
Geospatial Visual Analytics

Data Transformations



Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

Gennady Andrienko & Natalia Andrienko

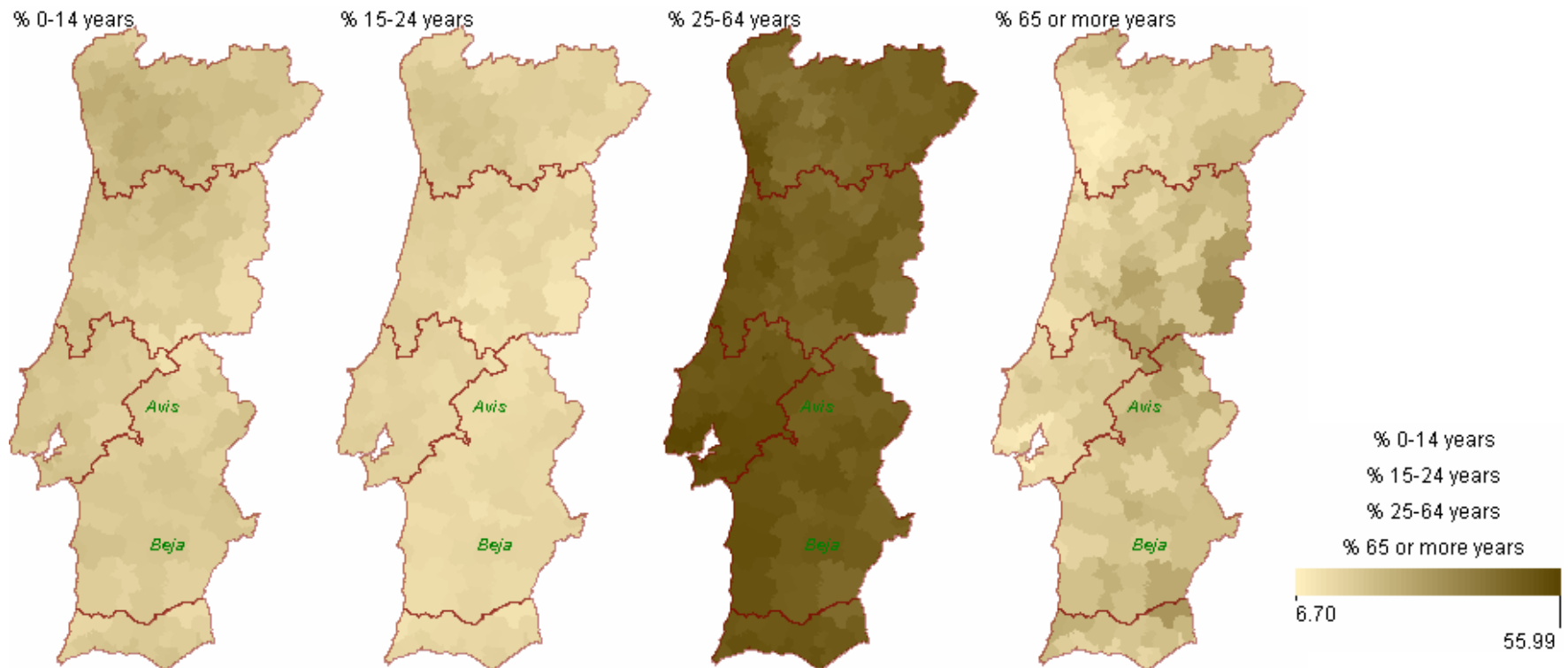
<http://geoanalytics.net>

Data transformations

- may help us to see what is originally not evident
- allow us to look at data from different perspectives and to investigate various aspects
- can reduce the amount of data we need to deal with
- can give us additional useful data
- may allow us to apply a richer set of analytical techniques

Transformation of attribute scales

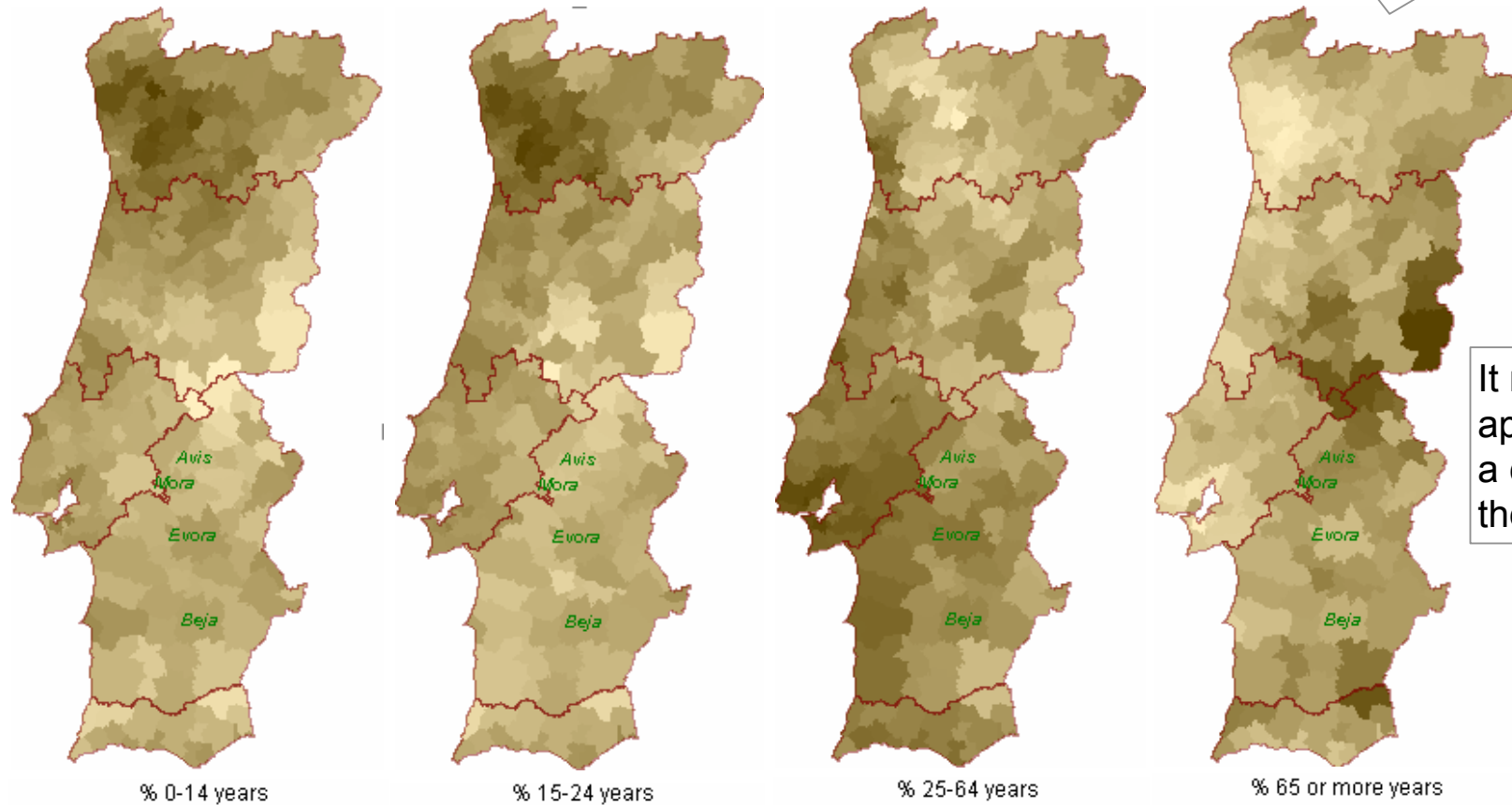
Population age structure in Portugal



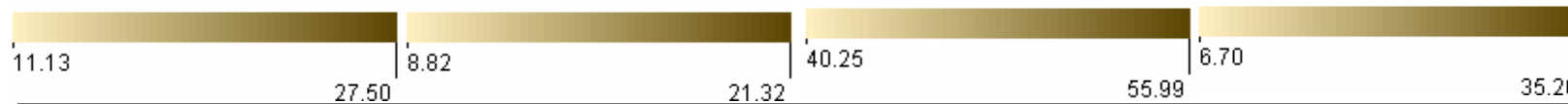
Population age structure in Portugal

An individual shading scale for each age group

The values in each district are shaded according to their relative positions within the value ranges of the attributes

