Geospatial Visual Analytics:
Interactive Maps and Multiple Coordinated Views

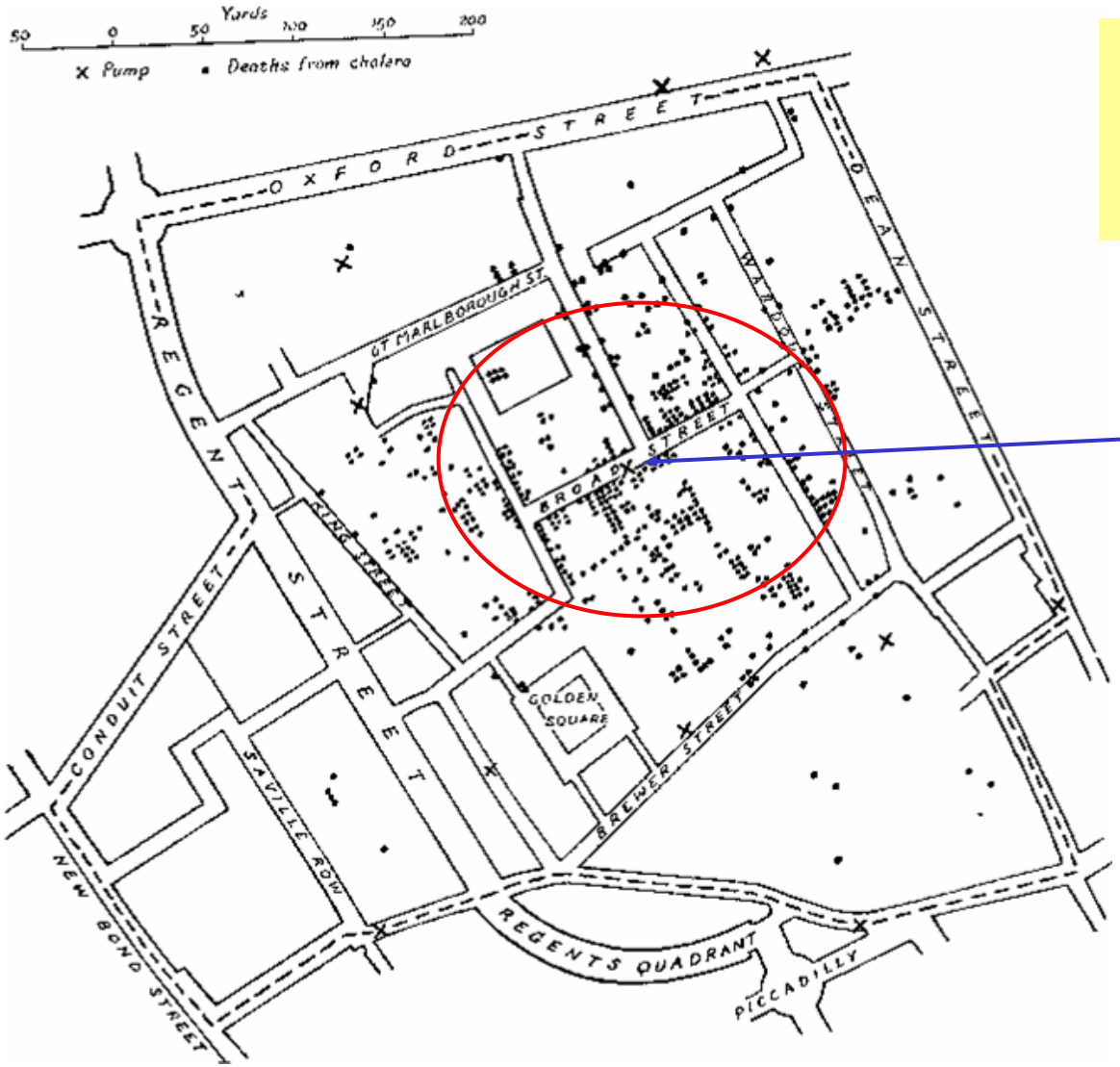
Fraunhofer Institut
Intelligente Analyse- und
Informationssysteme

Gennady Andrienko & Natalia Andrienko
http://geoanalytics.net

Maps: not only for orientation!

