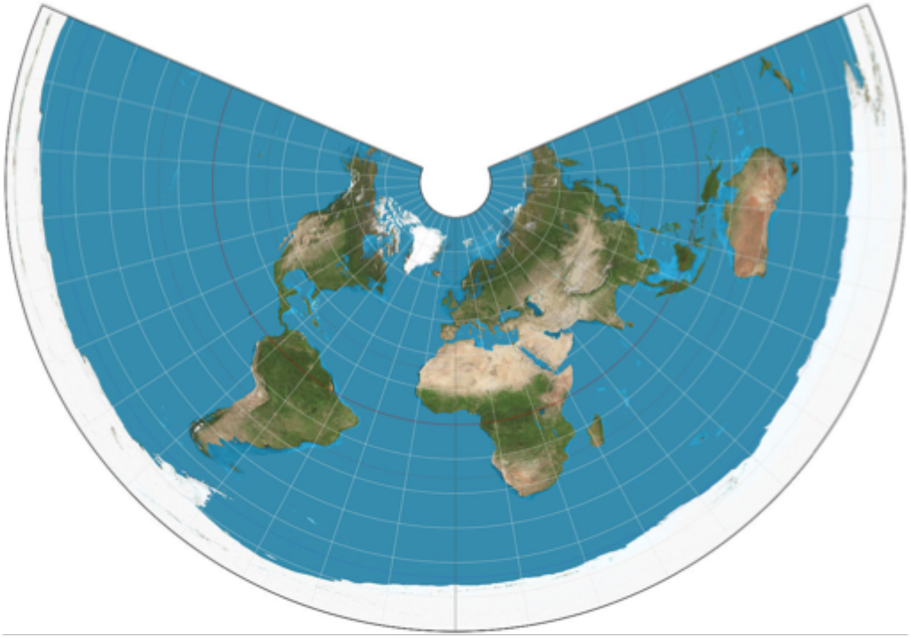
Plate Carree Equidistant

Type: Cylindrical

Used for: Radio and seismic mapping, navigation

Characteristics: The projection maps meridians to vertical straight lines of constant spacing (for meridional intervals of constant spacing), and circles of latitude to horizontal straight lines of constant spacing. The standard parallel is the equator

Equidistant map projections



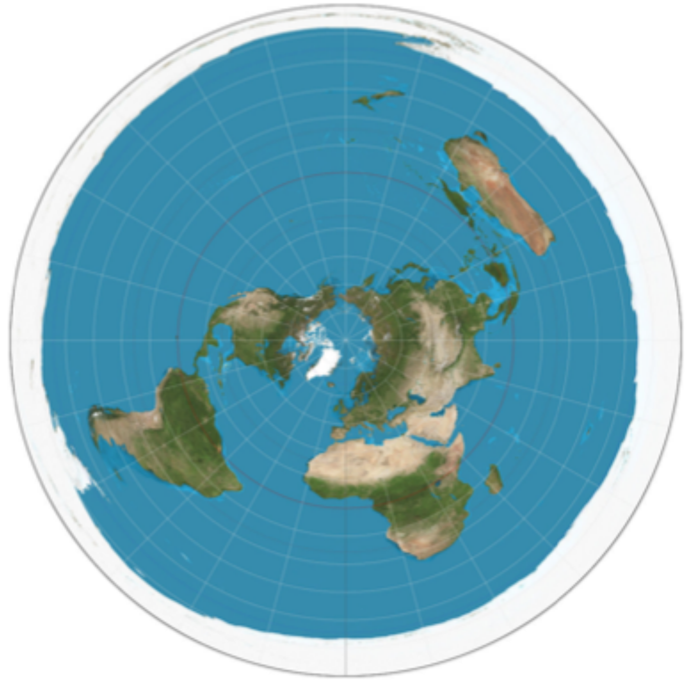
Equidistant conic

Type: Conic

Used for: maps of regions elongated east-to-west (such as the continental United States)

Characteristics: distances along the meridians are proportionately correct, and distances are also correct along two standard parallels that the mapmaker has chosen

Equidistant map projections



Azimuthal Equidistant

Type: Planar

Used for: the USGS in the National Atlas of the United States of America

Characteristics: Distances from centre are conserved.