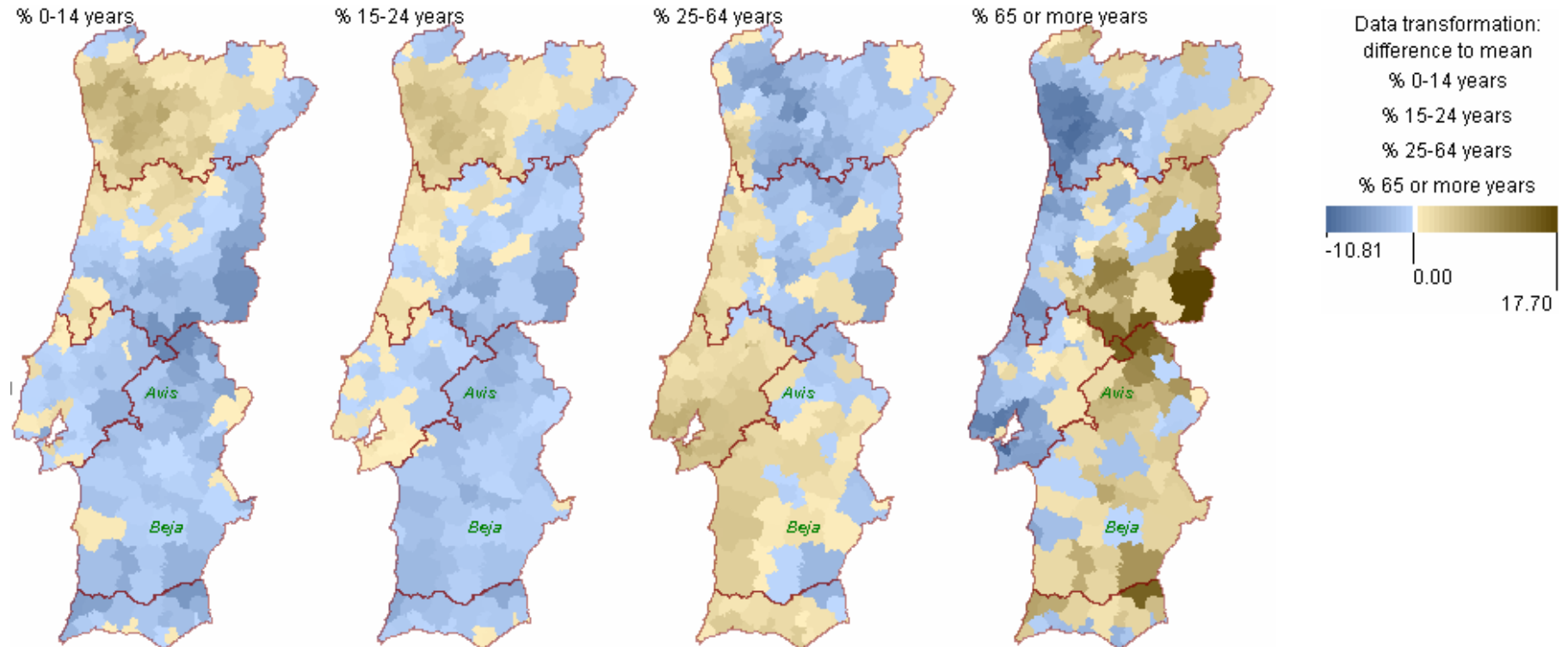


It may be appropriate to find a common basis for the comparison



Population age structure in Portugal

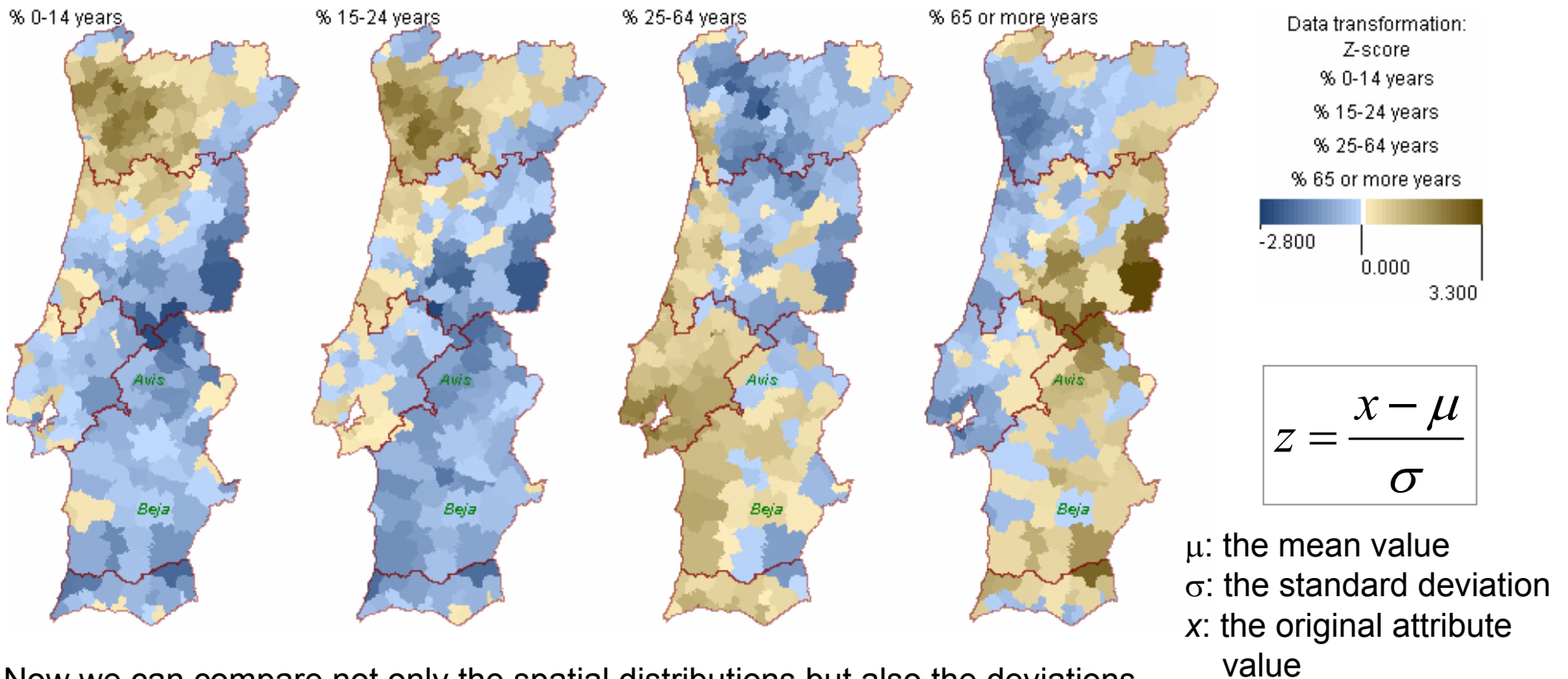
Comparison to country's mean values (absolute difference)



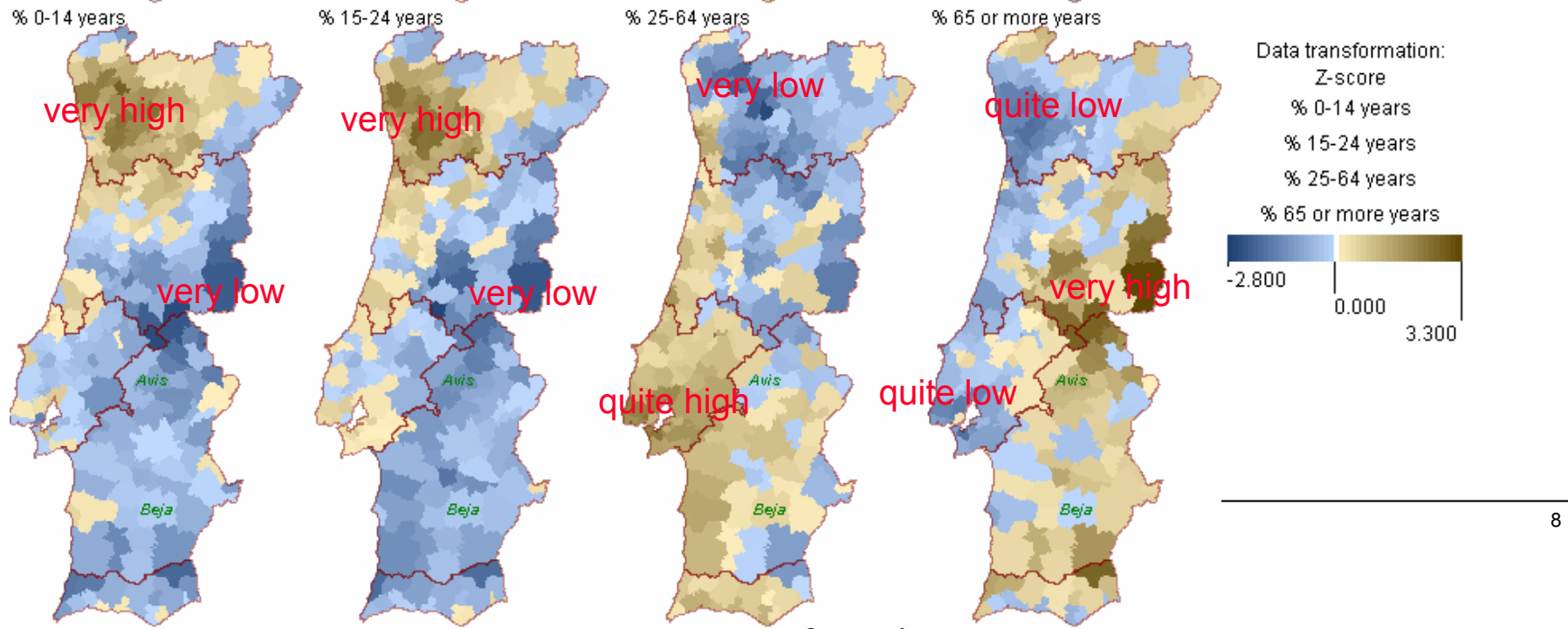
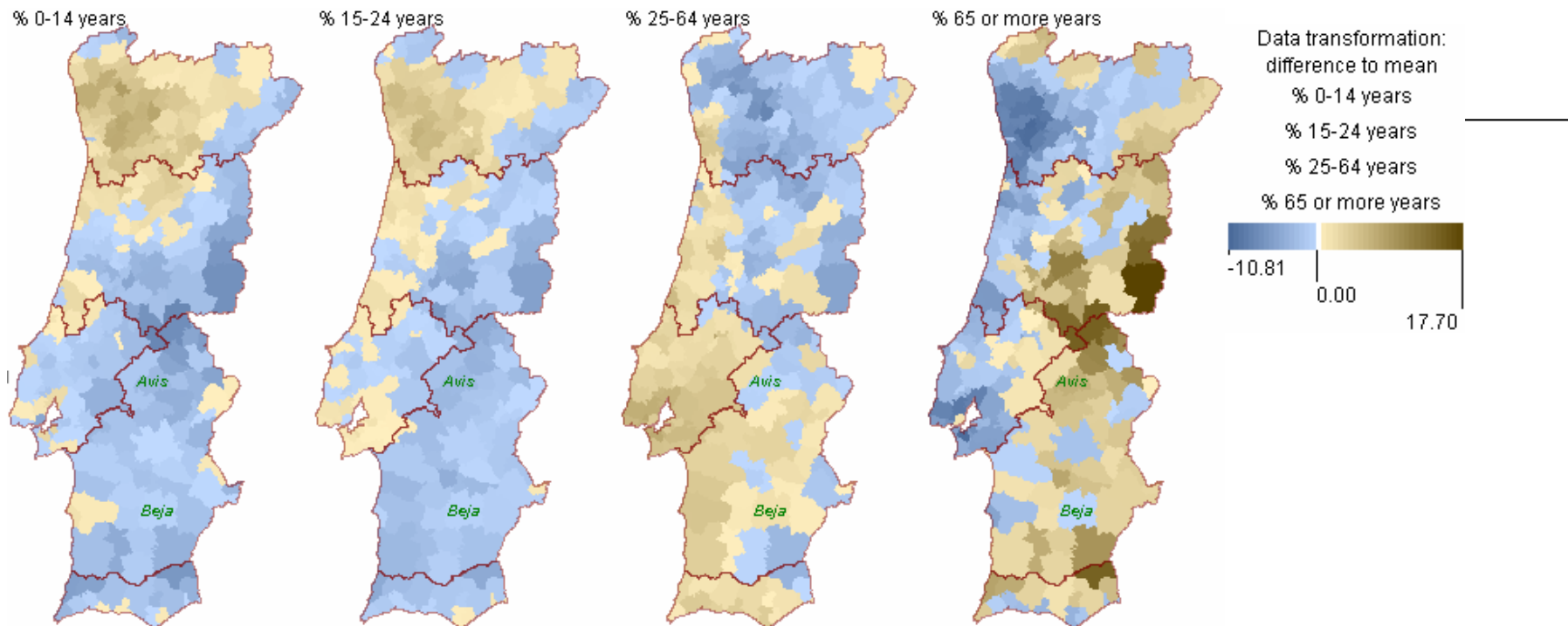
It is now easier to see both the similarities and differences in the spatial distributions. However, the absolute differences to the mean depend on the value ranges and may differ much between attributes.

Population age structure in Portugal

Z-scores (normalised differences from the country's mean values)

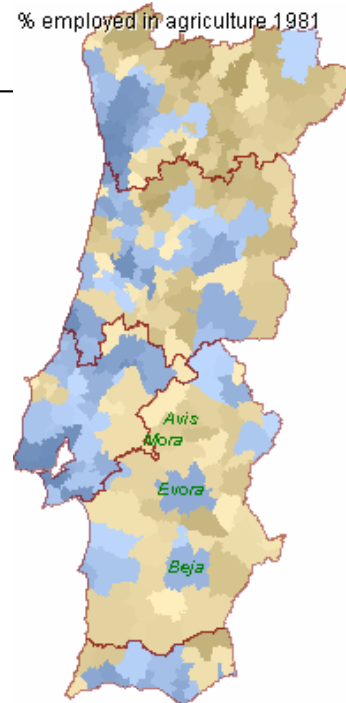


Now we can compare not only the spatial distributions but also the deviations from the mean values.

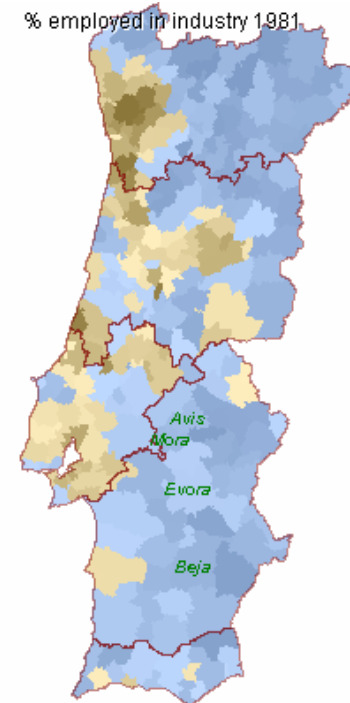


Here we use the transformation into z-scores (*standard normal transformation*) to explore the changes in the employment structure that occurred between the 1981 and 1991 censuses.

