Geodata models and core concepts for analysis

Spatial Data Analysis and Simulation modelling,
2020, Simon Scheider

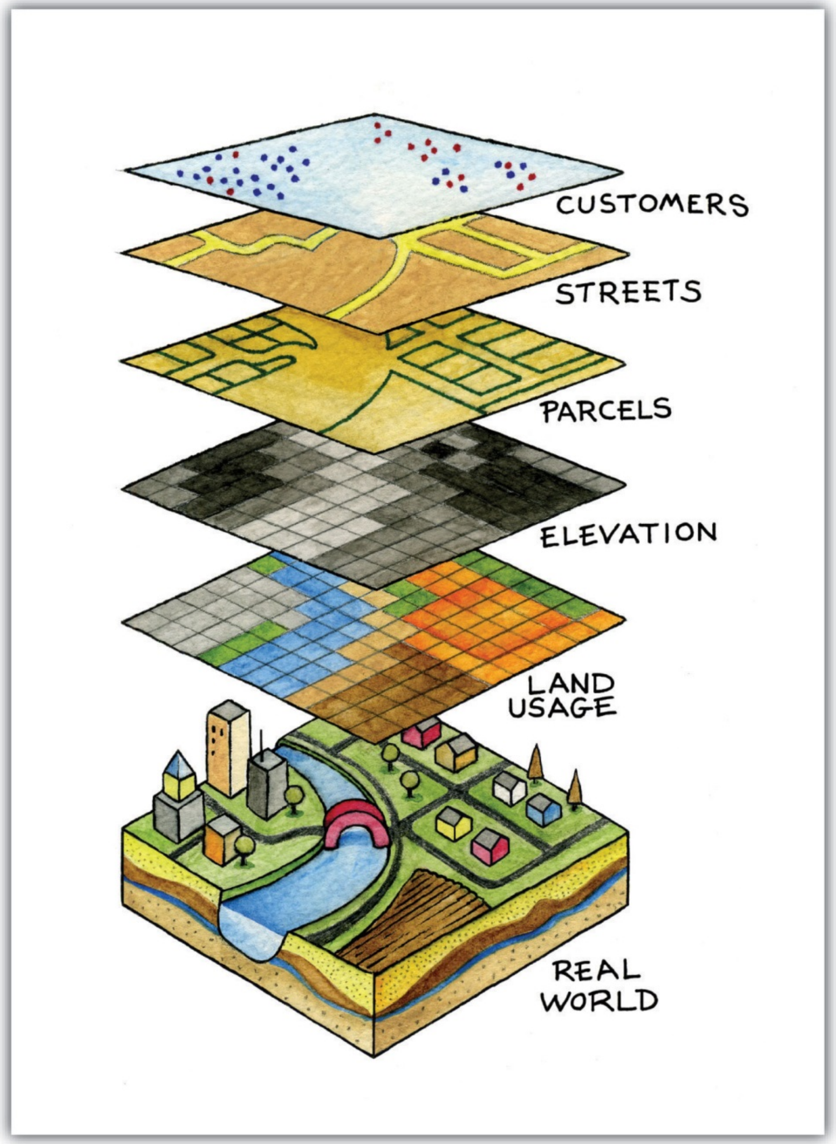
Outline

- Principles of spatial data transformation
  - Layer principle and overlay in GIS
  - Analysis process
  - Different types of overlay
  - Different methods for point data
- Geometric data models
  - Geometric layer models
  - Vector geometry models
  - Basic geometric manipulations
- Core concept data types
  - Core concepts
  - Core concept data types (CCD)
  - Examples
  - Constraints for possible transformations

Principles of spatial data transformation

Layer principle of GIS

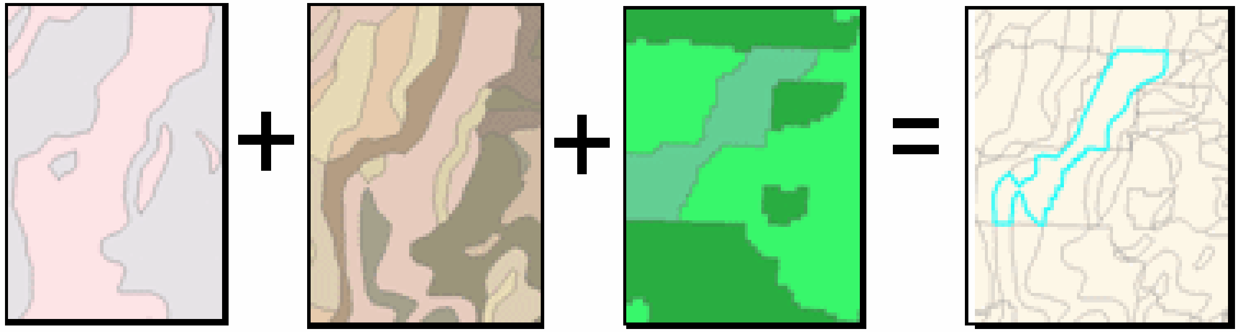
- Fundamental principle
- Layers (vector, raster) are overlaid
  - To derive new layers
  - To spatially analyse landscape
  - To aggregate and summarize data
- Overlay methods for vector and raster data differ, however (and are not always called overlay)



layers closely related to quality as well

Vector overlay analysis in GIS

- Development suitability analysis
- Steep slopes, soil, and vegetation type given as polygons



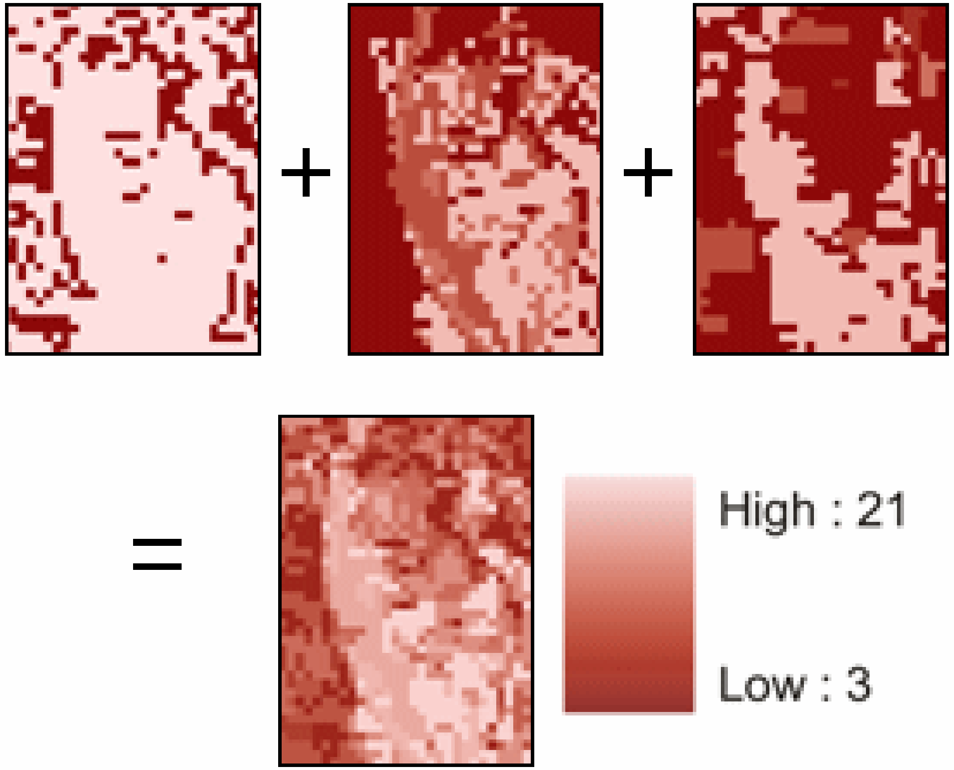
FID	Shape*	FID_soils	CODE	CLASS	FID_sl	SLOPE	FID_veg	DET_TYPE
3039	Polygon	508	38F	6	0	60	117	A
3040	Polygon	508	38F	6	0	60	119	SS
3041	Polygon	508	38F	6	0	60	157	U
3042	Polygon	508	38F	6	0	60	158	A
3043	Polygon	508	38F	6	0	60	160	FC

Raster overlay

same kind of outcome scenario

- Three raster layers (steep slopes, soils, and vegetation) are ranked for development suitability on a scale of 1 to 7.
- Sum = suitability for development

result is raster layer which displays the sum for each vell



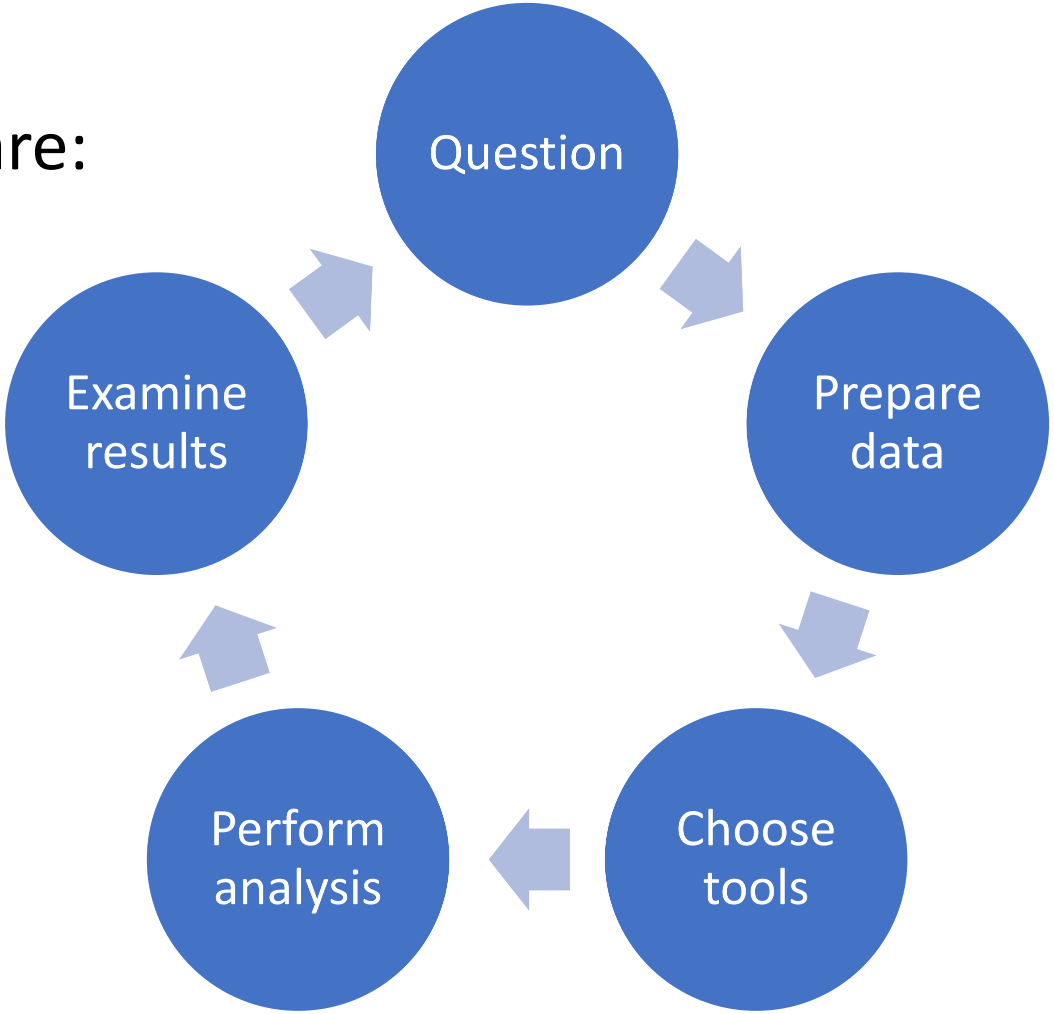
Analysis as a process

The five steps in the analysis process are:

- Frame the question
- Explore and prepare data
- Choose analysis methods and tools
- Perform the analysis
- Examine and refine results

How to decide on each step?