Equal-area map projections

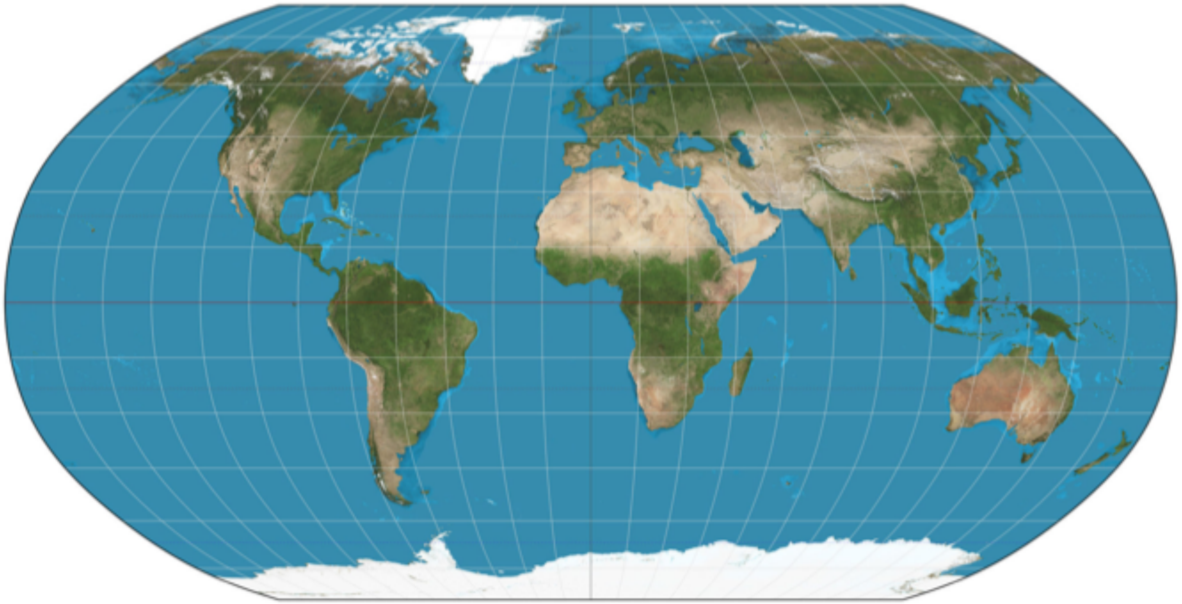
Mollweide

Type: Cylindrical

Used for: Global maps of the world or night sky.

Characteristics: Meridians are ellipses on the map plane

Compromise map projections



<u>Robinson</u>

Type: Pseudocylindrical

Used for: National Geographic

Characteristics: It was specifically created in an attempt to find a good compromise to the problem of readily showing the whole globe as a flat image