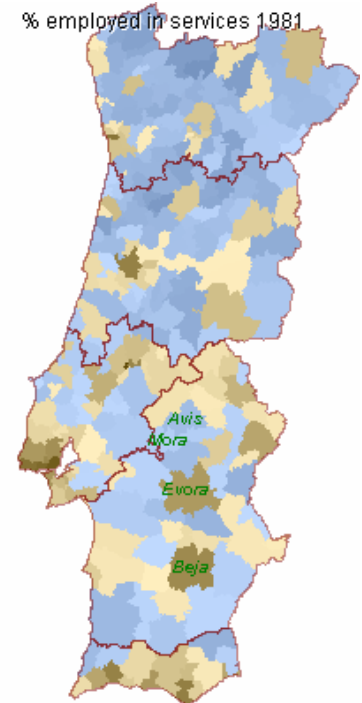
% employed in agriculture 1981



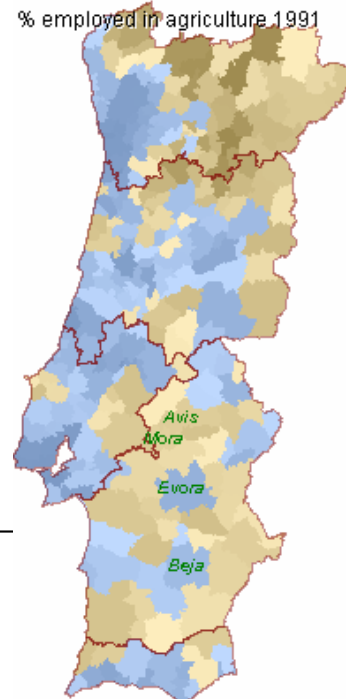
% employed in industry 1981



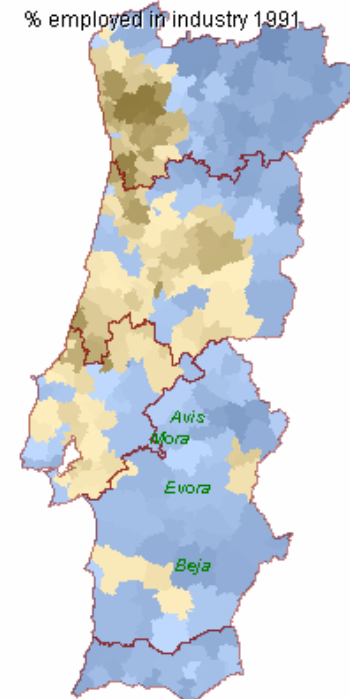
% employed in services 1981



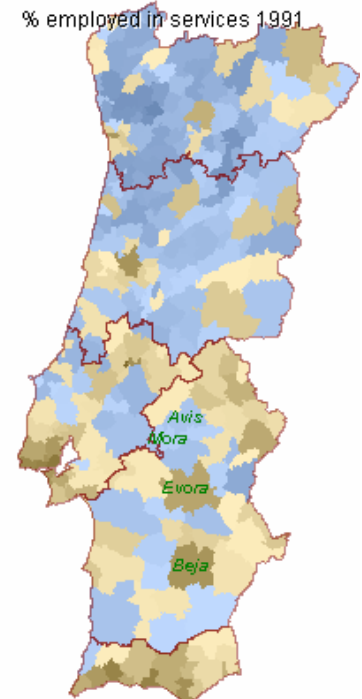
% employed in agriculture 1991



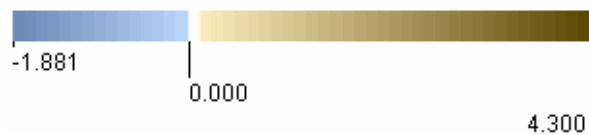
% employed in industry 1991



% employed in services 1991



Data transformation: Z-score
% employed in agriculture 1981
% employed in industry 1981
% employed in services 1981
% employed in agriculture 1991
% employed in industry 1991
% employed in services 1991

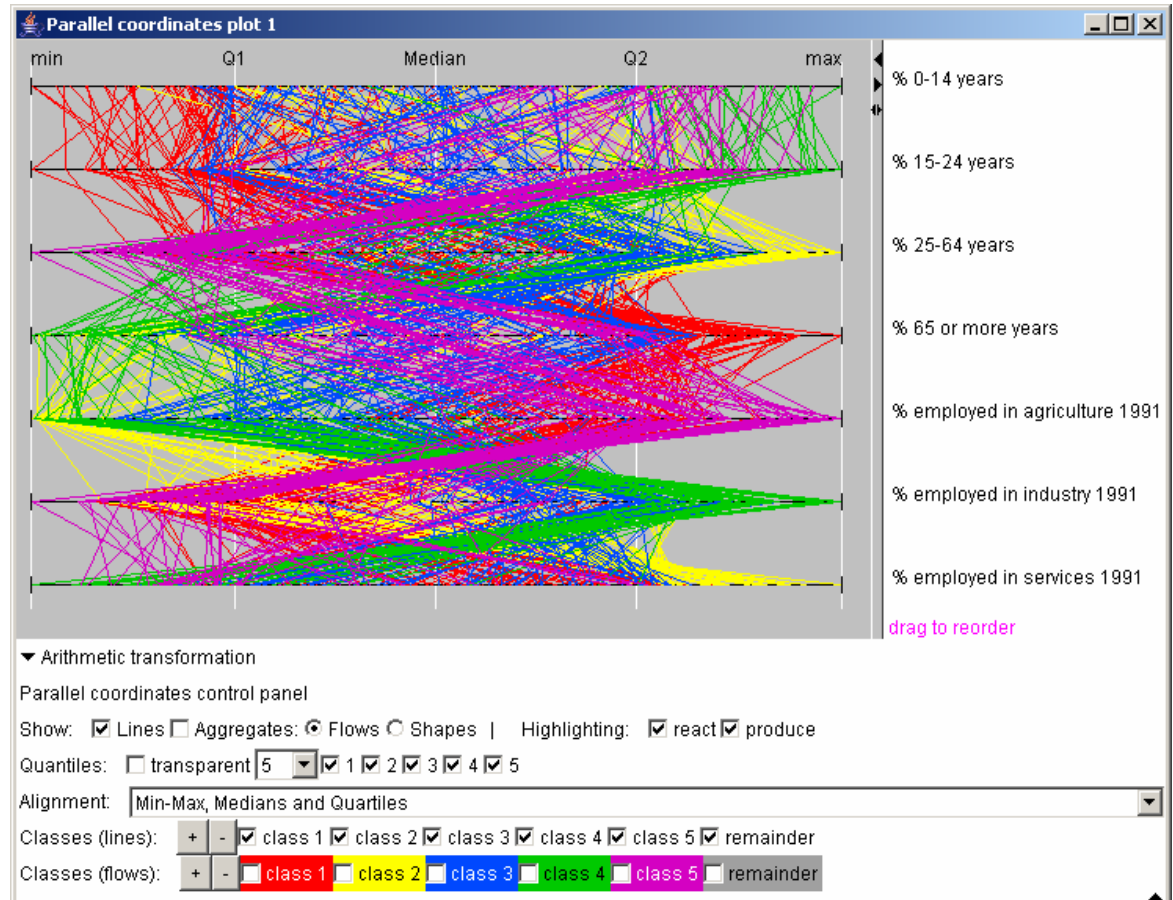
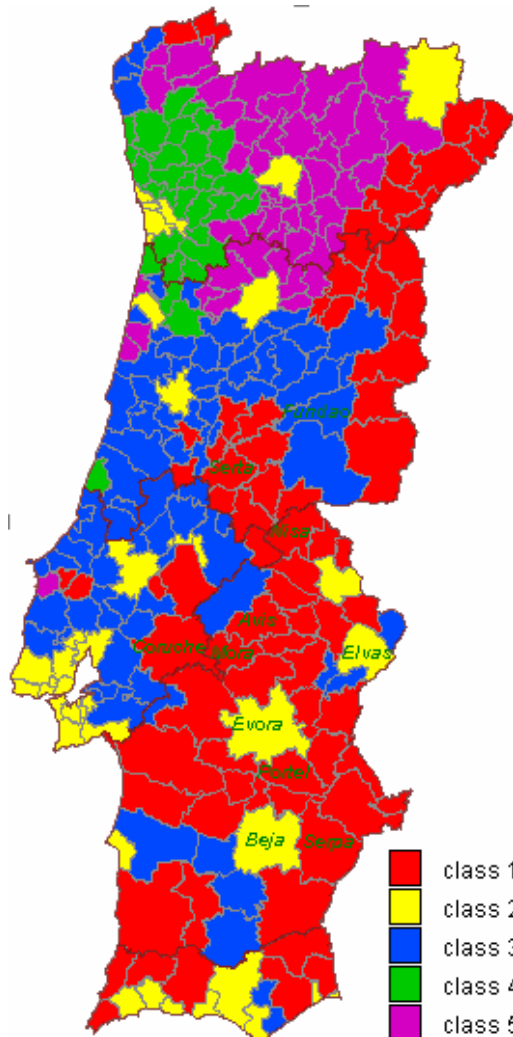


Bringing several attributes together by clustering

Goal of clustering: group objects by similarity of their characteristics expressed through multiple attributes.

We apply a clustering algorithm for grouping the districts according to the proportions of 4 age groups and proportions of working people employed in 3 areas of activities.

Divide Portugal into regions according to age and employment structure



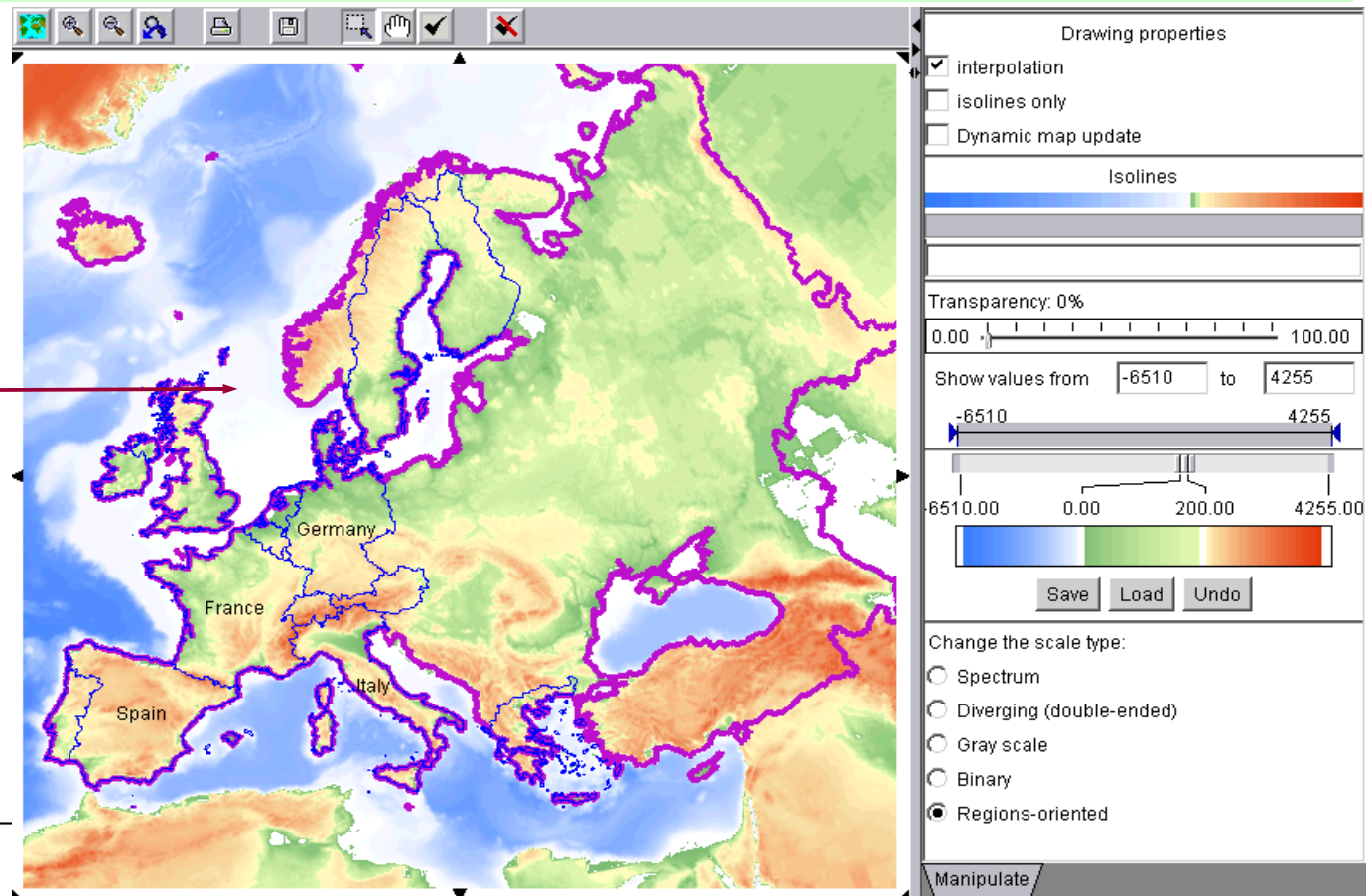
Tool: clustering
 ("simple k-means" method from Weka data mining toolkit)

Transformation of data formats for the exploration of raster data

Raster data

Typically, represent continuous phenomena. Values are measured at sample locations, usually in nodes of a regular rectangular grid.