question to be answered in this lecture



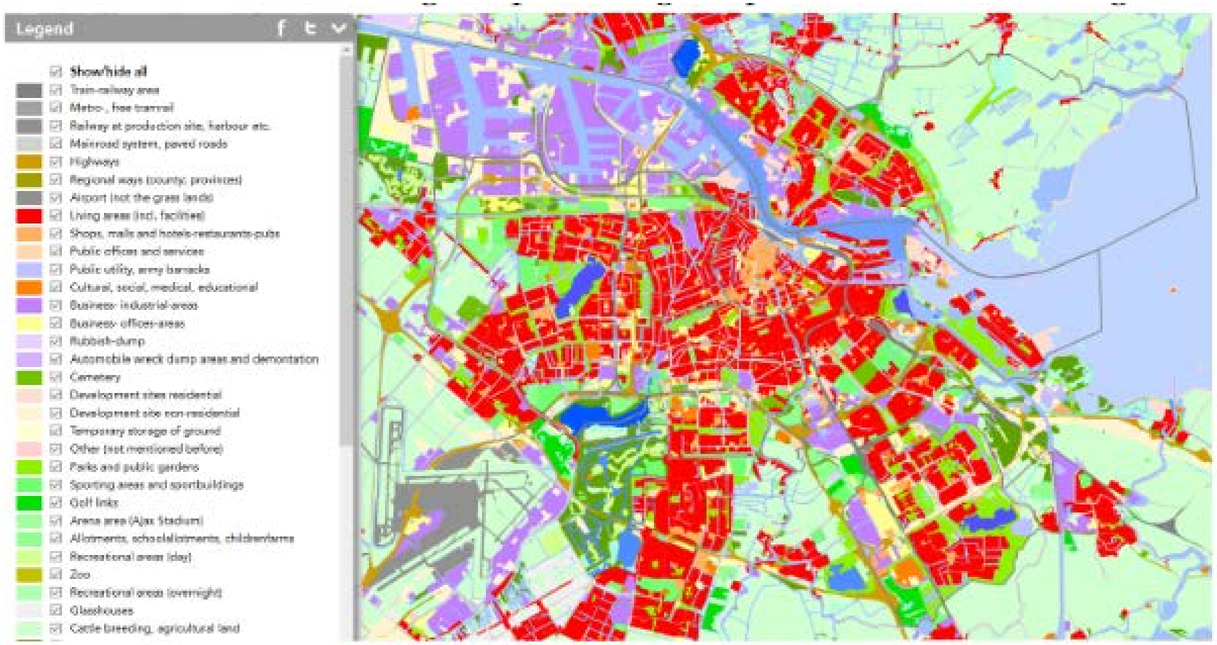
Breaking down an analytic question

into subquestions

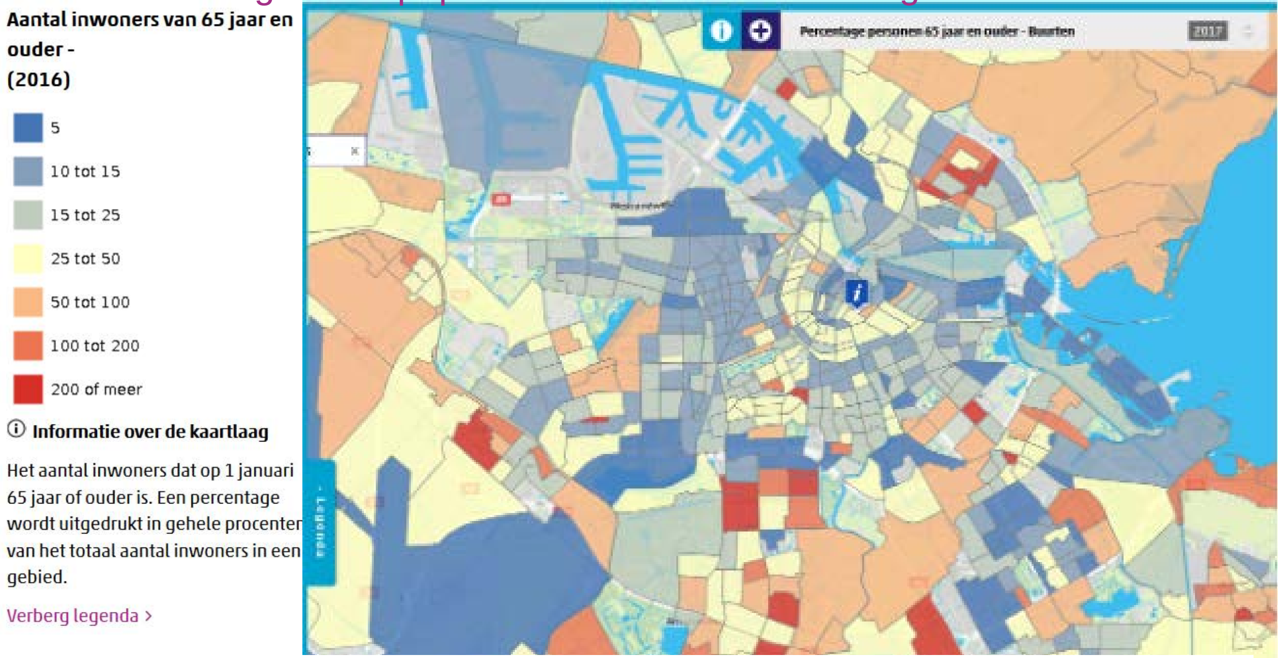
What is the proportion of green space in Amsterdam?

What is the amount of elderly people living in PC4 areas in Amsterdam?

making use of population statistics for the neighborhoods



Landuse map (Grondgebruik) of the Amsterdam Municipality

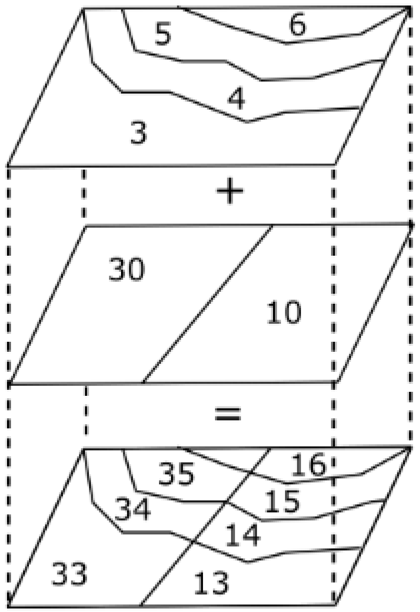


CBS Buurt (neighborhood) statistics

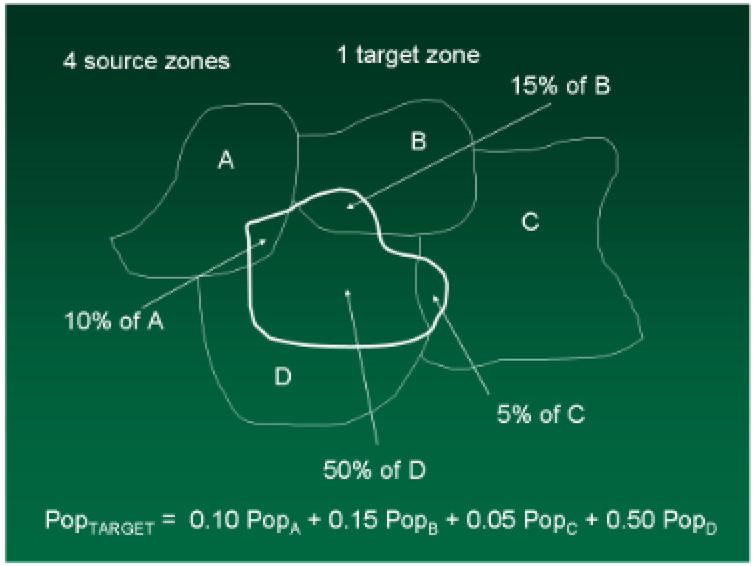
Selecting overlay methods for analysis

- Which overlay method could be used for assessing the amount of elderly people living in PC4 areas?

combines layers
geometrically (intersects)
attributes are summed up



(a) Vector overlay method.



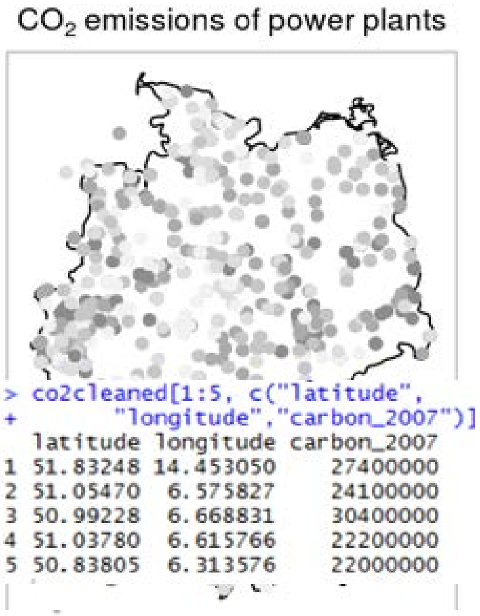
(b) Simple areal interpolation method. Image by kind permission of Michael Goodchild.

inout polygons contain population data - interest is accesing population for thick white line
weighted sum based on amount of overlap to estimate target polygon

Which one is adequate?

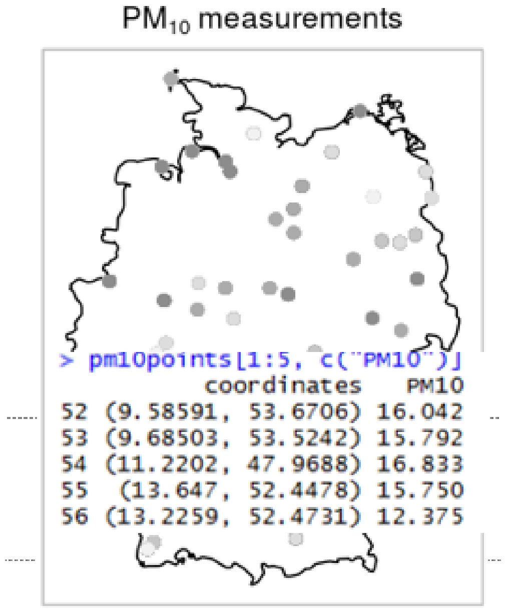
Methods for spatially combining polygon data. Which one is applicable for analysing a given polygon data set?

How much polluted is Germany?

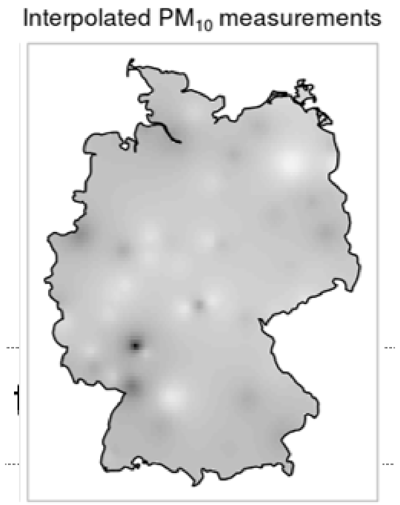


Given: some point datasets in R with pollution measurements.

How to assess pollution intensity of Germany?
By interpolation? Aggregation?

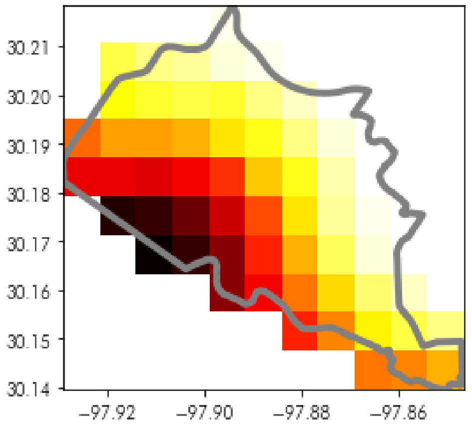


sum up?



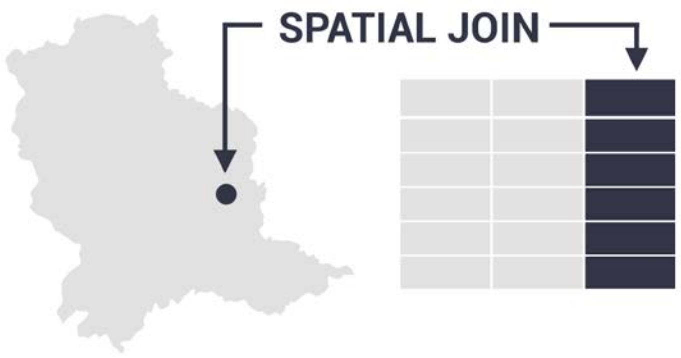
Aggregation or interpolation?

- … which GIS methods are appropriate?



Zonal
statistics?

continupus surface represented as raster

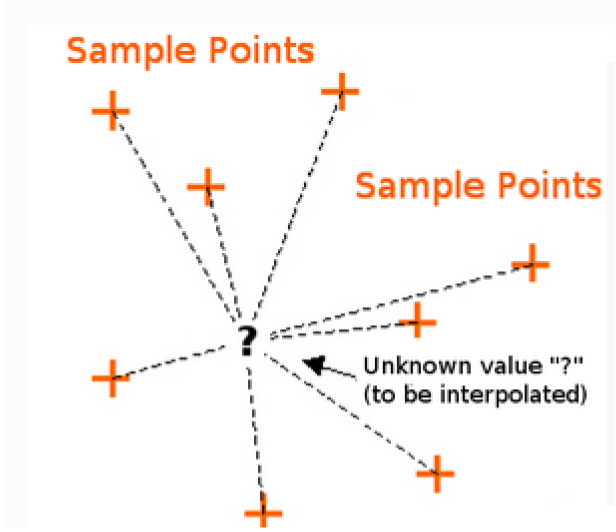


Spatial join?

next week

table of measurements

attribute/geometries are joined to target geometry



Point
Interpolation?