Università Ca' Foscari Venezia

Università Ca' Foscari Venezia

Coordinate reference systems

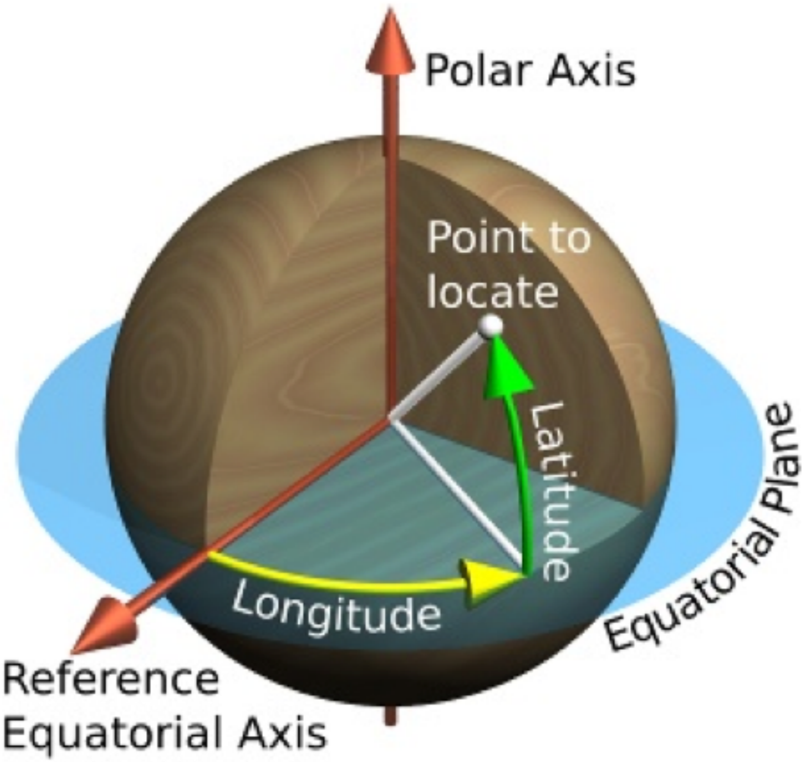
Locations on the Earth's surface are measured and represented in terms of **coordinates**.

A coordinate is a set of two or more numbers that specifies the position of a point, line, or other geometric figure in relation to some reference system, called **coordinate reference system** (CRS).

CRS can be divided into **geographic** coordinate reference systems and **projected** coordinate reference systems (also called Cartesian or rectangular coordinate reference systems)

Geographic CRS

Designed specifically to define positions on the Earth's surface (assumed to be an ellipsoid). It uses **two curved measurement scales** (latitude/longitude)



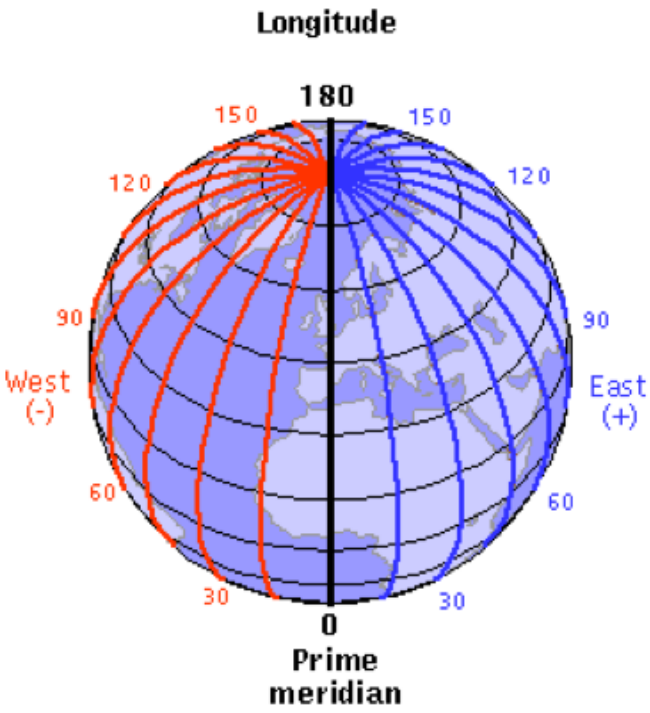
Geographic CRS

Longitude

E-W angle between the prime meridian and a second meridian that intersects the point of interest.

+180 (or 180° E) to -180° (or 180° W)

180° East and West longitude together form the International Date Line.