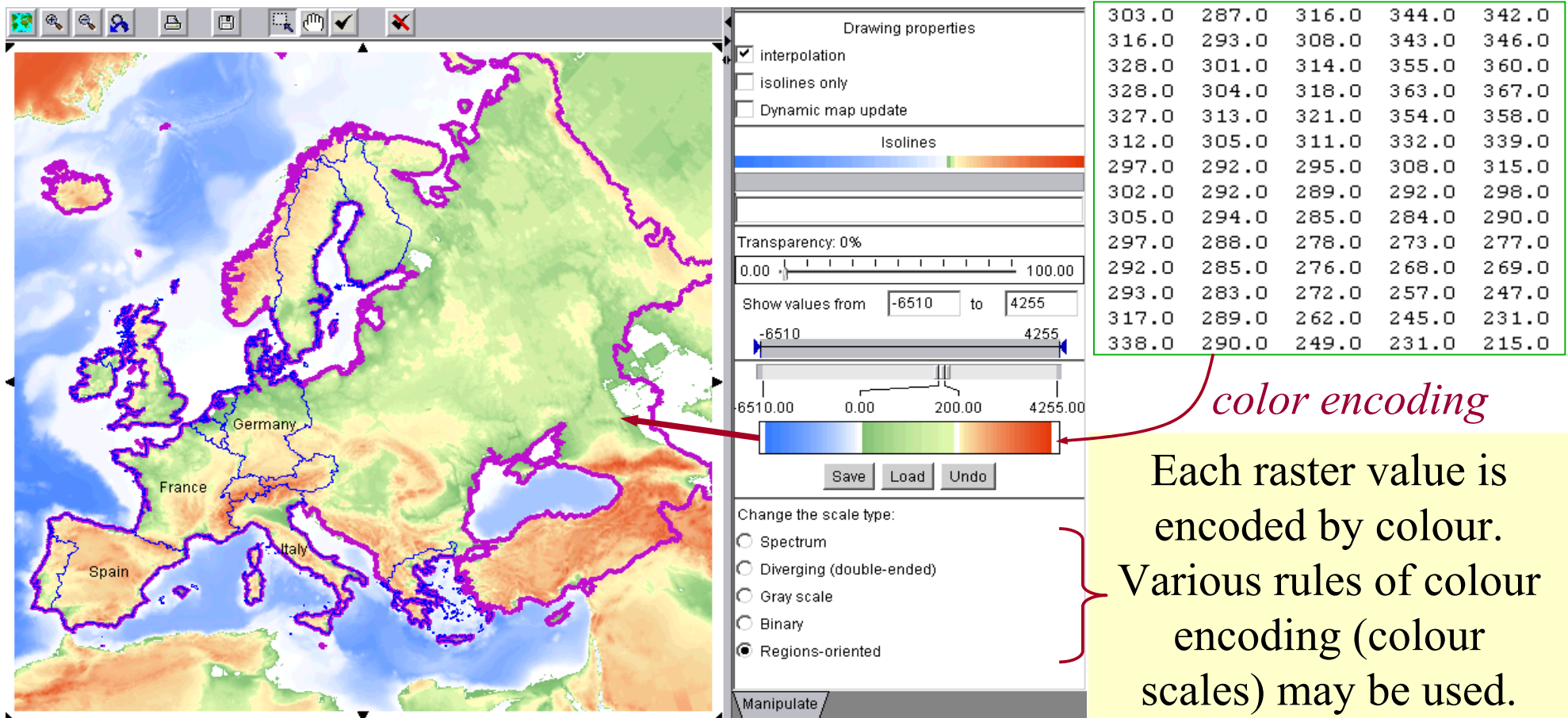
303.0	287.0	316.0	344.0	342.0
316.0	293.0	308.0	343.0	346.0
328.0	301.0	314.0	355.0	360.0
328.0	304.0	318.0	363.0	367.0
327.0	313.0	321.0	354.0	358.0
312.0	305.0	311.0	332.0	339.0
297.0	292.0	295.0	308.0	315.0
302.0	292.0	289.0	292.0	298.0
305.0	294.0	285.0	284.0	290.0
297.0	288.0	278.0	273.0	277.0
292.0	285.0	276.0	268.0	269.0
293.0	283.0	272.0	257.0	247.0
317.0	289.0	262.0	245.0	231.0
338.0	290.0	249.0	231.0	215.0



Raster Data Properties

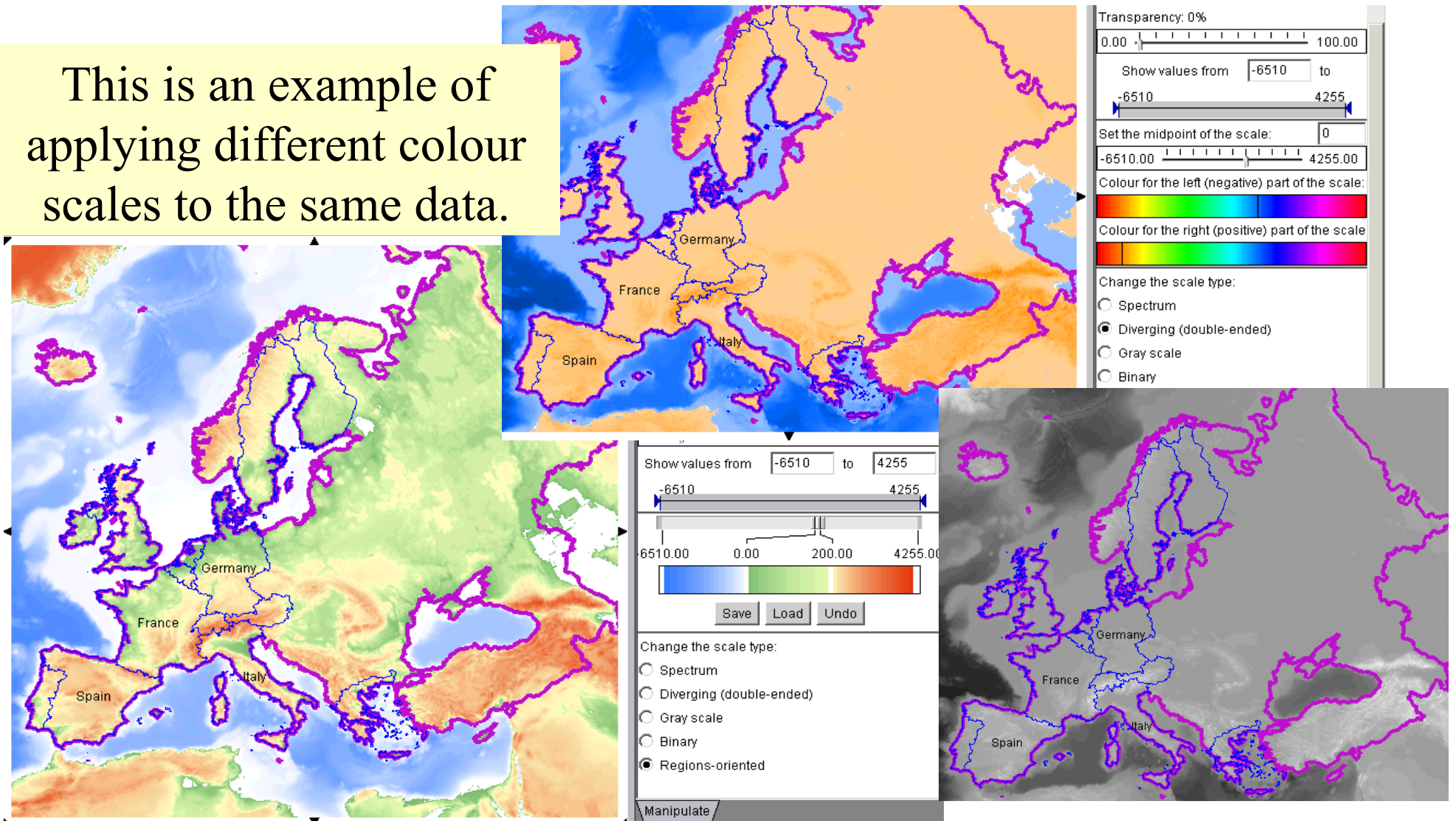
- One raster contains values of a single attribute.
- The values refer to regularly spaced locations or uniform rectangular cells rather than to any real spatial objects or to administrative units.
- Values between raster nodes are derived by means of interpolation.

Raster Data Visualisation (1)

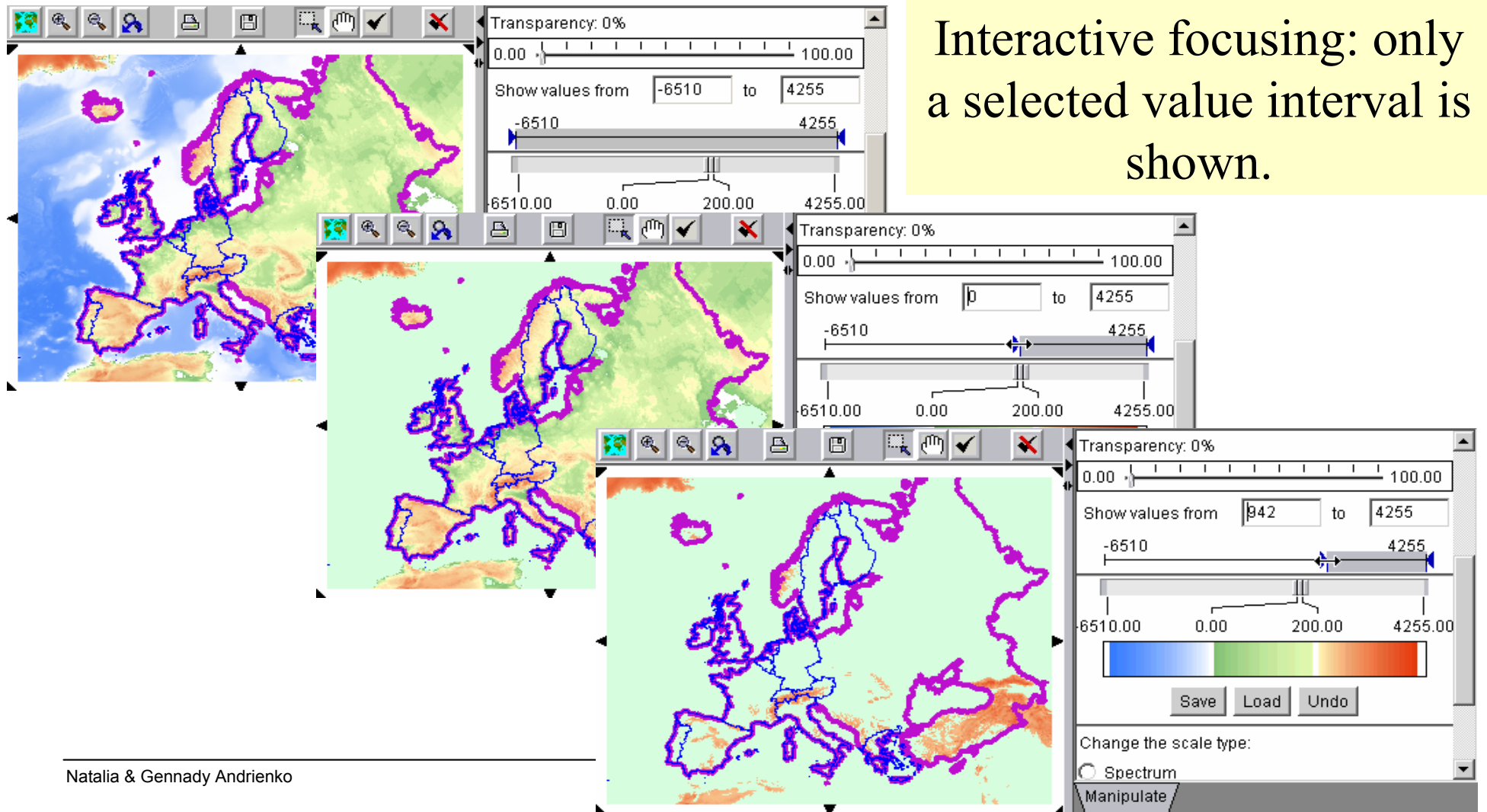


Raster Data Visualisation (2)

This is an example of applying different colour scales to the same data.