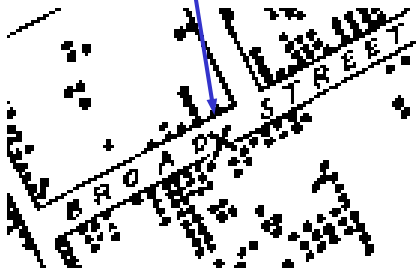
- People live in geographical space. Most of people's decisions and actions depend on
  - where the things are;
  - how are their locations related.
- Maps allow people to perceive the space beyond the directly observable extent.
- A map serves as a model of reality and helps to detect patterns existing in the reality

Example: Dr. John Snow's discovery



Map of locations of deaths from cholera
London, September 1854

infected water pump



Interactive maps

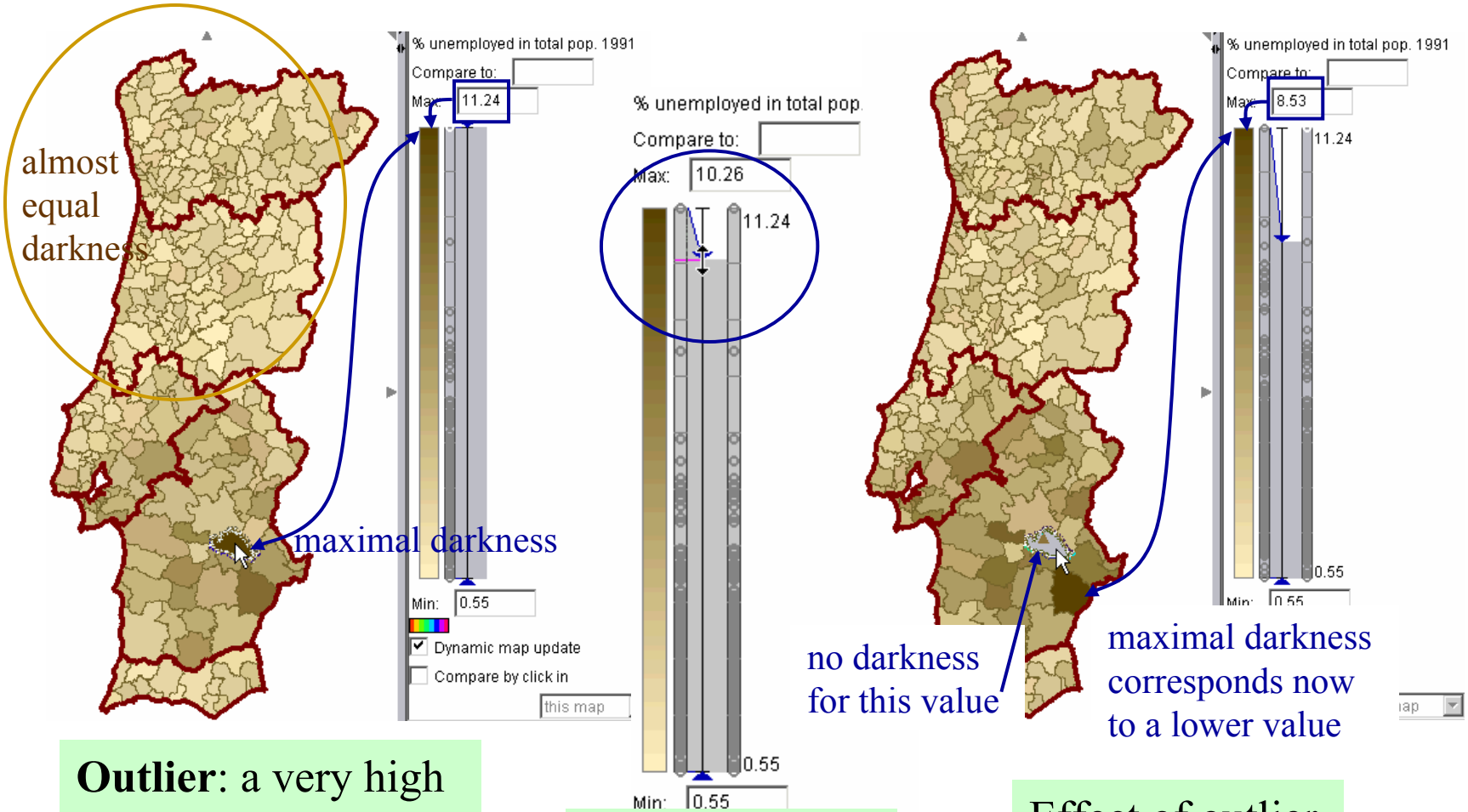
- Interactive maps can change in response to user's actions
- Many interactive maps are available on the Web, e.g. street maps, tourist maps, election maps, …
- Interaction techniques are used to
  - compensate for the display deficiencies, e.g. limited size (zoom and pan, showing additional information related to mouse position, …)
  - increase the display expressiveness
  - enable more sophisticated analyses