Geometric data models

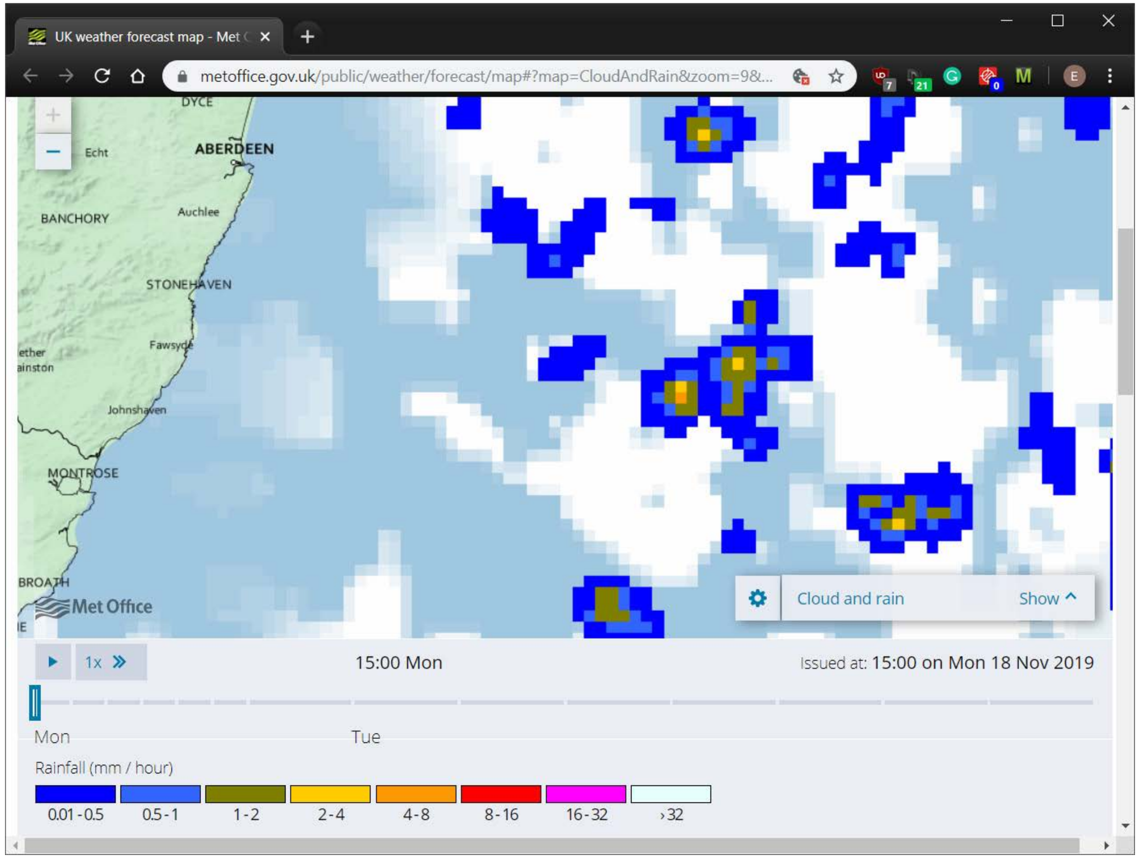
resselation is a region layer where the layers are non overlapping and covering the whole area

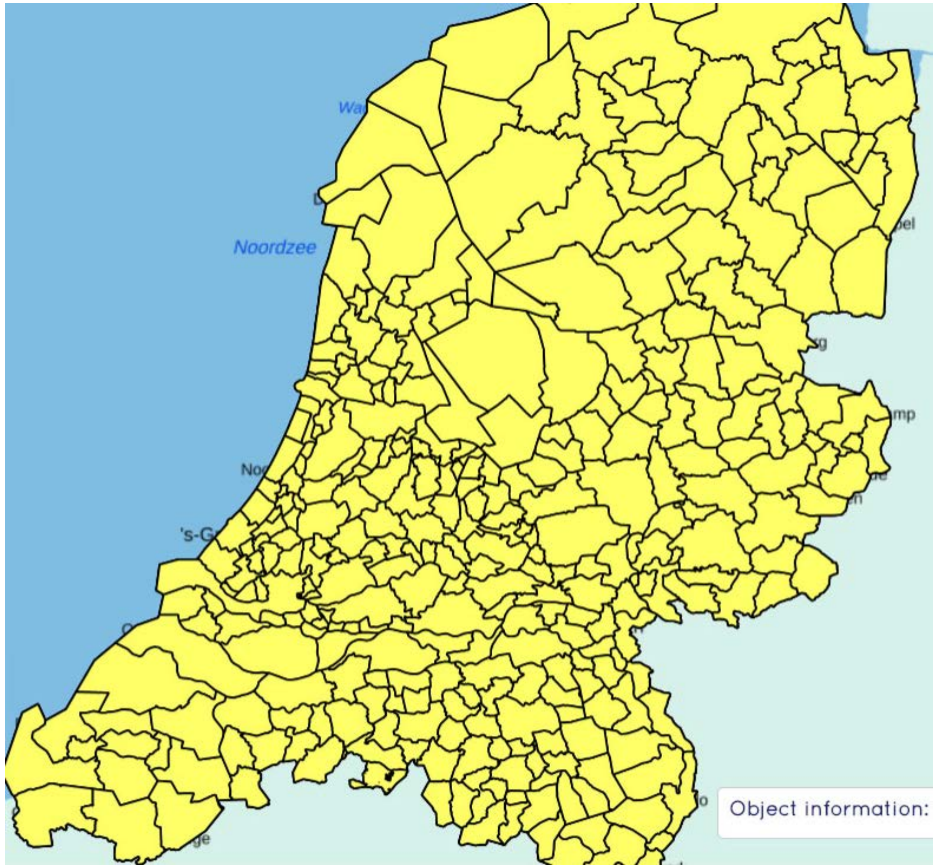
Tessellations

joined covering and non overlapping layer of polygons

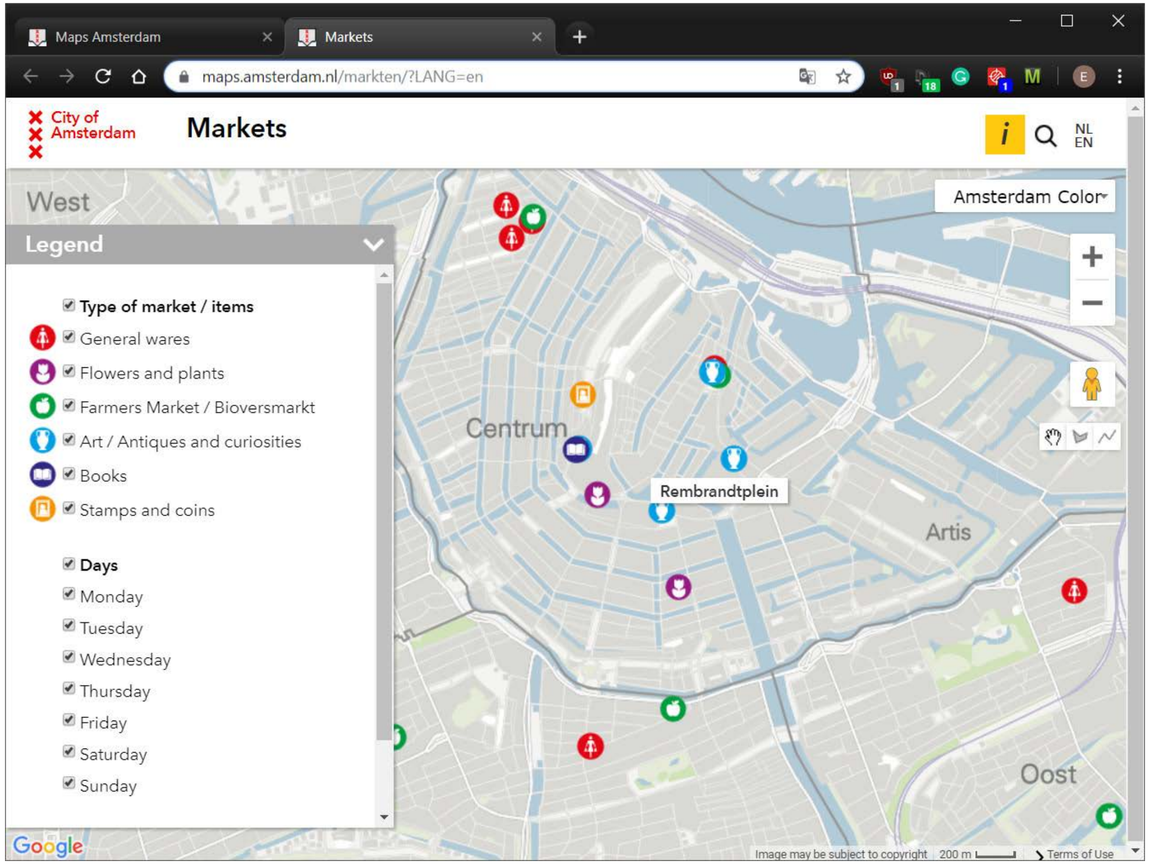
both are resselations

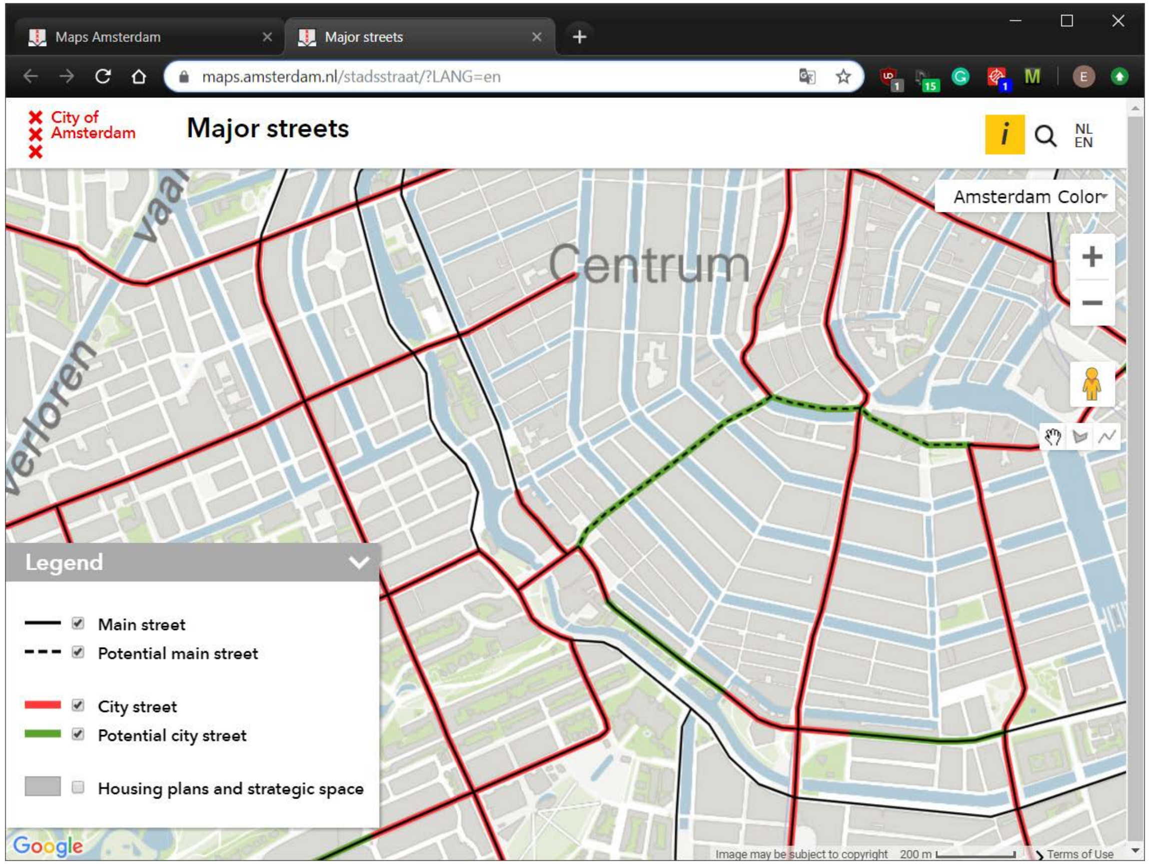
- Regular (Raster) and irregular (Vector)





Point and Line datasets



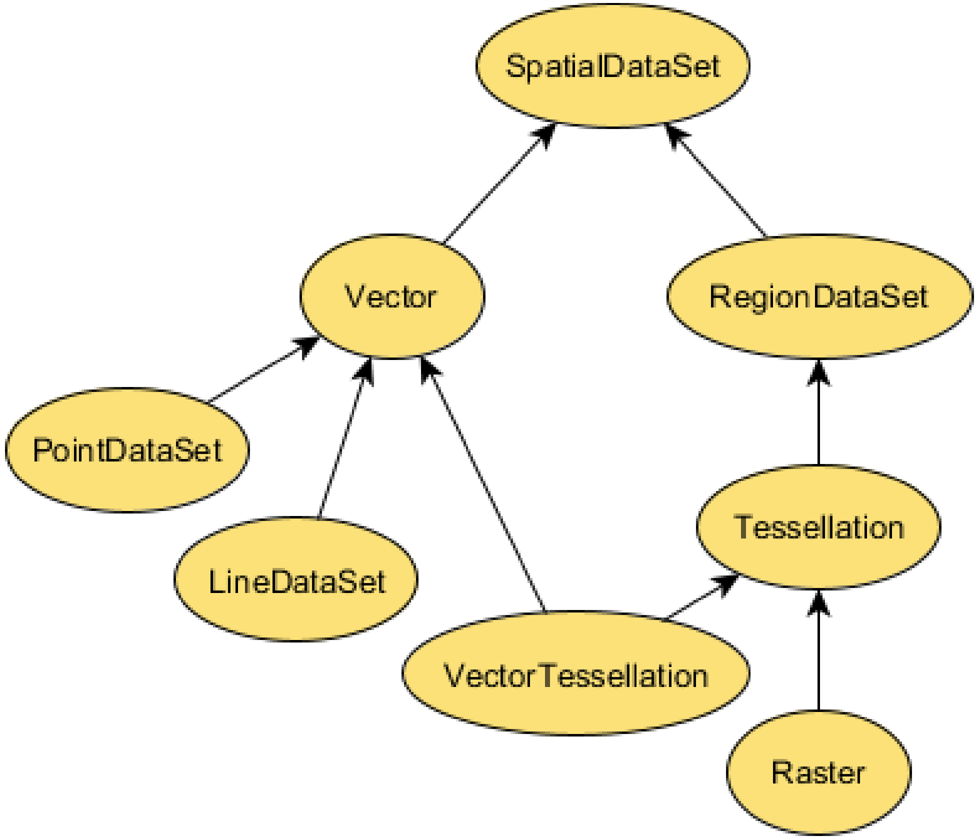


Generalized geometric layer model

buildings in a city would be an incomplete layer. This example is a region dataset so building dataset doesn't cover the whole area.