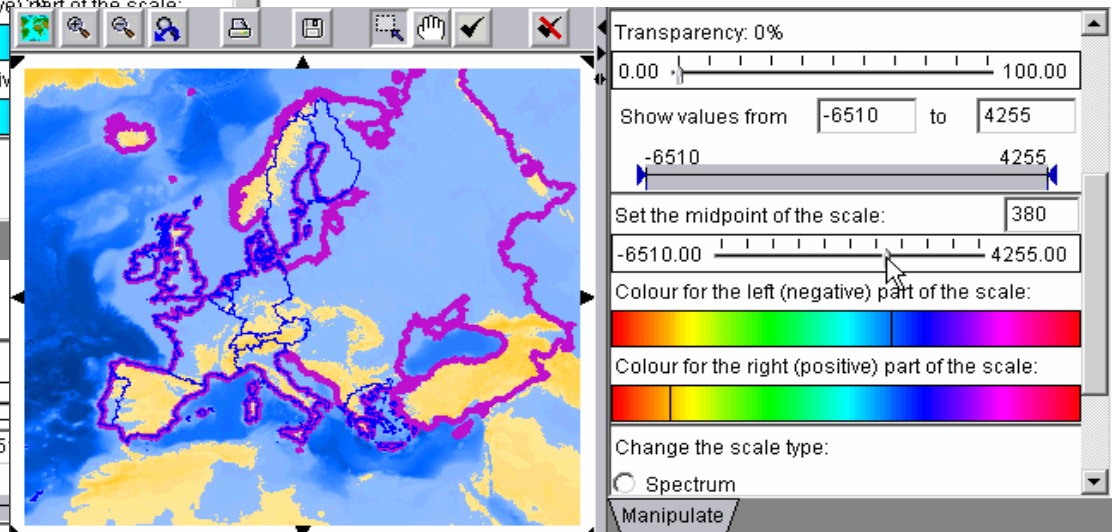
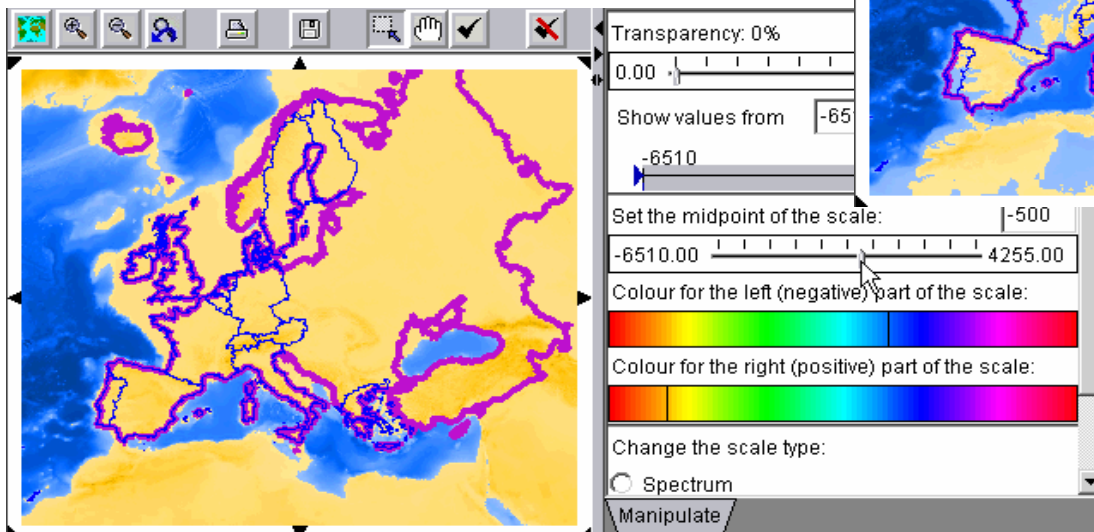
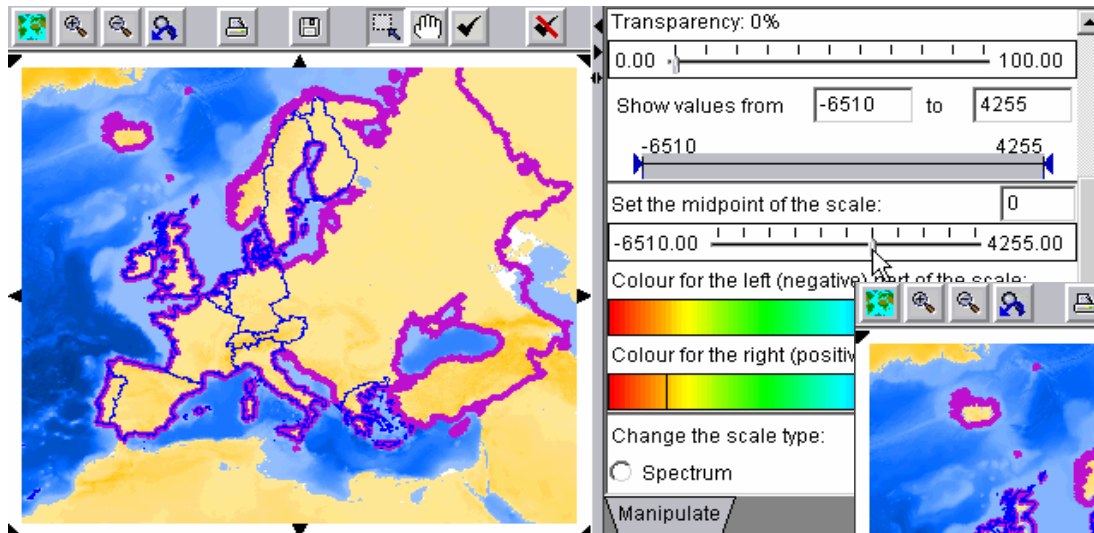
Manipulating Raster Visualisation 1



Interactive focusing: only a selected value interval is shown.

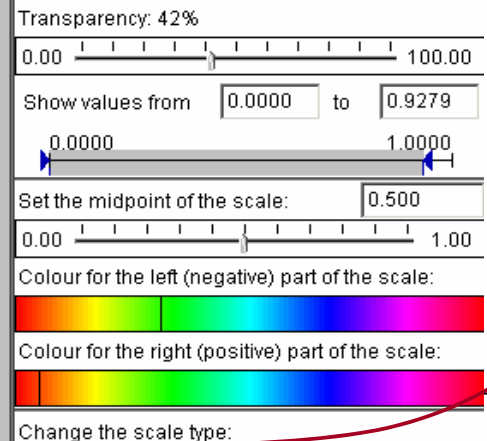
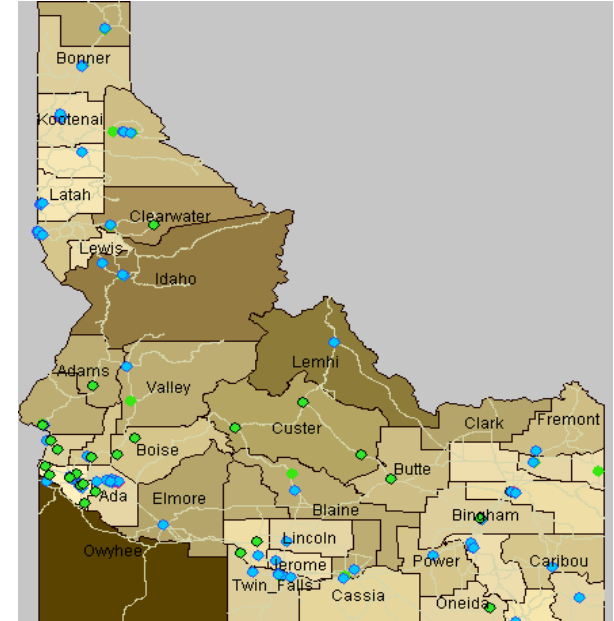
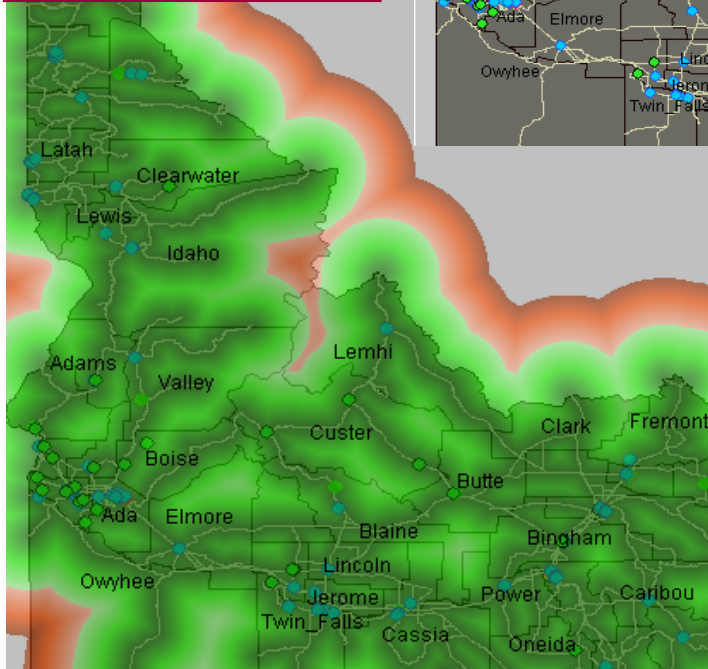
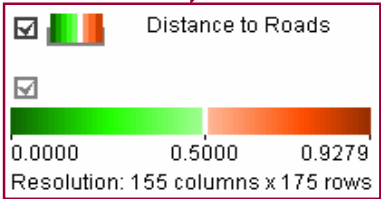
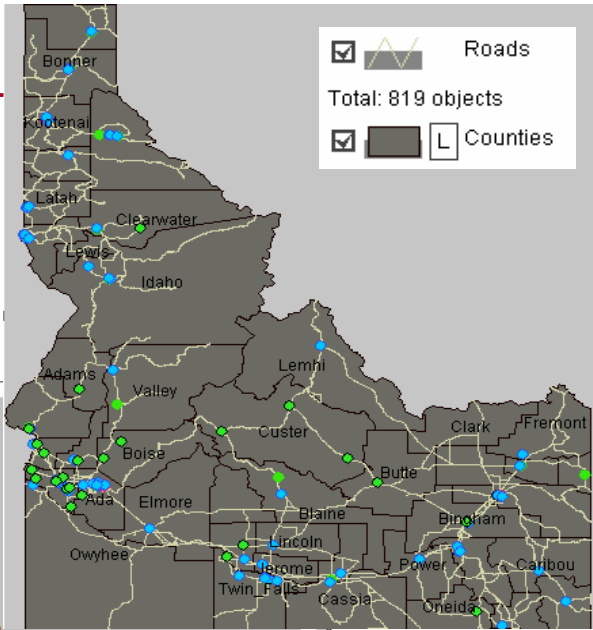
Manipulating Raster Visualisation 2

Interactive variation of the midpoint of the diverging colour scale.