Università Ca' Foscari Venezia

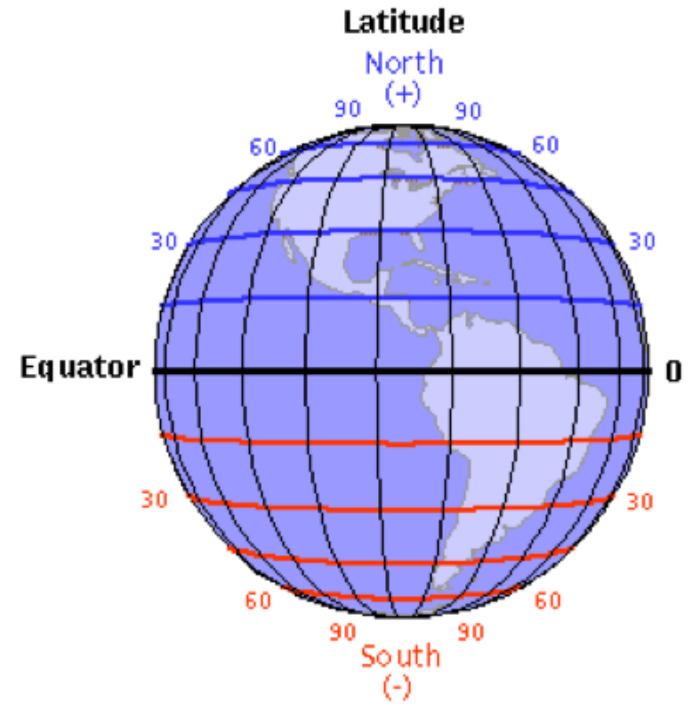
Geographic CRS

Latitude

N-S angle subtended at the center of the Earth between two imaginary lines, one that intersects the equator and another that intersects the point of interest.

+90 (or 90° N) to -90° (or 90° S)

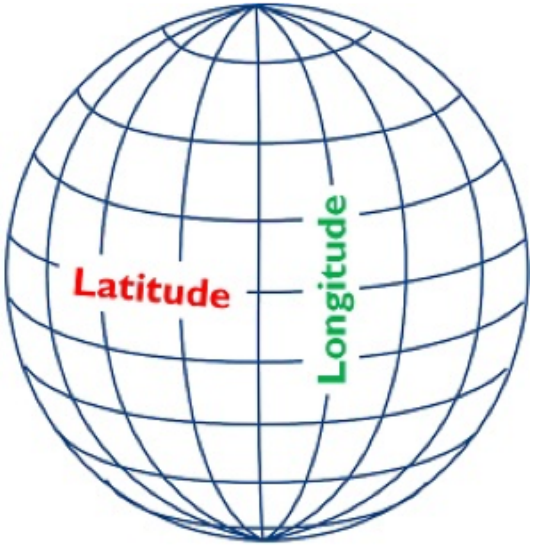
A line of latitude is also known as a parallel.



Geographic CRS

Latitude/Latitude

- the distance between the lines of latitude is the same (60 nautical miles)
- Only at the equator, the distance represented by one line of longitude is equal to the distance represented by one degree of latitude
- Distance represented by one line of longitude decreases to zero at the poles

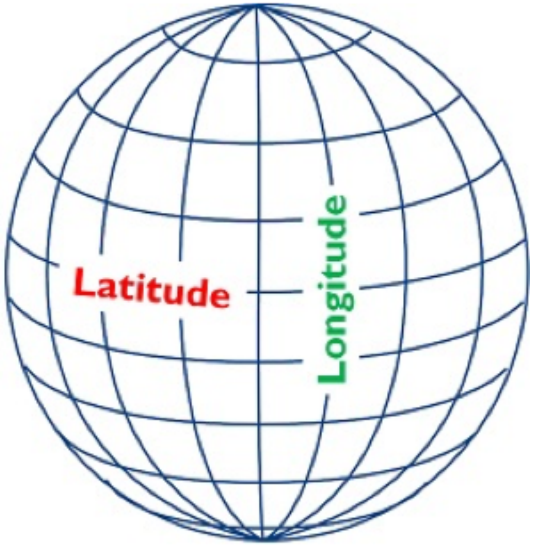


Geographic CRS

Measuring Latitude/Latitude

- Decimal degrees: 39.868055 -75.251388
- Degrees, minutes, seconds:

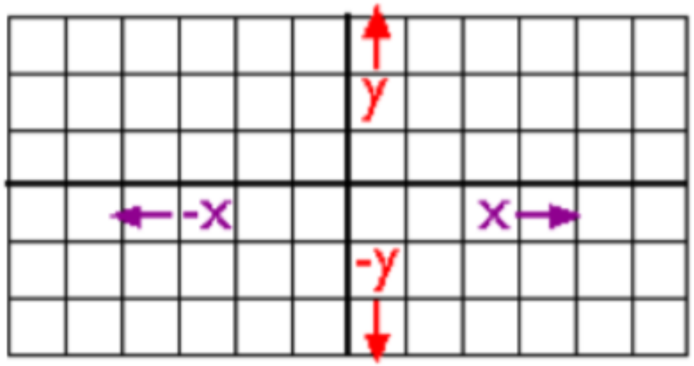
39° 52’ 5’’ N, 75° 15’ 5’’ W



Università Ca’ Foscari Venezia

Projected CRS

A Projected coordinate system is simply a grid.
The point at which both x and y equal zero is called the origin of the coordinate system.
All other positions are specified as distances of the projection along the axes relative to the origin.

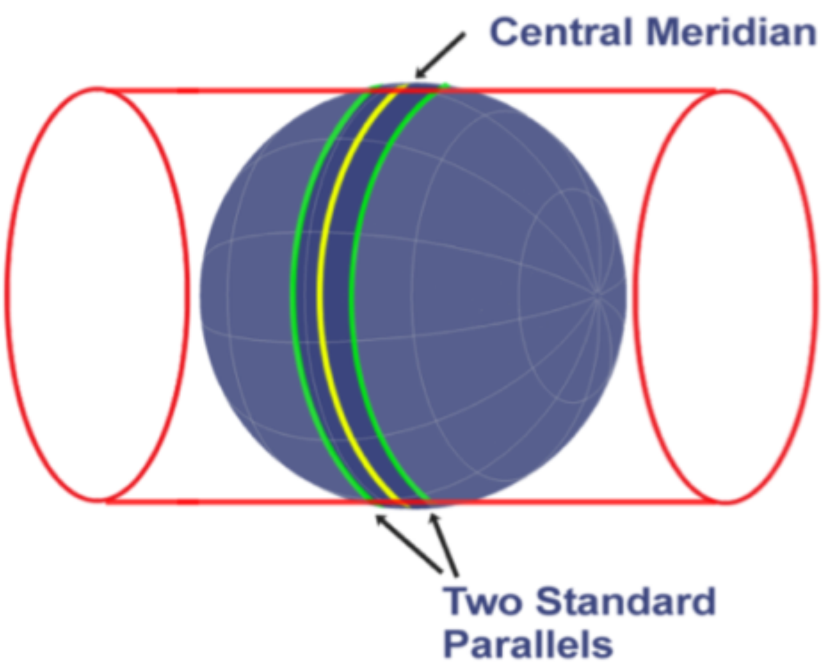


Simple to be used but it is not simple to map the whole earth to a 2D plane (as we have seen before…)