- *Region*: dataset where the geometric primitives are regions (polygons or cells)
- *Tessellation*: A specific form of *Region* Regions are non-overlapping, and fill the entire extent of the dataset without gaps
- *Raster*: A special kind of *Tessellation* , where the regions are all squares (cells)
- *VectorTessellation*: Irregular tessellation

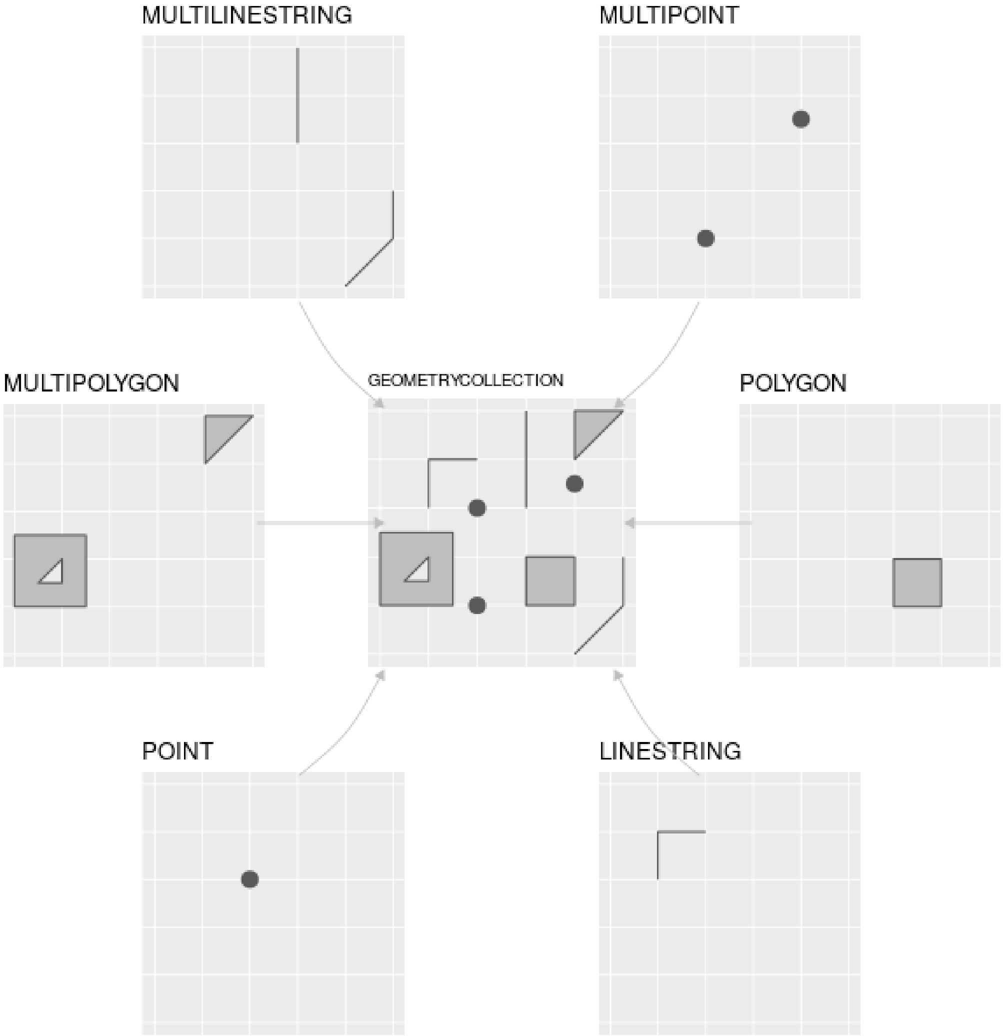
hierarchy of different datasets in geodata



Vector geometry model (Simple Features)

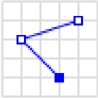
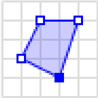
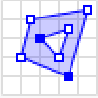
models of how it is encoded

(WKT standard)



important!

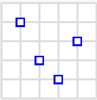
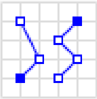
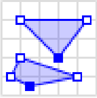
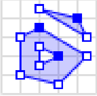
Geometry primitives (2D)

Type	Examples	
Point	POINT (30 10)	
LineString	LINESTRING (30 10, 10 30, 40 40)	
Polygon	POLYGON ((30 10, 10 20, 20 40, 40 40, 30 10))	
	POLYGON ((35 10, 10 20, 15 40, 45 45, 35 10), (20 30, 35 35, 30 20, 20 30))	

with hole

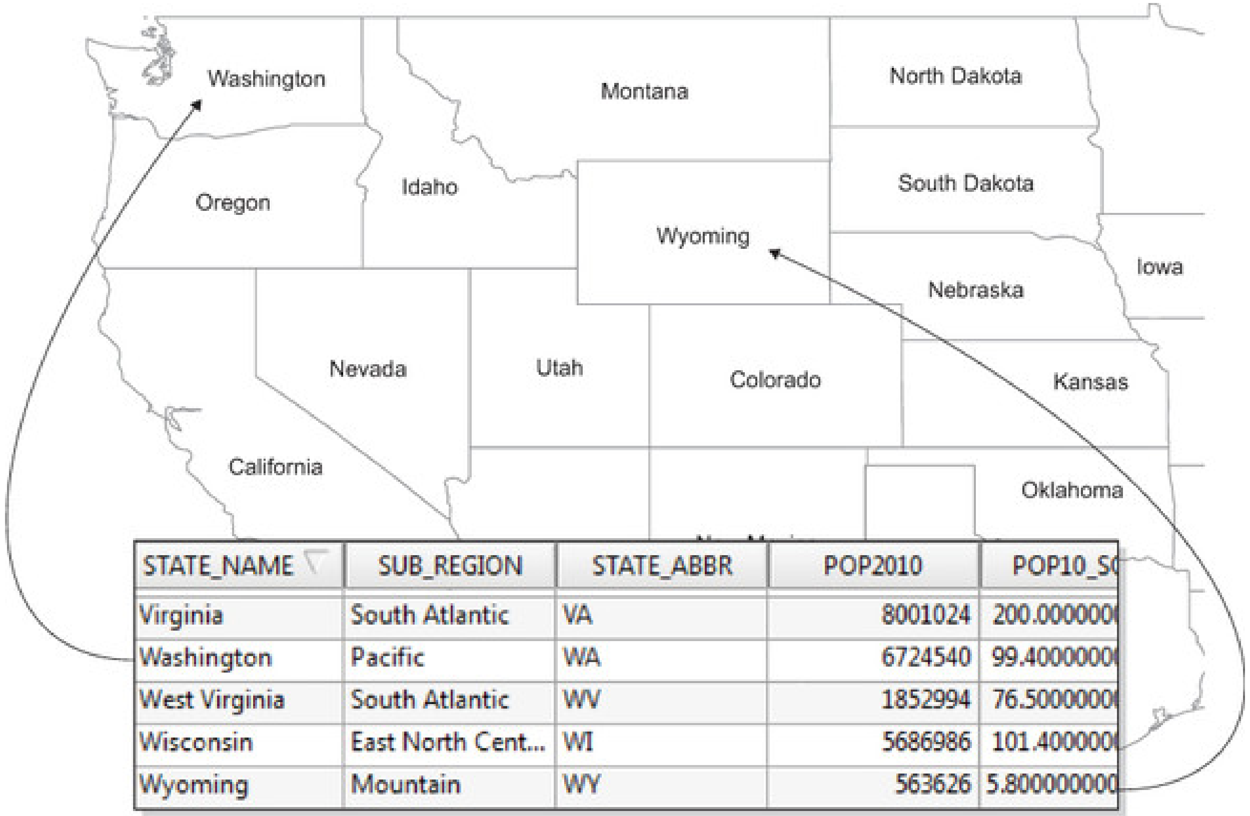
Multipart geometries (2D)

collections

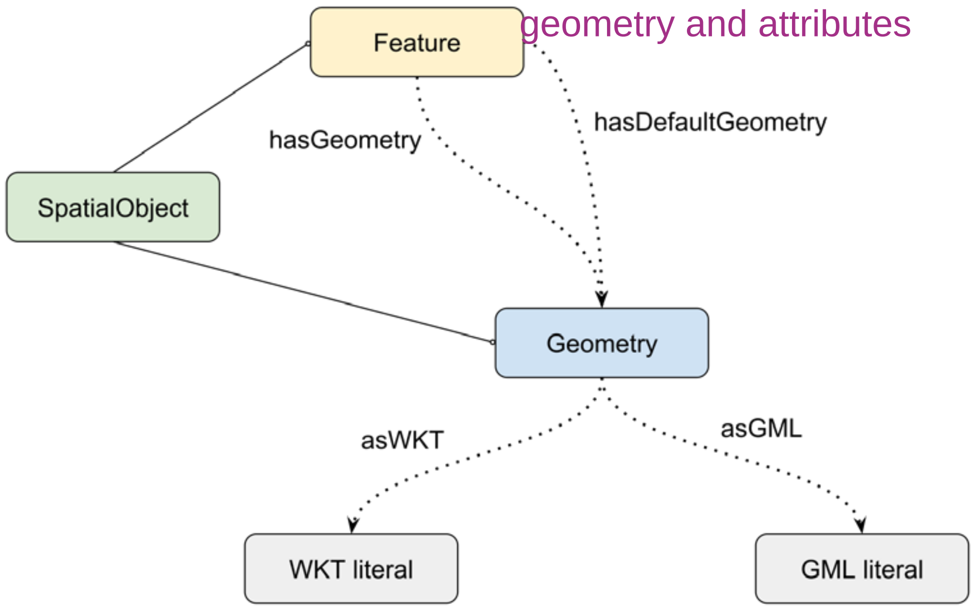
Type	Examples	
MultiPoint	MULTIPOINT ((10 40), (40 30), (20 20), (30 10))	
	MULTIPOINT (10 40, 40 30, 20 20, 30 10)	
MultiLineString	MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))	
MultiPolygon	MULTIPOLYGON (((30 20, 10 40, 45 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5)))	
	MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 45 20, 30 5, 10 10, 10 30, 20 35), (30 20, 20 25, 20 15, 30 20)))	

Vector data model (Geometry + attributes)

- The association of attributes and spatial geometries makes geodata special. Captured by OGC’s notion of a *feature (here: GeoSPARQL)*



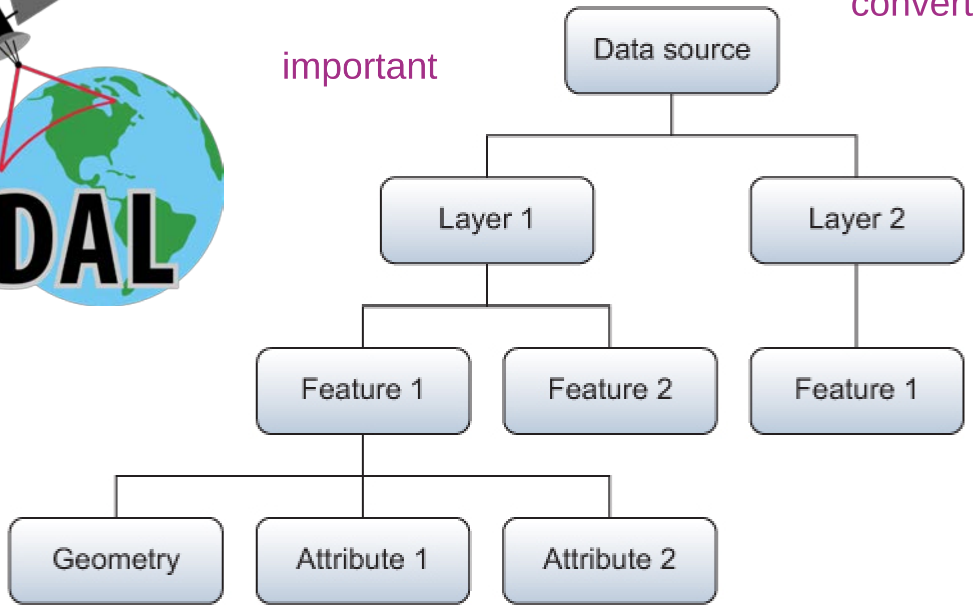
STATE_NAME	SUB_REGION	STATE_ABBR	POP2010	POP10_S
Virginia	South Atlantic	VA	8001024	200.000000
Washington	Pacific	WA	6724540	99.4000000
West Virginia	South Atlantic	WV	1852994	76.5000000
Wisconsin	East North Cent...	WI	5686986	101.400000
Wyoming	Mountain	WY	563626	5.800000000