Derivation of additional attributes from raster data

Grid from vector



	Average distance to roads
Owyhee	0.27951637
Lemhi	0.19848134
Idaho	0.1943852
Clearwater	0.1561323
Custer	0.1436425
Clark	0.13796966
Elmore	0.13754171
Valley	0.13168082
Washington	0.12958139
Blaine	0.12722465
Boundary	0.12600437
Adams	0.123795286
Minidoka	0.12363555

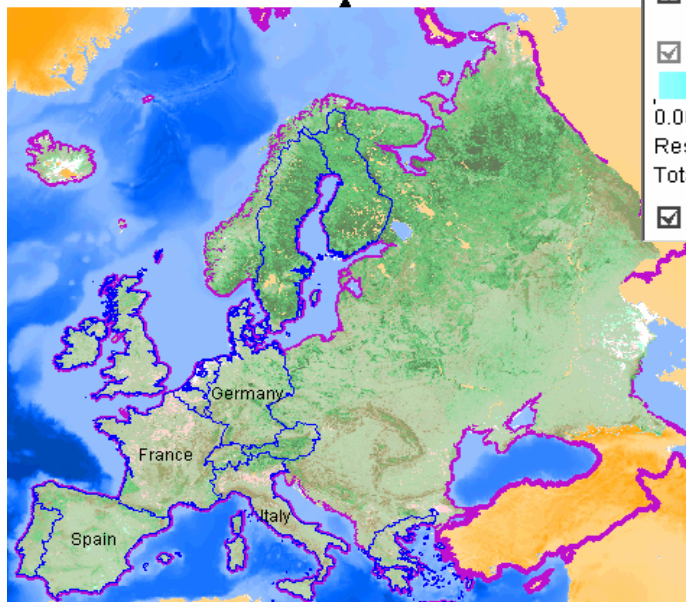
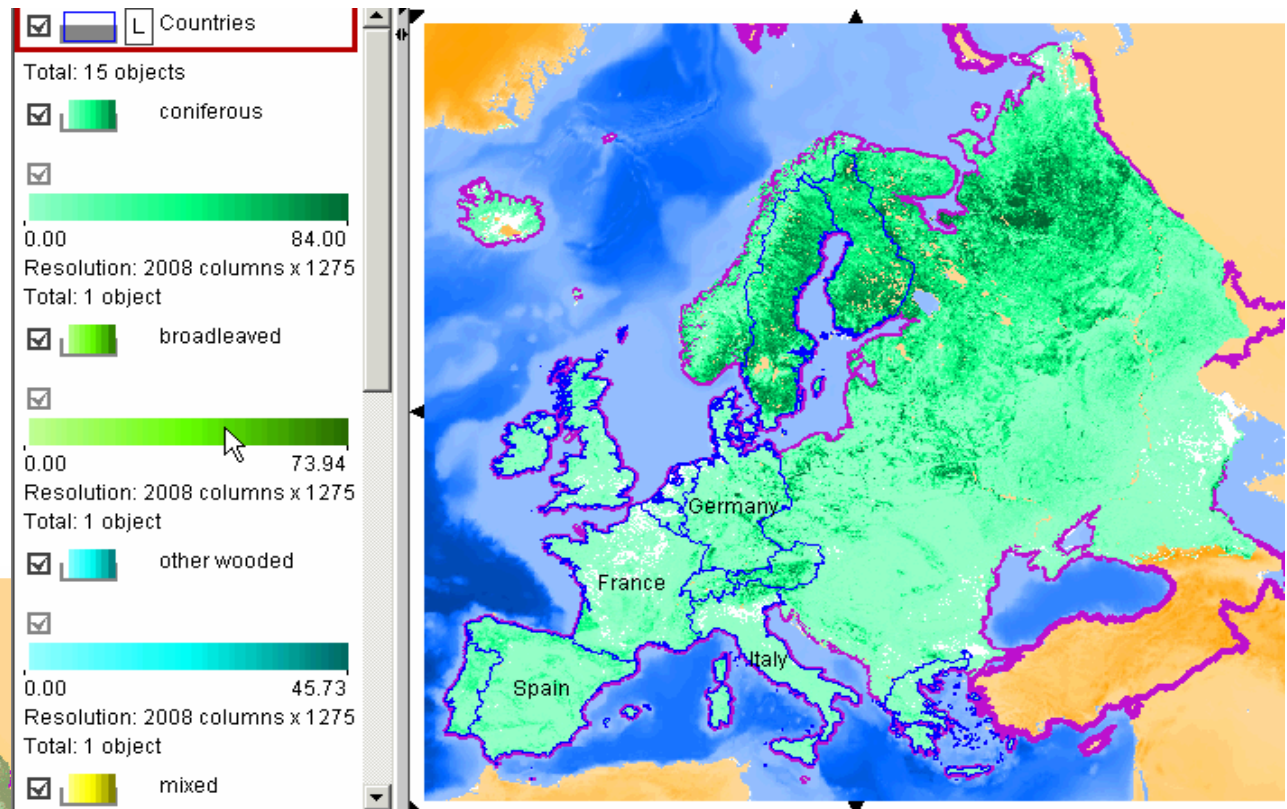
Visualise the attribute

Attribute from grid

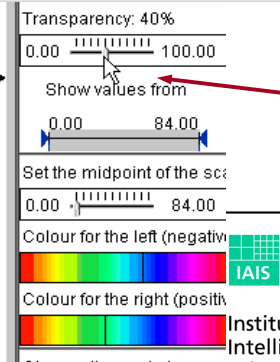
Some Problems with Rasters...

How to visualise and analyse multiple rasters?

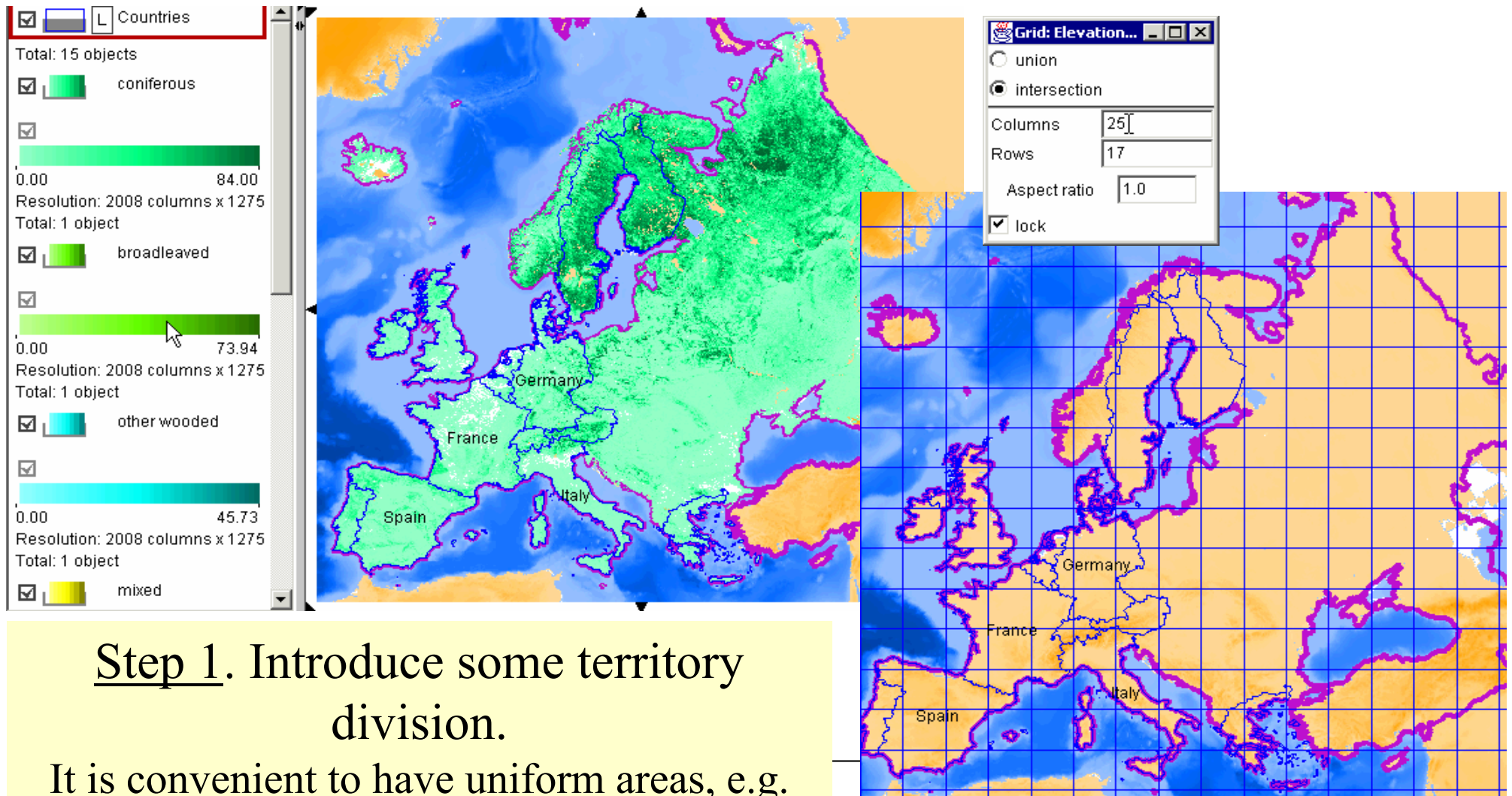
We can see only the topmost raster layer...



Transparency does not really help



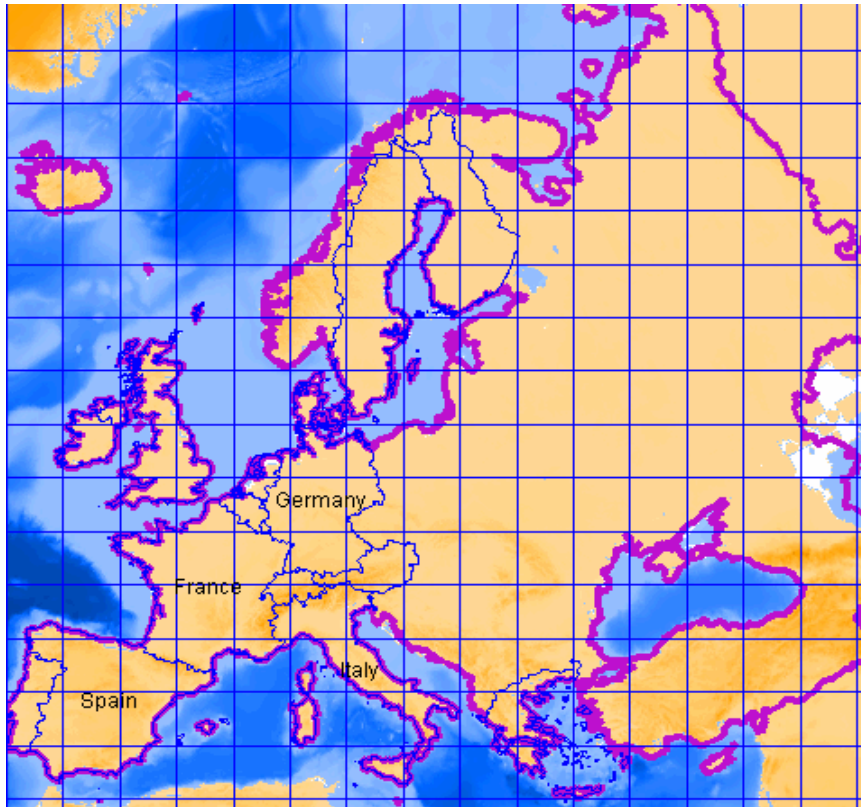
Interactive Raster-Table Bridge (1)



Step 1. Introduce some territory division.

It is convenient to have uniform areas, e.g. equal-size rectangles.