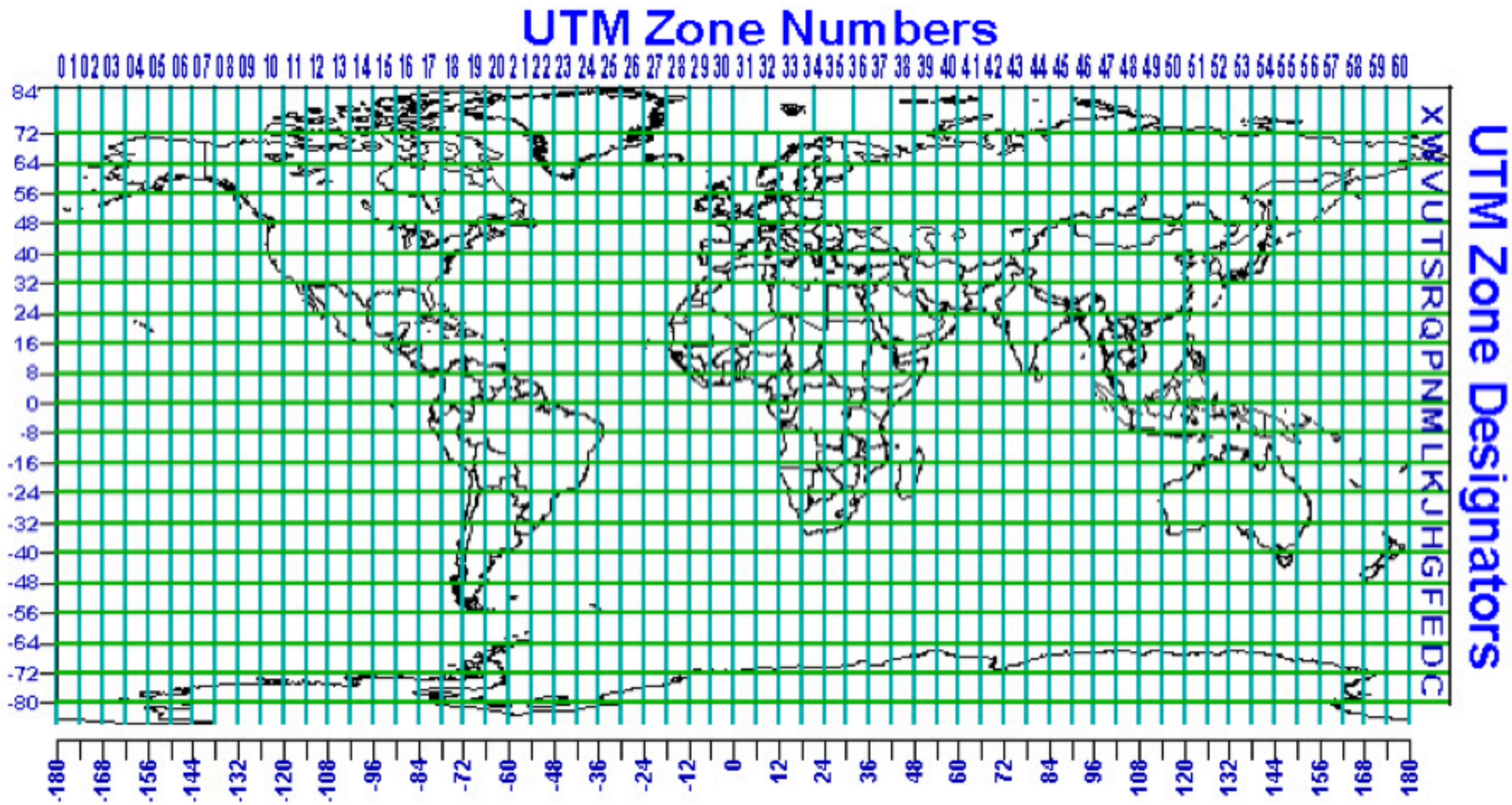
Projected CRS

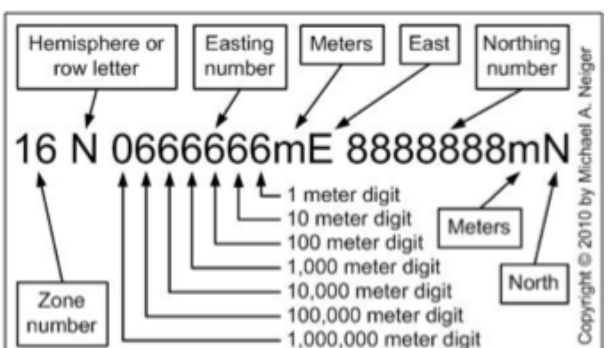
Universal Transverse Mercator (UTM)

The system use multiple map projections. It divides the Earth into 60 zones, each being a six-degree band of longitude, and uses a secant transverse Mercator projection in each zone



Universal Transverse Mercator





The point of origin of each UTM zone is the intersection of the equator and the zone's central meridian. To avoid dealing with negative numbers, the central meridian of each zone is defined to coincide with 500000 meters East

Geodetic Datum

The absolute (though not the relative) position of a point on a surface, depends upon the shape of the surface

Horizontal datums define the geometric relationship between a coordinate system grid and the Earth's surface.

Coordinate system grid is aligned with an ellipsoid that approximates the Earth's shape. The most common is called **WGS84**, based on a reference ellipsoid developed in 1984