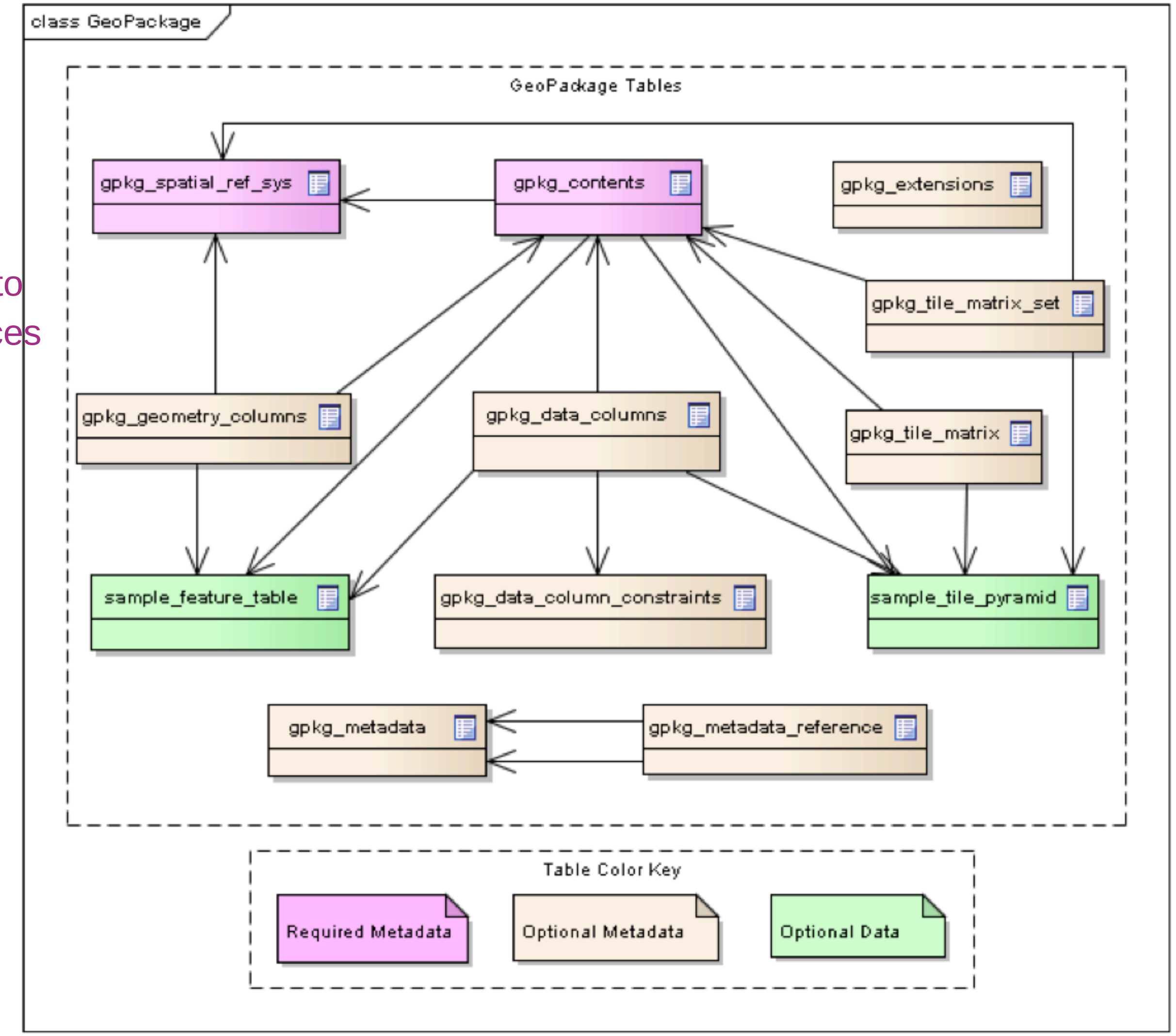
Vector data model (Geometry + attributes)

- GDAL/OGR is a universal library for dealing with vector (and raster) formats https://gdal.org/ lab 2 makes it possible to convert data sources

important



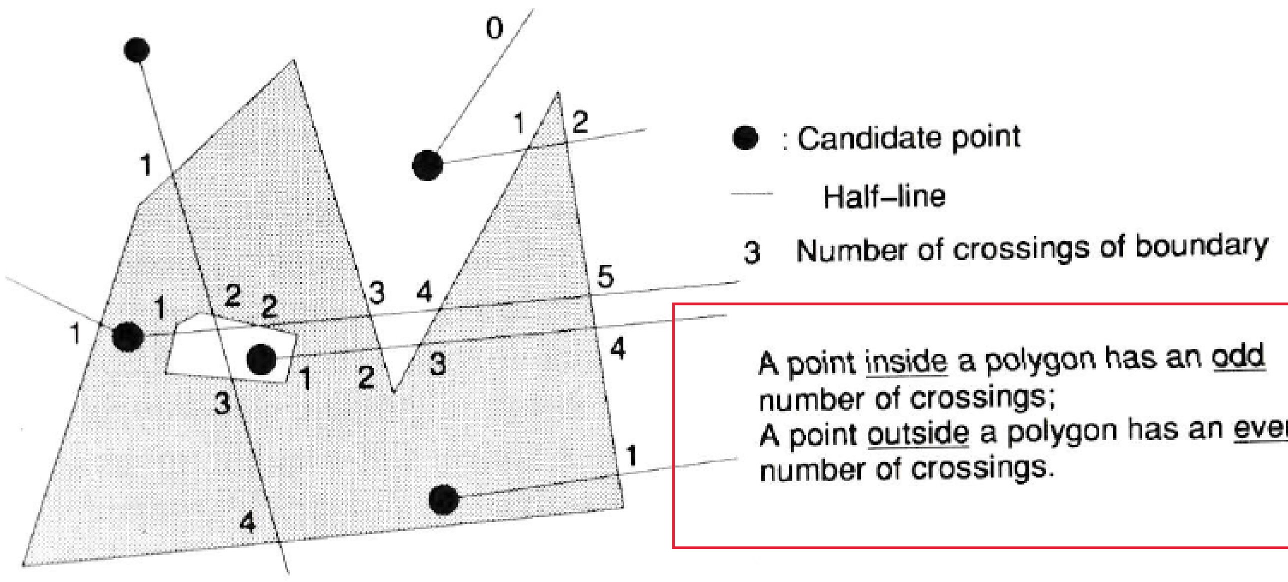
OGR class structure



Basic geometric operations: Point in Polygon and Centroid

- To find out whether a point is located inside a polygon
- To find out the center of gravity

combining geometries
by constructing a ray and then
counting the number of
intersections



A point inside a polygon has an odd number of crossings;
A point outside a polygon has an even number of crossings.

Figure 7.7 Illustration of the half-line theorem (point-in-polygon rule).

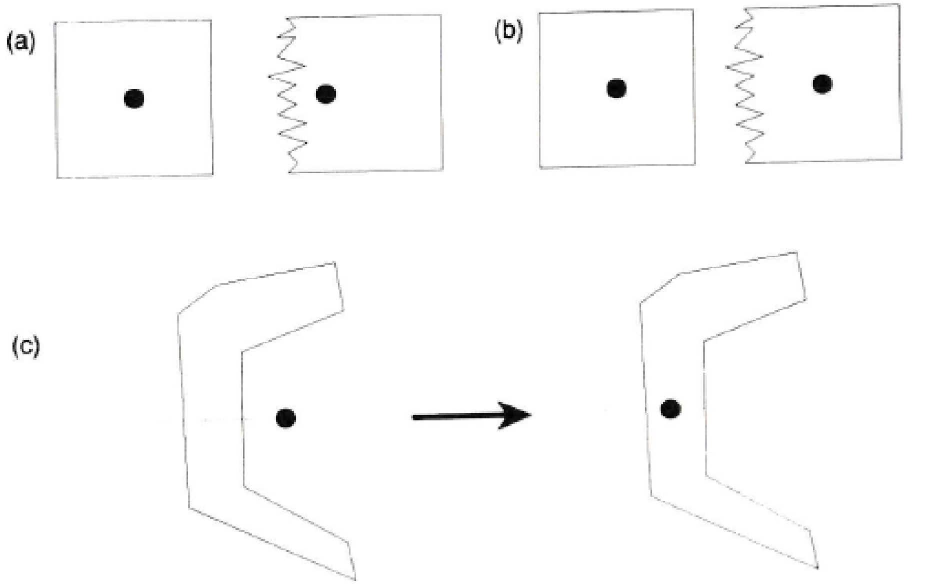


Figure 7.8 Definition of centroids. (a) Examples of centroids. (b) Examples of centroids defined as centres of gravity. (c) Moving an outside centroid.

Laurini/Thompson 1996

Basic geometric operations: Geometric intersections

- Compute the difference, union and intersection of polygons

divide the polygons into slabs

split polygons into slabs according to point out of which they exist, then count them and subtract the ones where they intersect

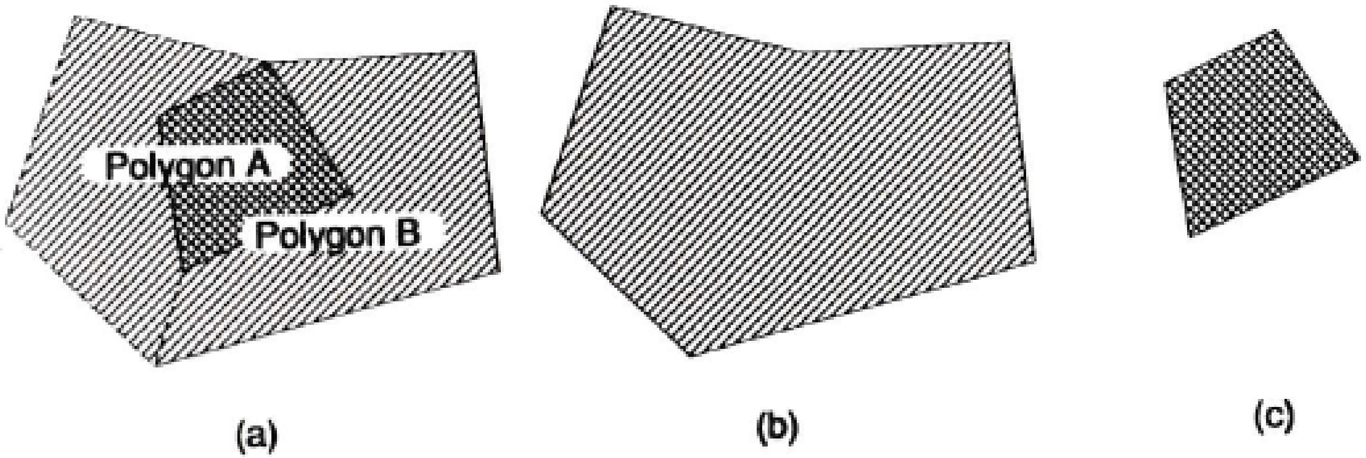


Figure 7.10 Union and intersection of polygons. (a) Two intersecting polygons. (b) Union of A and B. (c) Intersection of A and B.

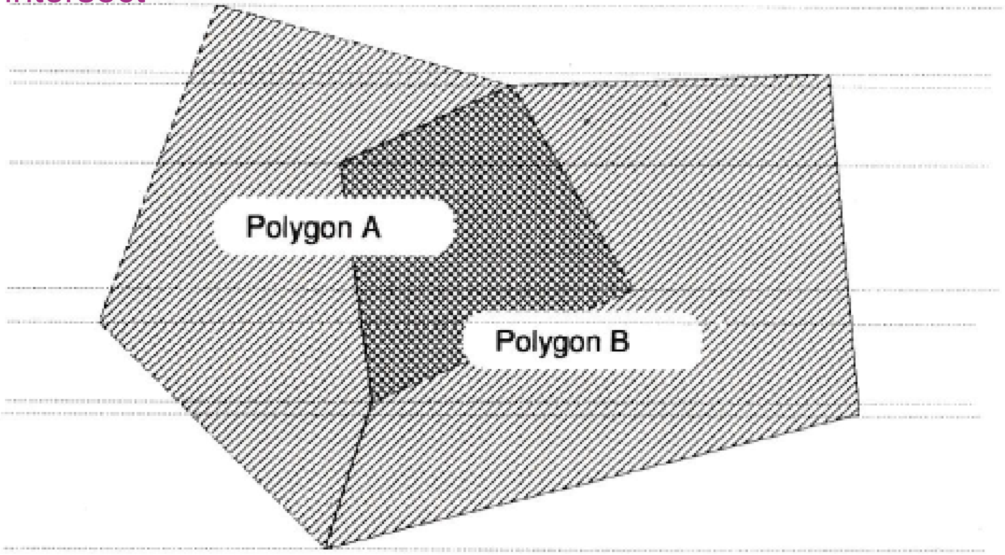


Figure 7.11 Splitting two polygons into slabs.