Interactive Raster-Table Bridge (2)

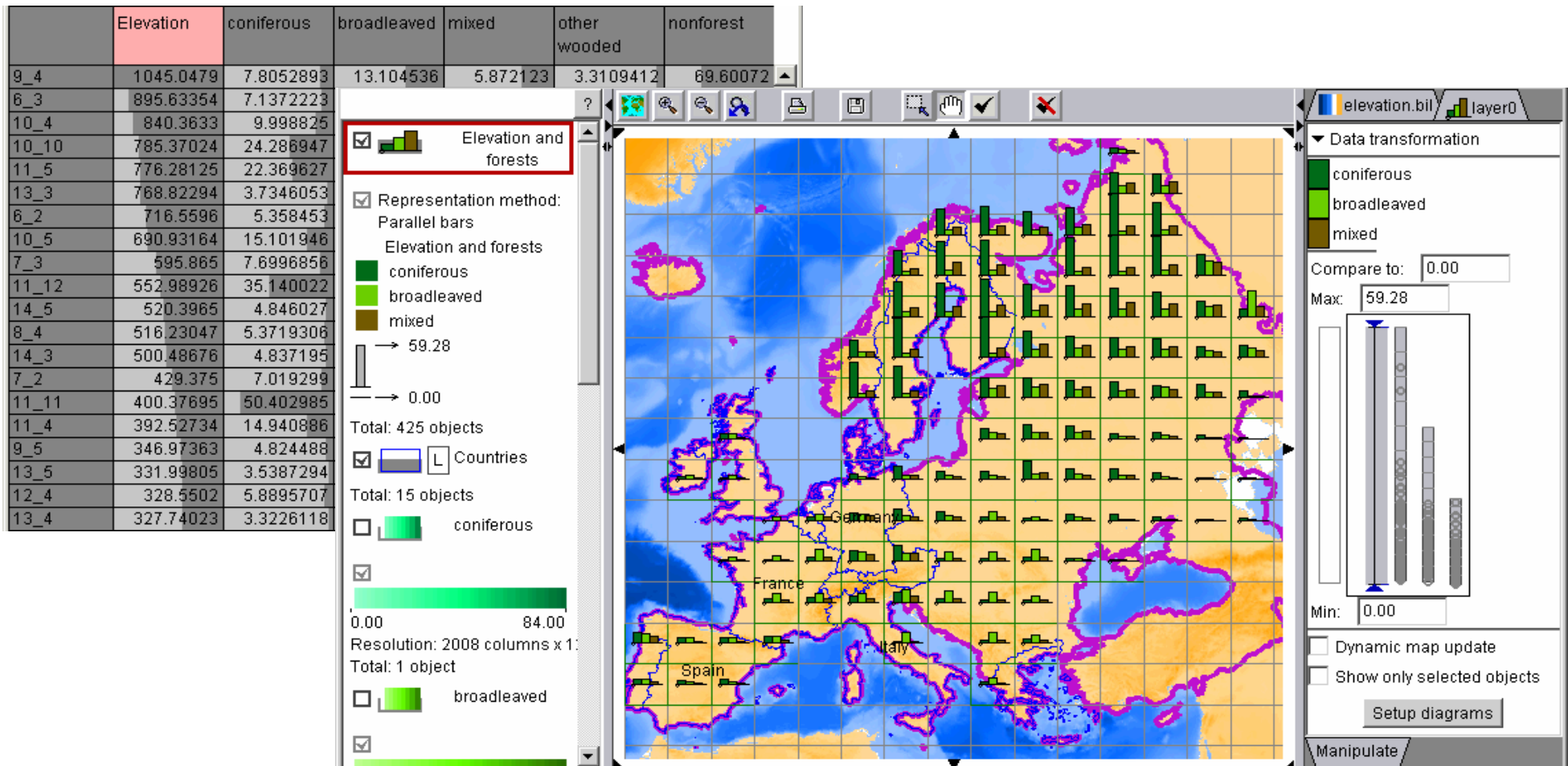


	Elevation	coniferous	broadleaved	mixed	other wooded	nonforest
9_4	1045.0479	7.8052898	13.104536	5.872123	3.3109412	69.60072
6_3	895.63354	7.1372223	7.4486217	2.7511675	20.408121	61.864834
10_4	840.3633	9.998825	12.937923	5.360093	8.05673	63.349785
10_10	785.37024	24.286947	6.970478	7.457732	2.552092	58.571743
11_5	776.28125	22.369627	11.365161	10.576998	2.7555962	52.743294
13_3	768.82294	3.7346053	14.303546	4.090813	9.760096	67.81471
6_2	716.5596	5.358453	4.5545135	1.521516	28.371563	59.962135
10_5	690.93164	15.101946	12.330335	8.694795	2.7672572	60.91224
7_3	595.865	7.6996856	8.176775	2.6910162	17.918682	63.10536
11_12	552.98926	35.140022	6.285668	8.712789	2.0118082	47.764194
14_5	520.3965	4.846027	17.05582	6.5686135	2.3491848	68.97509
8_4	516.23047	5.3719306	13.094806	2.9343753	2.6425462	75.4853
14_3	500.48676	4.837195	11.139068	4.2039137	6.8322487	72.657166
7_2	429.375	7.019299	3.918073	1.3403729	30.82854	56.655293
11_11	400.37695	50.402985	7.506863	13.255096	2.081928	26.623074
11_4	392.52734	14.940886	18.63655	9.580678	6.1169367	50.472378
9_5	346.97363	4.824488	16.084341	5.6440325	2.5189521	70.59787
13_5	331.99805	3.5387294	13.52229	4.317274	2.6508849	75.72668
12_4	328.5502	5.8895707	15.993362	6.596373	3.126134	68.179375
13_4	327.74023	3.3226118	14.132109	4.1035585	3.0310817	75.17449

Step 2. Derive attributes of the rectangular areas from the rasters

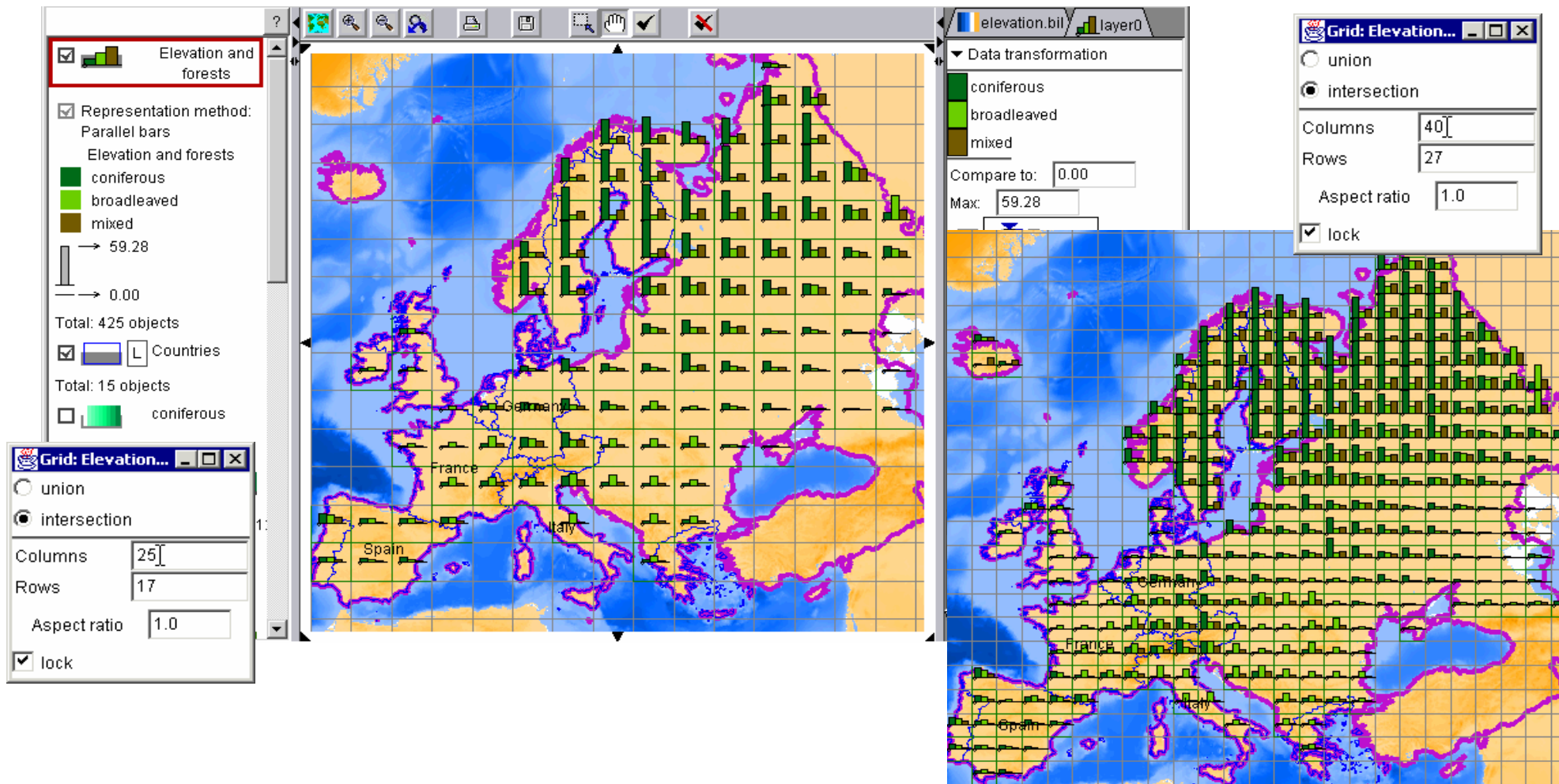
...like we did before for the counties of Idaho

Interactive Raster-Table Bridge (3)



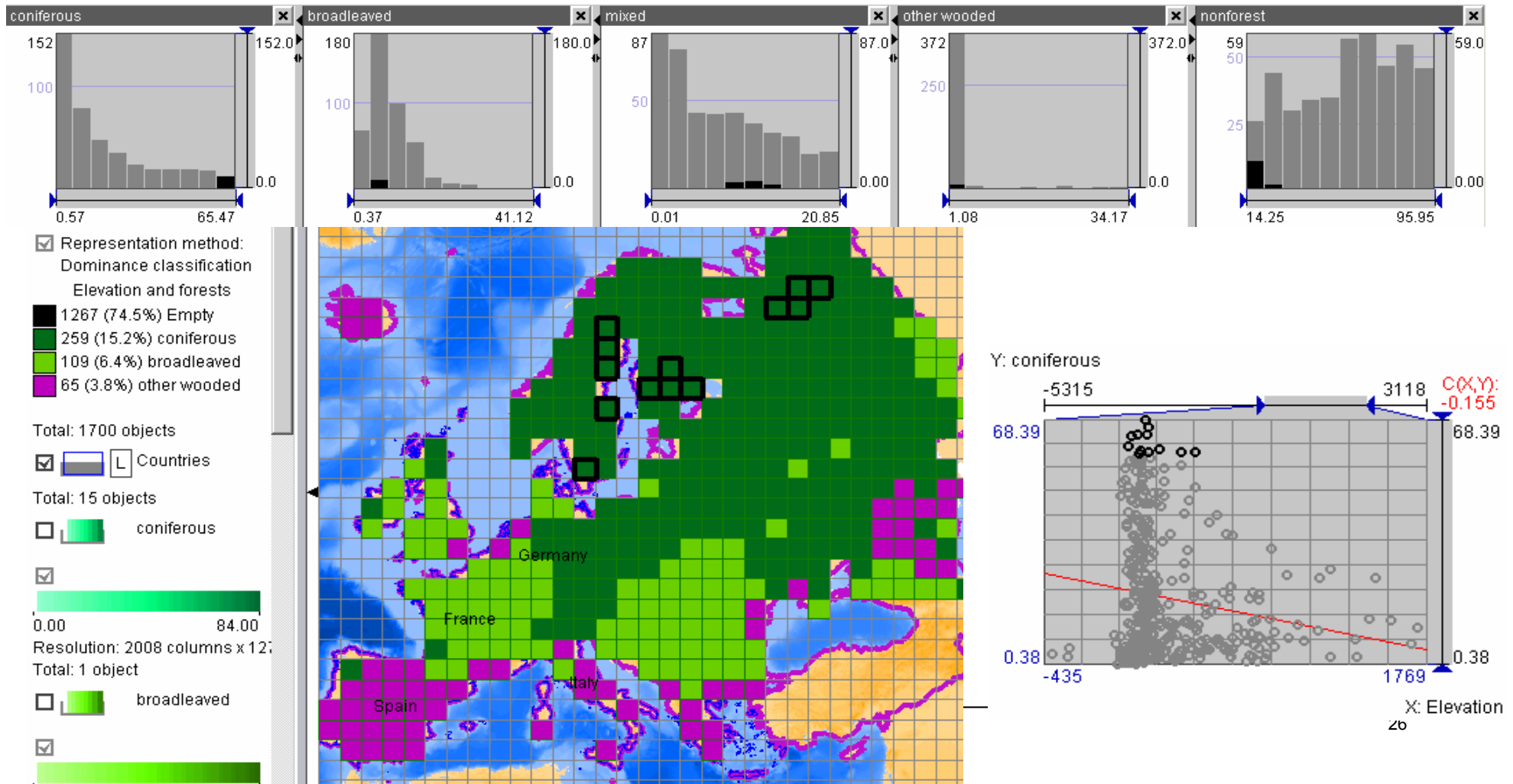
Step 3. Visualise and analyse the resulting attributes as usual

Interactive Raster-Table Bridge (4)

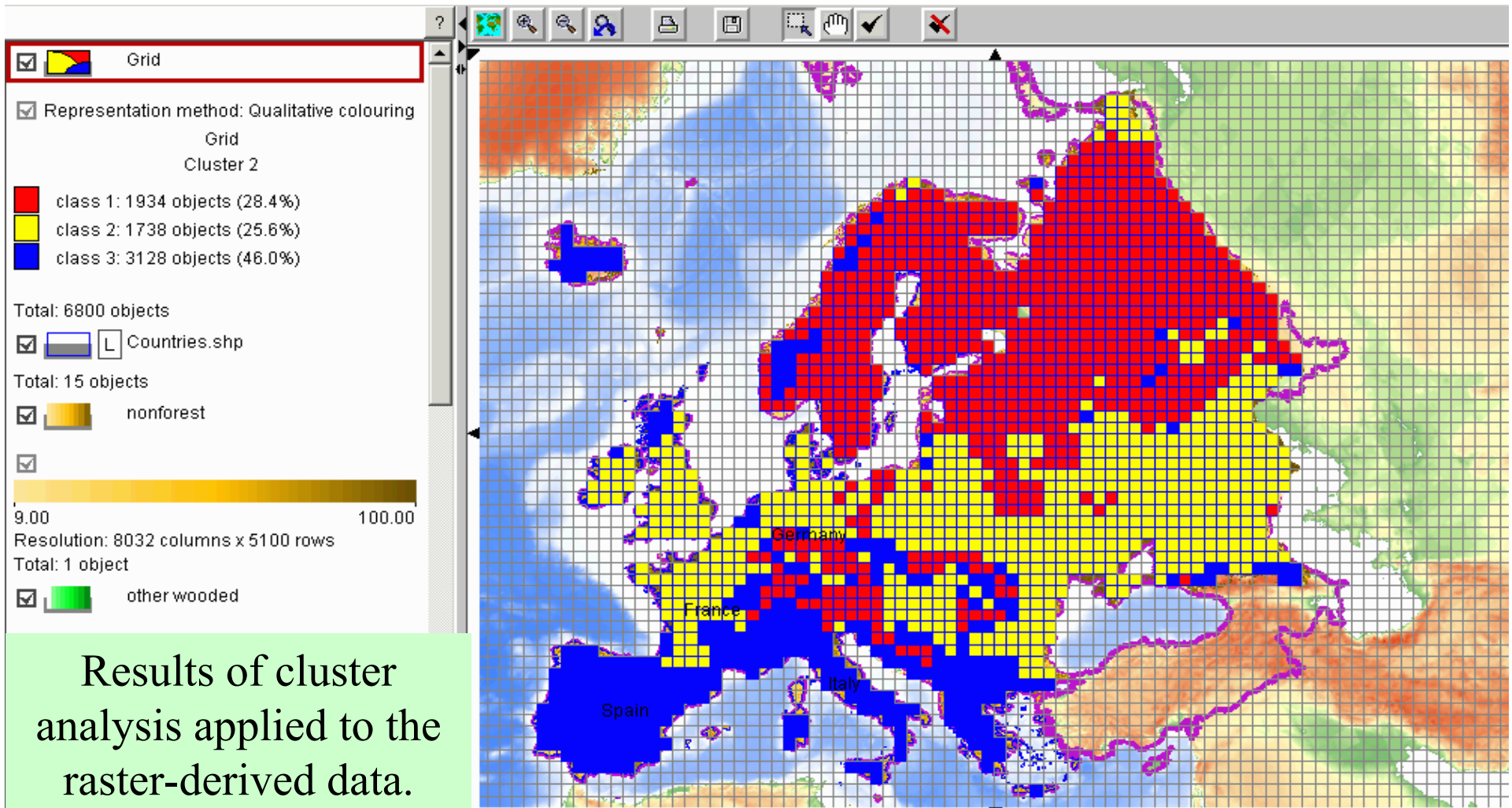


The sizes of the rectangles (raster-table bridge resolution) can be interactively changed. The attributes are automatically re-computed and the displays updated.

Further Possibilities for Analysis of Raster-Derived Attributes



Yet More Analysis: Data Mining

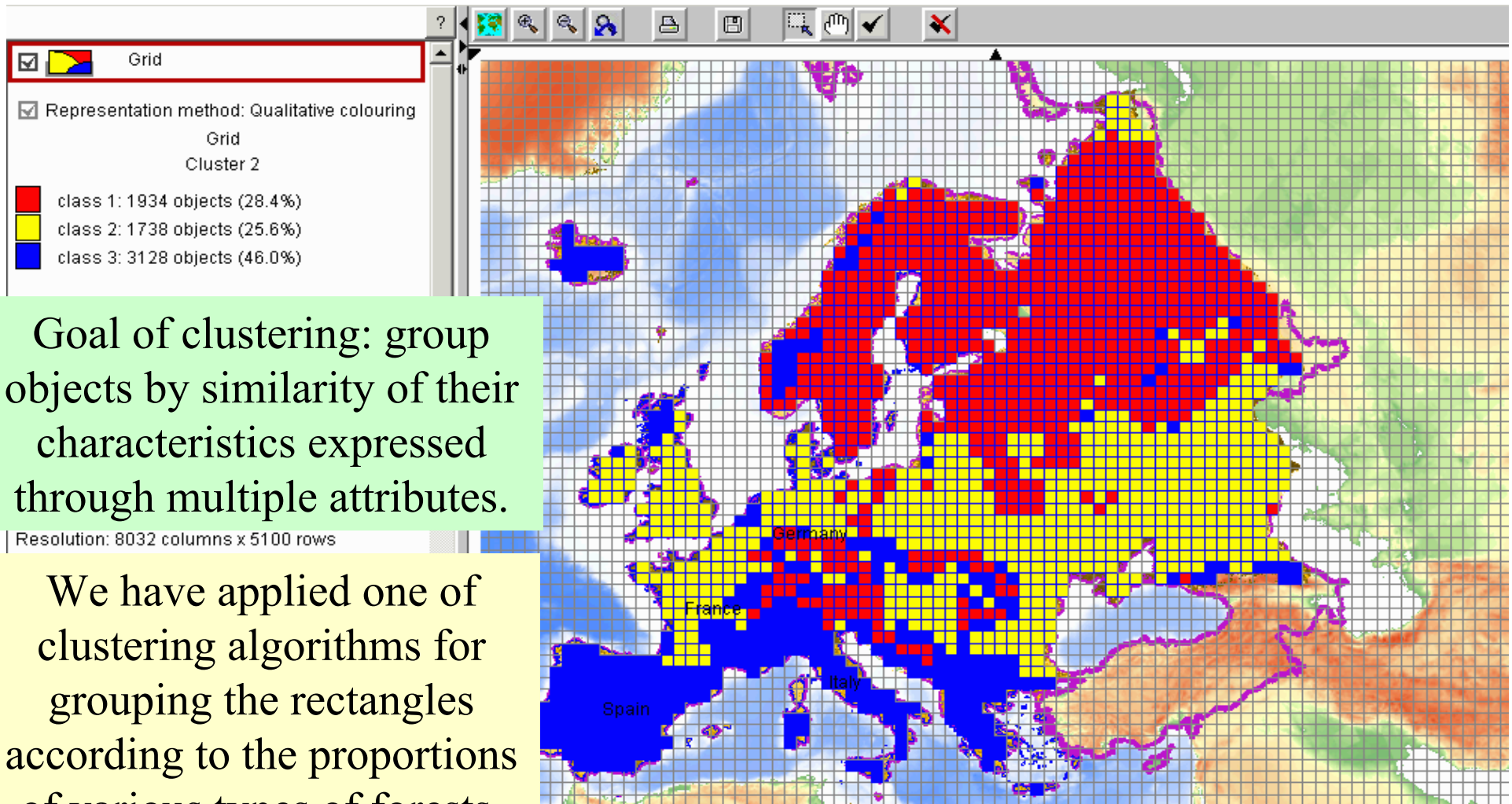


Results of cluster analysis applied to the raster-derived data.

Natalia & Gertraud Anthonerko



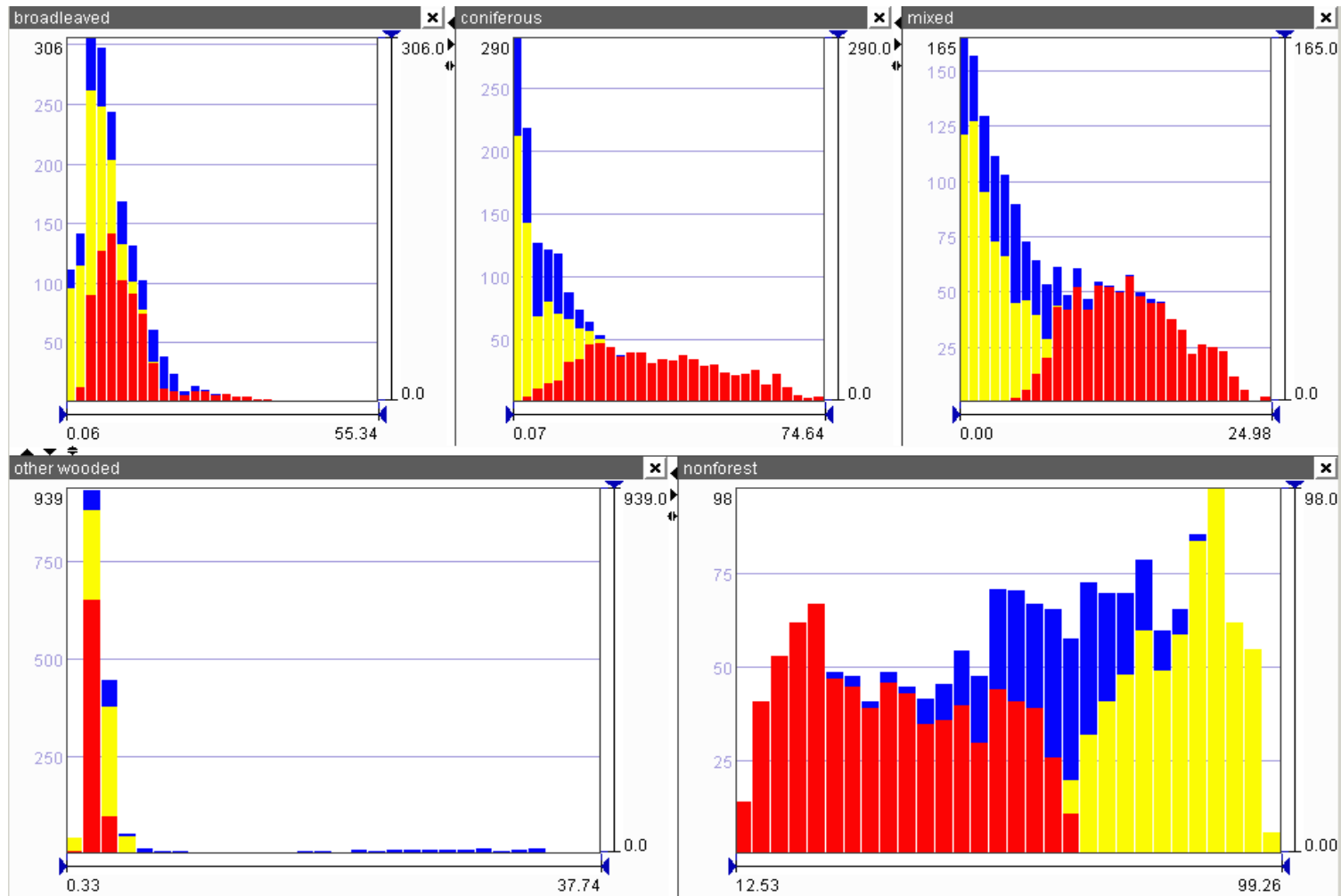
Cluster Analysis



Goal of clustering: group objects by similarity of their characteristics expressed through multiple attributes.

We have applied one of clustering algorithms for grouping the rectangles according to the proportions of various types of forests.

Visualisation and Interpretation of Cluster Analysis Results



Summary

- Data transformations play an important role in visual analysis

We considered several examples; more to come in the next lectures

- Absolute attribute values → relative (difference, ratio); normalisation
 - standard normal transformation (z-score): comparison of spatial distributions of different attributes
- Clustering based on attribute values: characterise objects or areas in terms of multiple attributes; divide territory into regions
- Computation of new attributes from vector and raster data: additional data useful for analysis
- Integration of data from several raster layers into a multi-attribute table: allows us to use additional analytical facilities (visualisation by diagrams, various kinds of attribute transformations, statistical graphics, clustering)

See also

- Natalia and Gennady Andrienko
Exploratory Analysis of Spatial and Temporal Data
A Systematic Approach
Springer-Verlag, December 2005

section 4.5, pp.259-335