Laurini/Thompson 1996

Core concept data types

Core concepts of spatial information (Kuhn 2012)

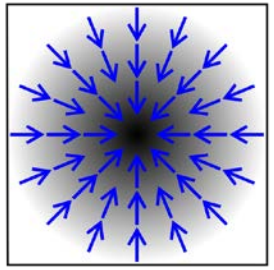
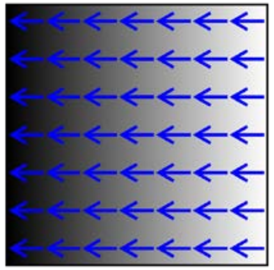
Cognitive lenses through which the environment can be studied
(… like “cell” in biology or “value” in economics)

Provides constraints for…

1. Analysing geodatasets with tools
2. Posing analytical questions
3. Finding answers with GIS workflows

Core concepts (= what geodata represents)

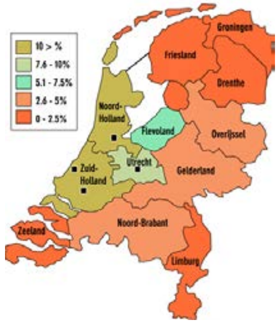
Field

- continuous phenomenon
- Space -> Quality (value field)
- boundaries are irrelevant

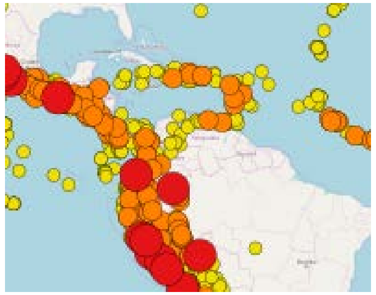
like temperature
discrete phenomena
functions of objeect identifier and their qualities

Object



- discrete phenomenon with qualities
- Object -> Space (projected in space, not time)
- boundaries are relevant
- Object -> Quality

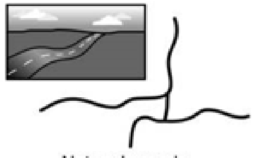
Event




- like objects, but...
- Event -> Time, Space (projected in time and space)

events are bound to a time (time component)

Network

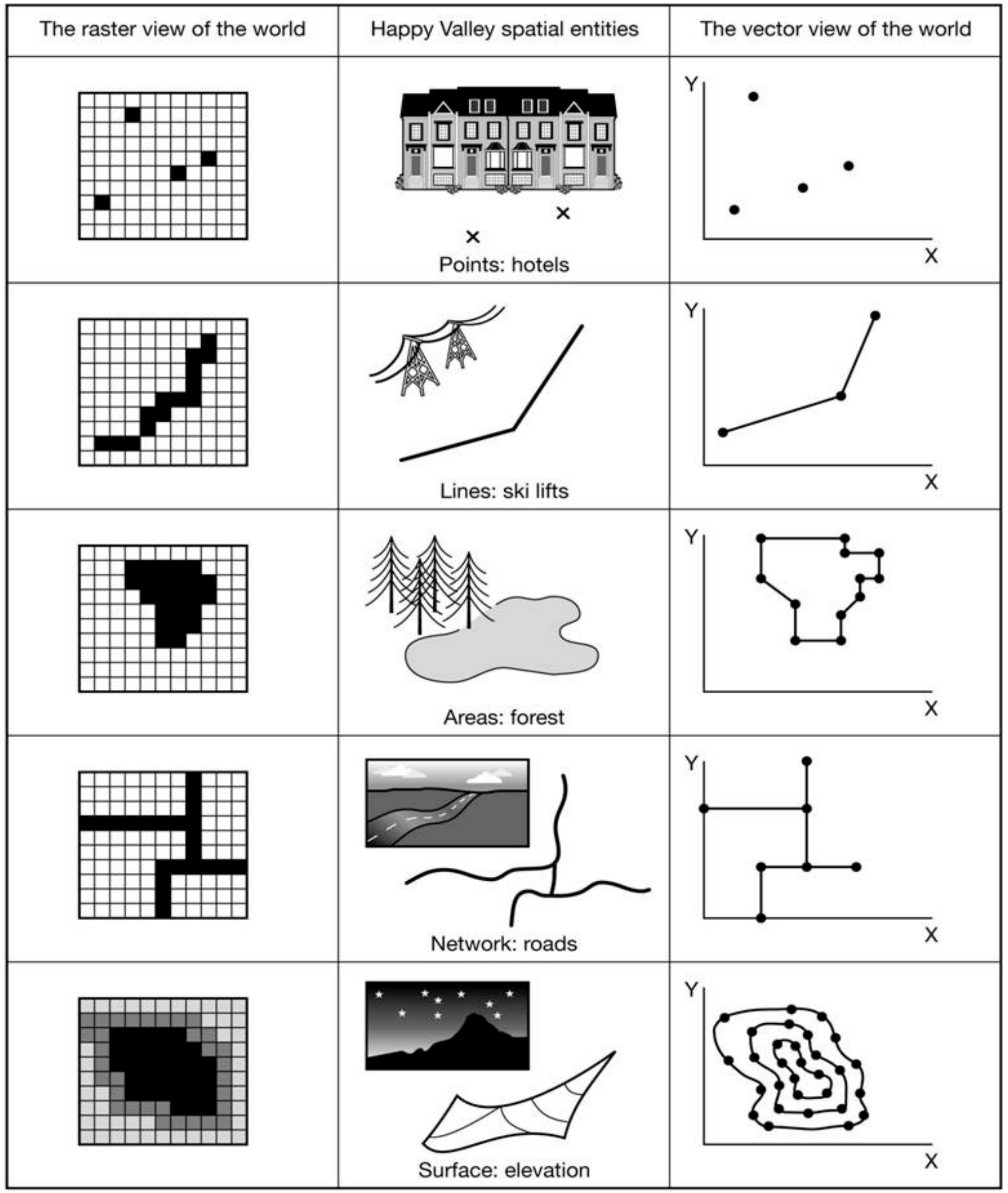


- Quantified relations between objects
- Object x Object -> Quality

capture the relations between data (street flow, water) each concept can be represented in raster and vector format

Raster/vector data model (= how geodata represents)

- Concept != data model
- Raster – Vector data model
- Can both represent fields, and objects, and networks ...
- So why do we use one over the other?
- Concept is "added" by human interpreter of data
- Concept is needed to understand analysis and possible transformations



How geodata models represent core concepts

once dimension is geometric layer type
one is the core concepts

Vector

Vector Tessellations

Raster

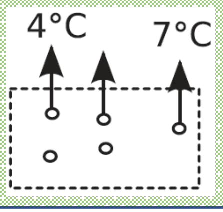
polygons

Field

concept

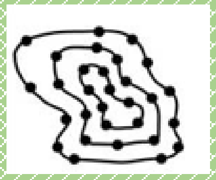
can be represented by different things

Point measures

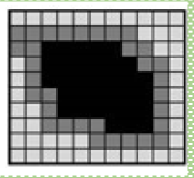


Coverages

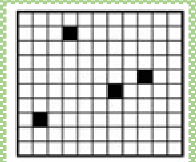
Contours



Field Raster



Existence Raster

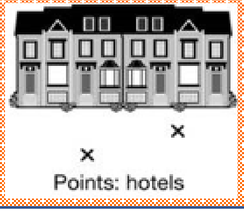


not an object more like a patch (no discrete object)
homogen. values

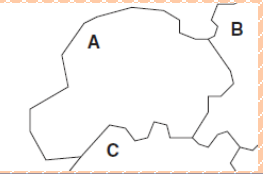
Object

object vectors like
admistrative regions

Object Vectors



Lattice
(Statistical regions)



lattice is used to denote
a tessalation that represents objects
so not just about geometric property but also that
each of the geo. properties represents an object that has
a meaning (like a municipality or country)

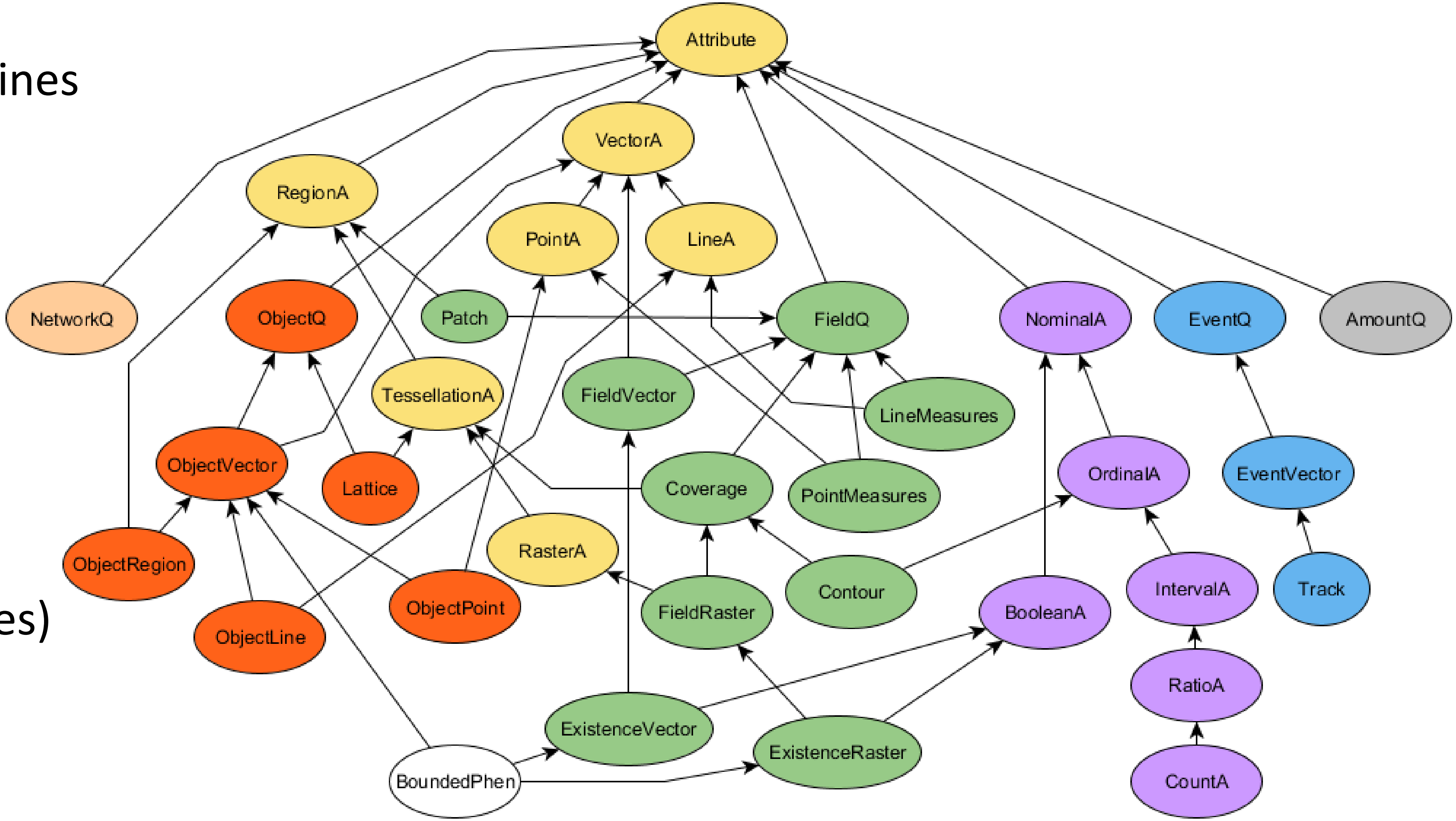
combines the layer types

Core concept data types (CCD) ontology

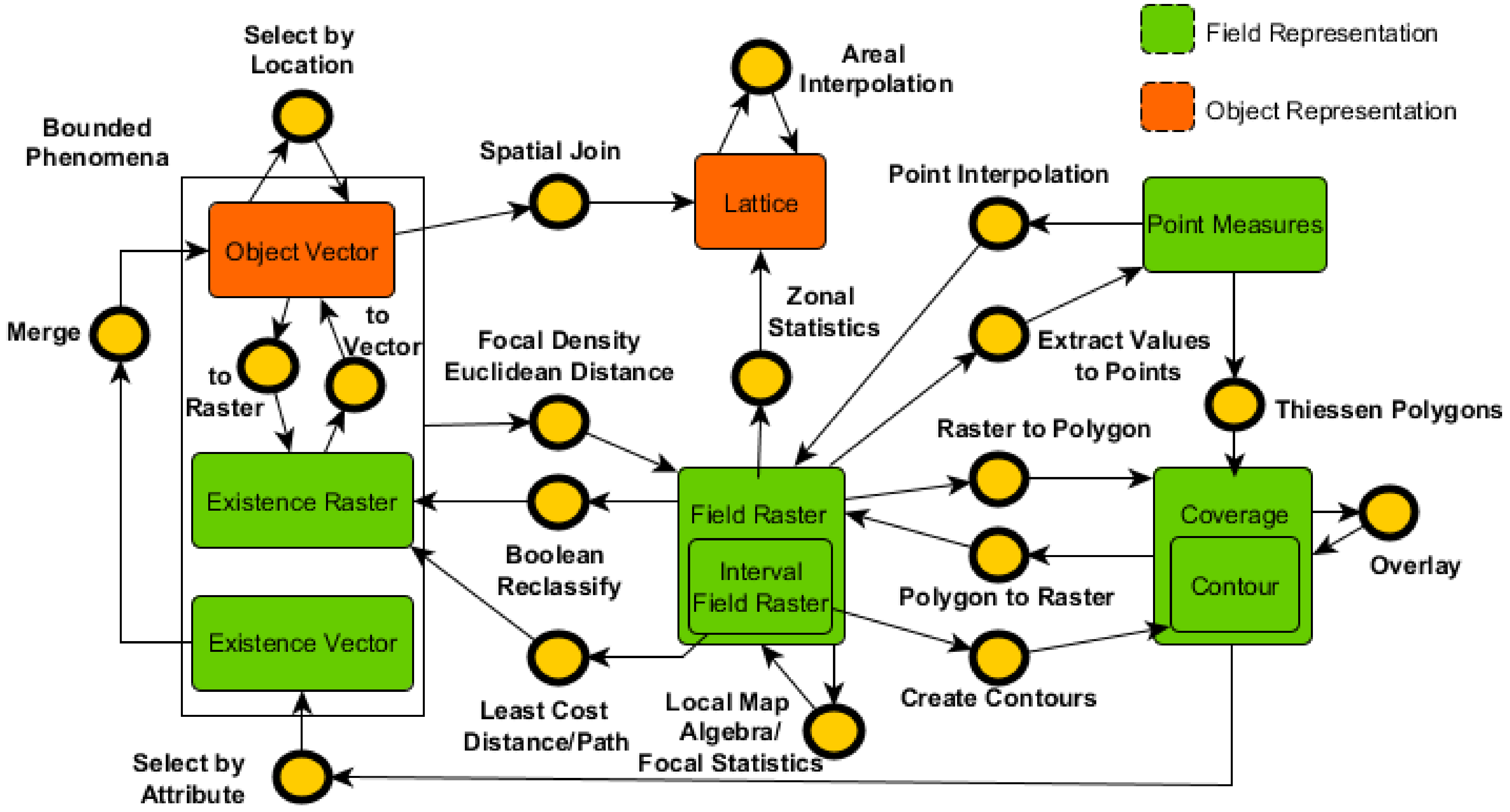
The ontology combines three dimensions:

1. Layer types
2. Core concepts field, object, event, network
3. Levels of measurement

(see lab for examples)



In which ways can geodata be transformed?



CCD Ontology

datasets can be described with these concepts

The ontology
can be used to
annotate geodata
resources.
(Scheider et al. 2020)

Examples
from the Amsterdam
data portal

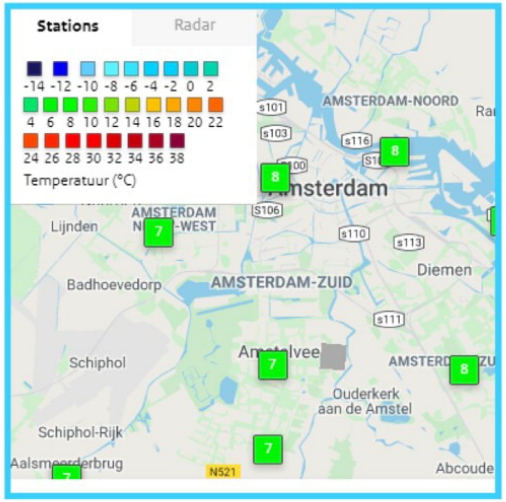
https://maps.amsterdam.nl/open_geodata/

ObjectPoint

layer showing buildings as objects

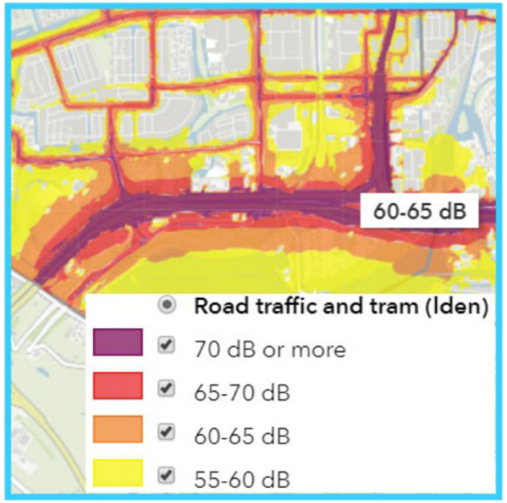
EventPoint

WW2 attacks



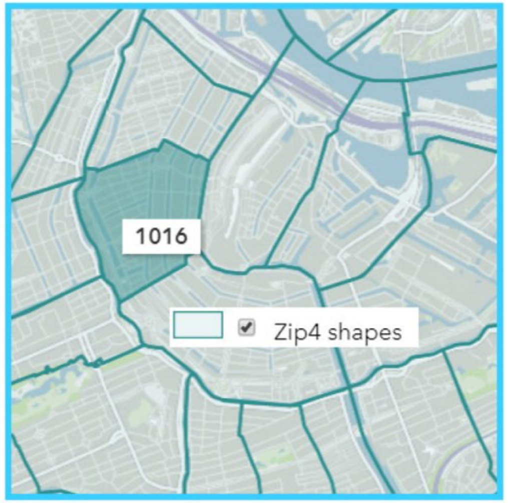
PointMeasures

Coverage



Contour

continuous noise field



Lattice

boundaries of objects

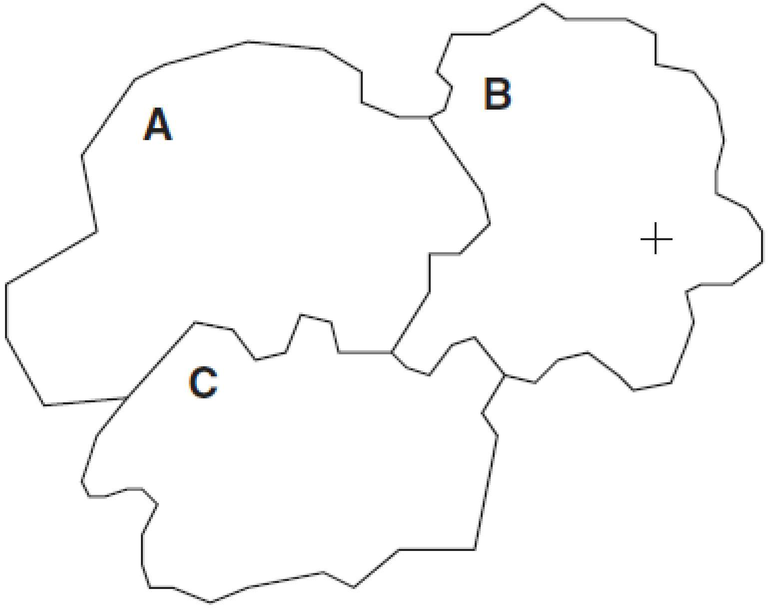
The difference between Coverages and Lattices

both are

Tessellations are regions that fully cover space without overlap. For three polygons A, B, and C, land cover type and average elevation values are given in the table. At an arbitrary location inside a polygon, for example, the location marked with a +, a *coverage* yields the quality for that location, such as its land cover type. For a *lattice*, such as average elevation, the elevation value at this particular location (+) is not available.

Coverage: self-similar (the attribute value of the whole applies also to parts)

Lattice: not self-similar



attribute value of the whole applies to each point within the layer

not self similar every point inside the polygon has it's own value

Coverage

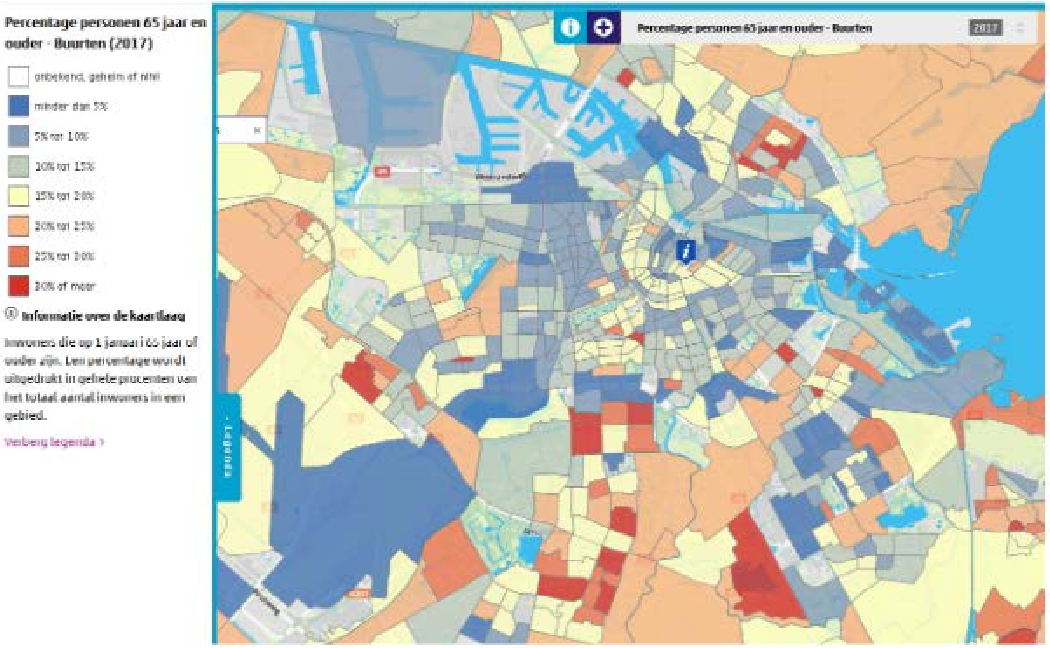
Lattice

	land cover type	average elevation (m)
A	Forest	631
B	Urban	220
C	Water	42
+	Urban	

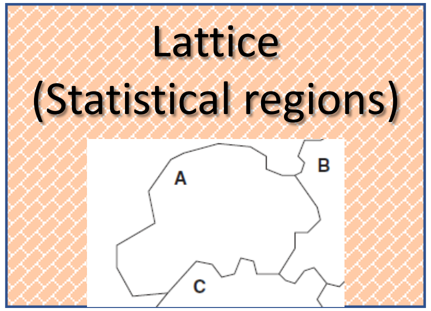
Which overlay method could be used for assessing the amount of elderly people living in PC4 areas?

Answer:

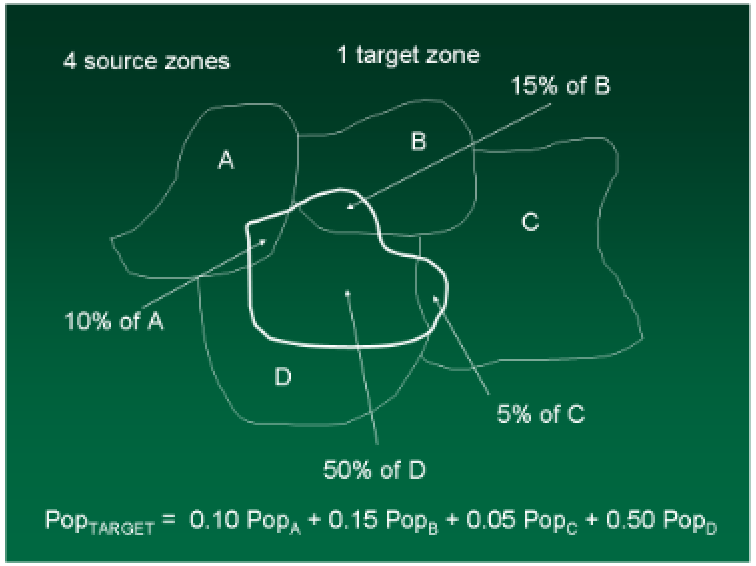
need to make use of areal interpolation



(a) CBS Buurt statistics, showing the percentage of persons over 65 in neighborhoods.



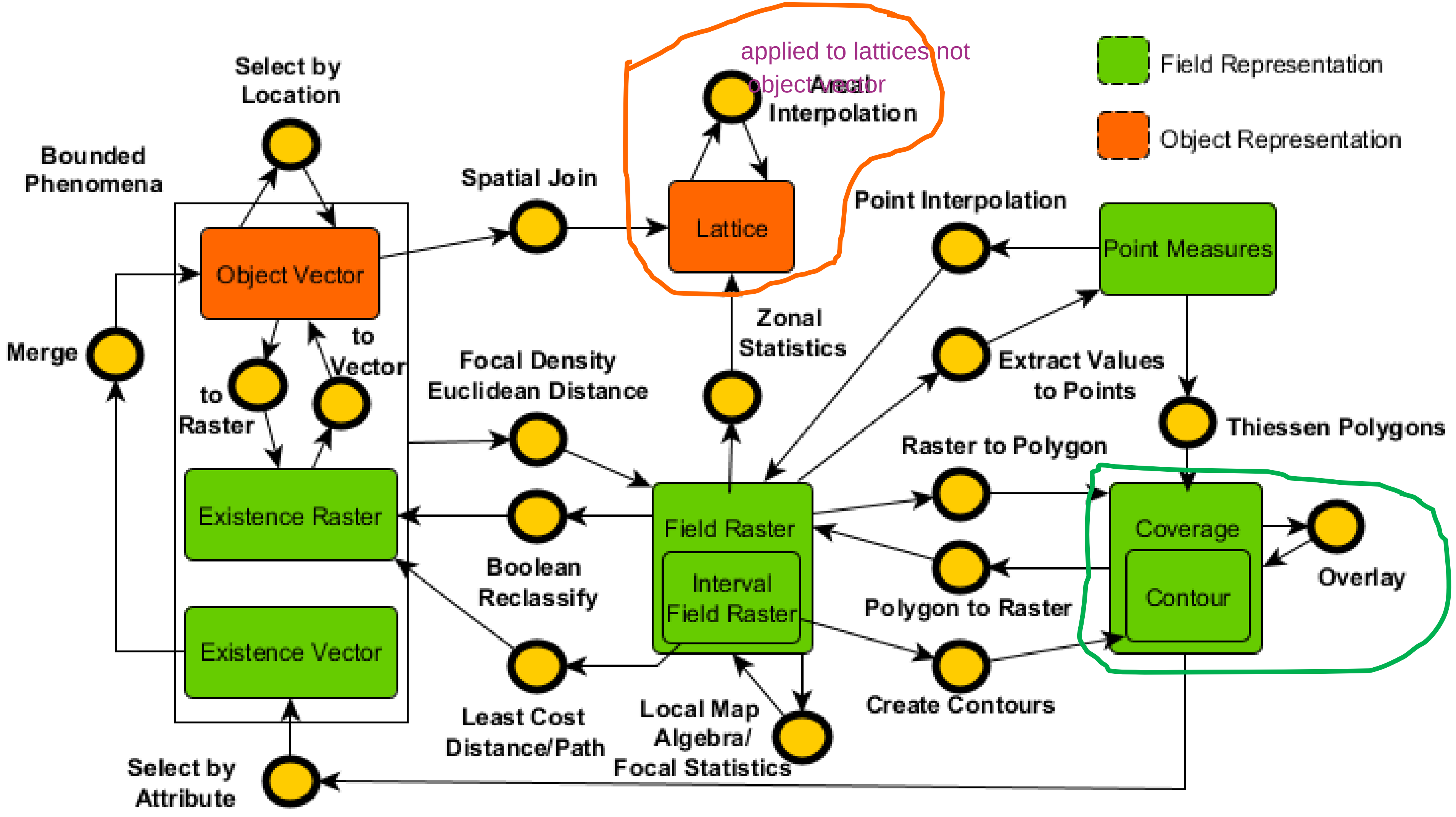
Lattices are not self-similar, thus overlay needs to account for this



(b) Simple areal interpolation method. Image by kind permission of Michael Goodchild.

cbs lattice

In which ways can geodata be transformed?



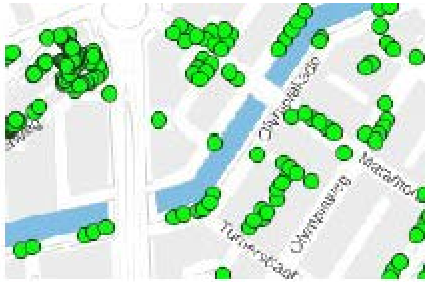
The difference between ObjectPoints, PointMeasures, LineMeasures and Contours

ObjectPoints are point representations of objects

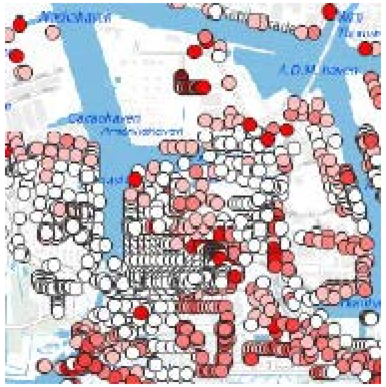
PointMeasures are pointwise measurements of fields

LineMeasures are linewise measurements of fields

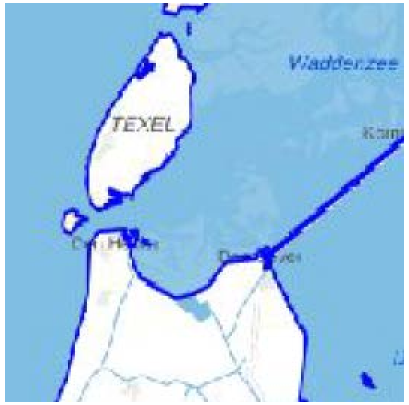
Contours are tessellated regions of field intervals



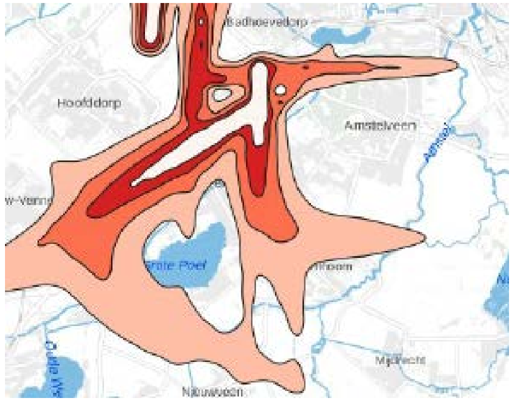
Trees: ObjectPoint



Temperature: PointMeasures

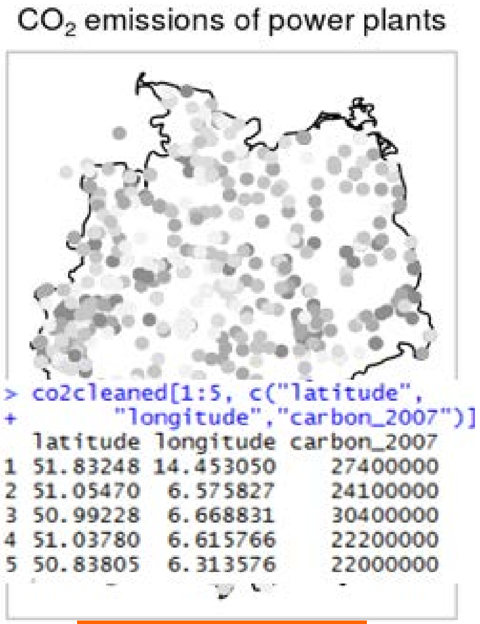


Coast: LineMeasures



Noise: Contours

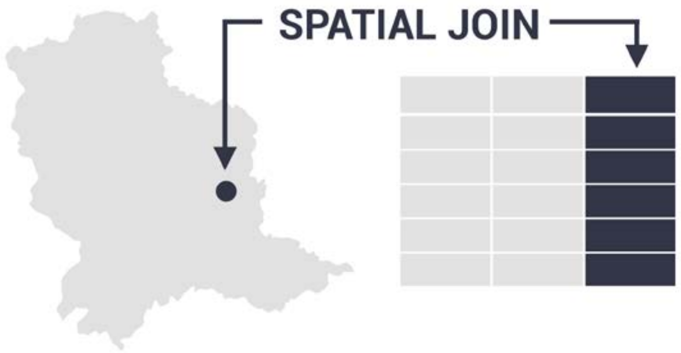
How much polluted is Germany?

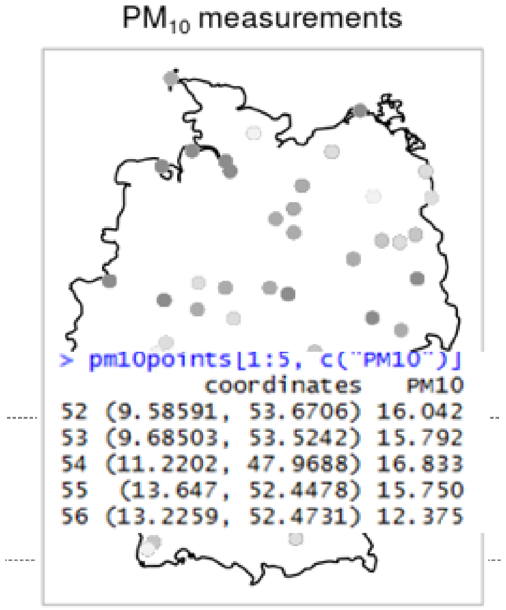


can use spatial join

object measures
ß cannot be interpelated

ObjectPoints

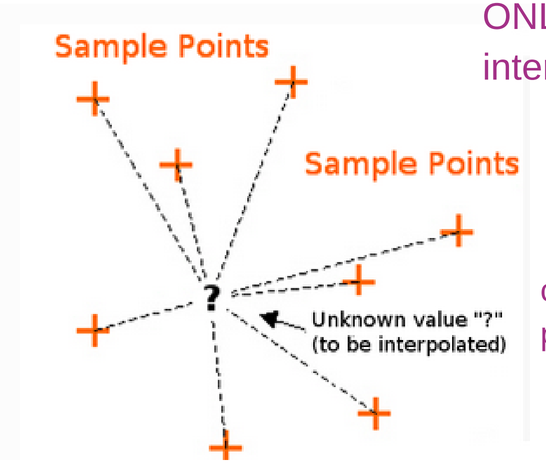




cannot use spatial join

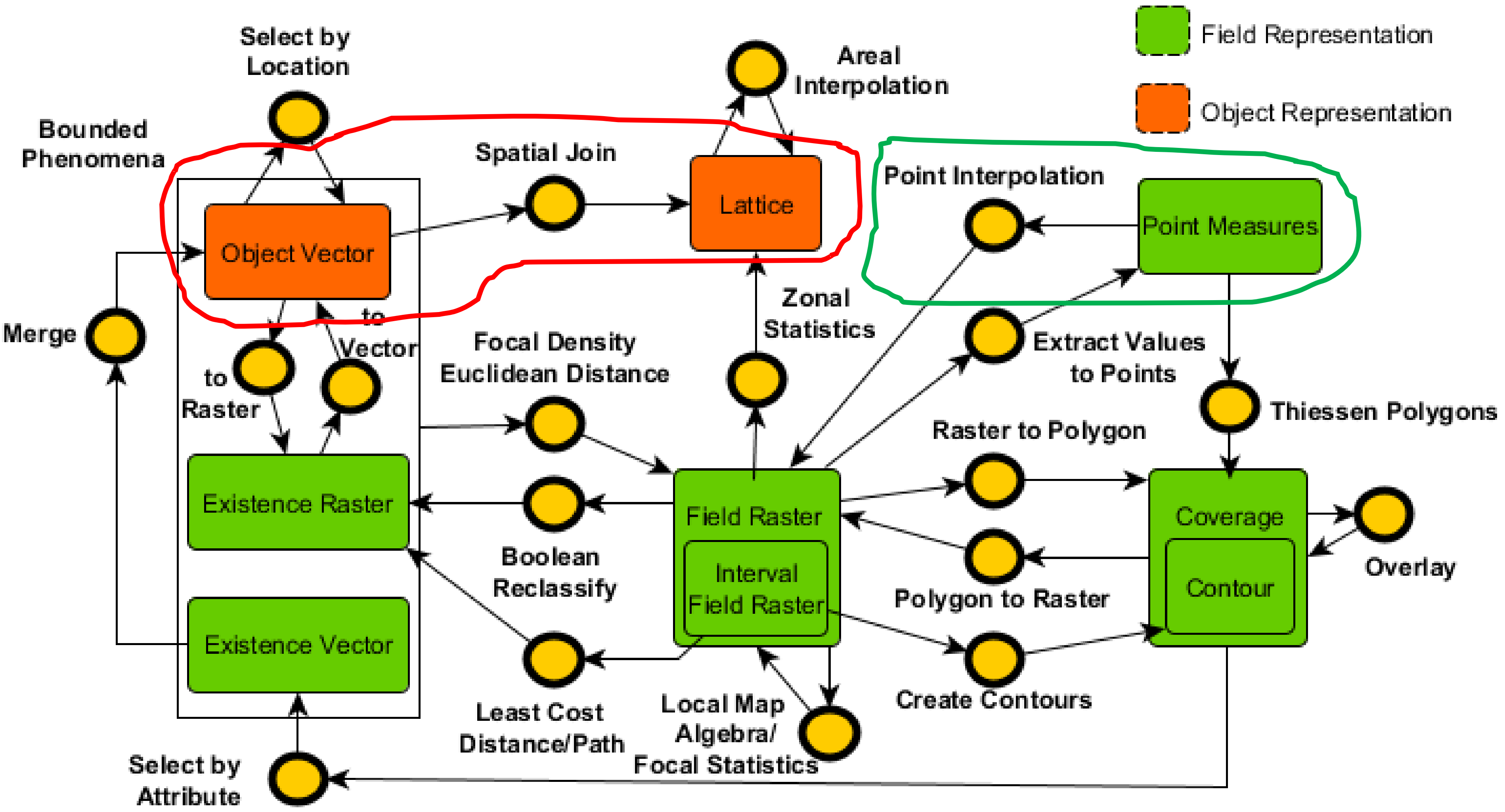
PointMeasures

point measures can ONLY do point interpolation



continuous field with point mesaure

In which ways can geodata be transformed?



Questions?
(Q&A session)

References

- Mandatory readings:
  - Chapter 2 “Measurement frameworks” in Chrisman 2002: Exploring geographic information systems, 2nd edition)
  - Chapter 7: Manipulations (interpolations, geometric operations, transformations) in Laurini and Thompson 1992: Fundamentals of Spatial Information Systems
- https://www.ogc.org/standards/sfa
- Kuhn, W. (2012). Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12), 2267-2276.
- Scheider, S., Meerlo, R., Kasalica, V., & Lamprecht, A. L. (2020). Ontology of core concept data types for answering geo-analytical questions. Journal of Spatial Information Science, 2020(20), 167-201 (http://www.josis.org/index.php/josis/article/viewArticle/555)