Typical interactive operations

- Select information layers, e.g. on a tourist map: accommodation, museums, restaurants, nightlife
- Select time moments or intervals in displays of time-related information, e.g. election year
- Change the spatial scale, e.g. states or counties
- Change the theme, e.g. president elections or governor elections, absolute values or differences in comparison to the previous time
- Choose the visualization method, e.g. area painting or proportional symbols

Examples of analytical interactions

Removing Outliers (1)

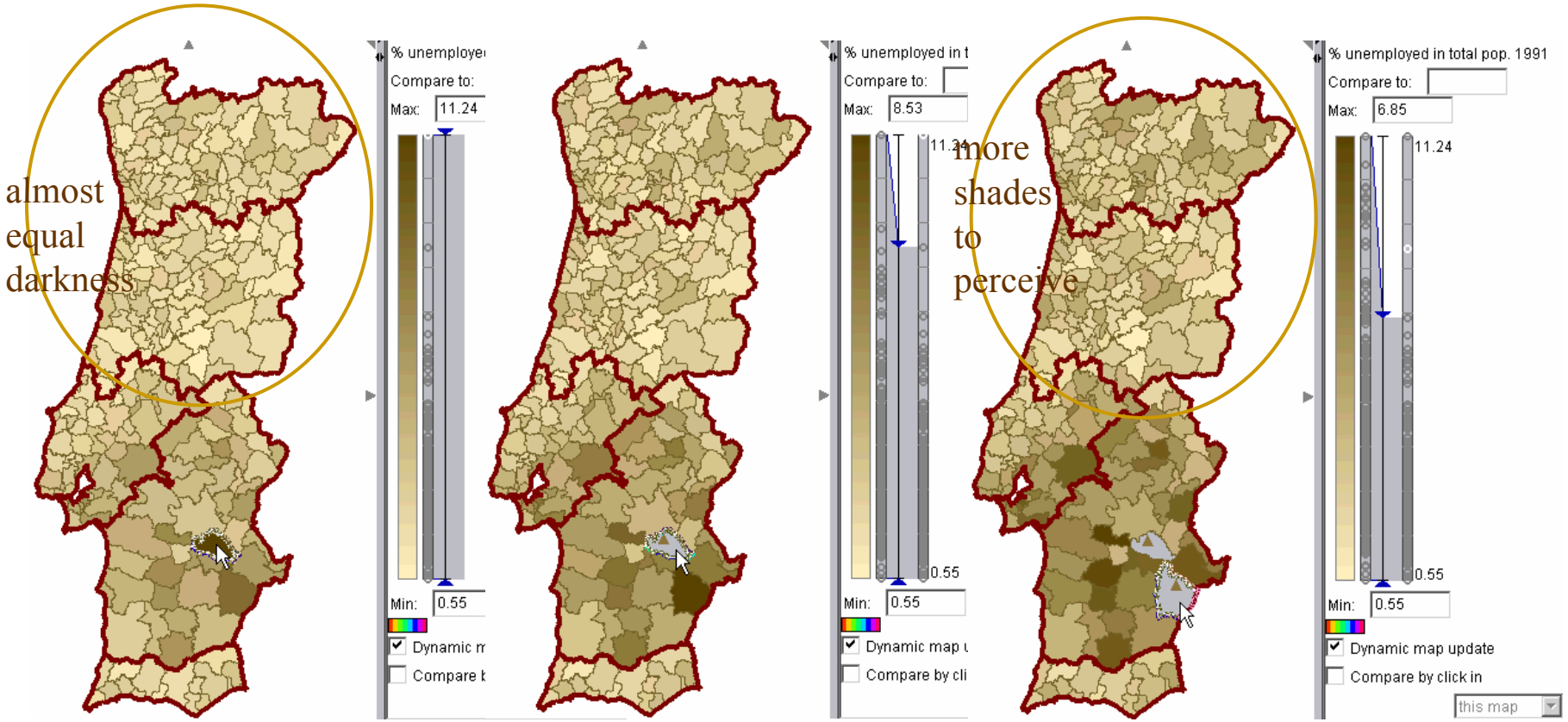


Outlier: a very high (or very low) value, far apart from others

Interactive outlier removal

Effect of outlier removal

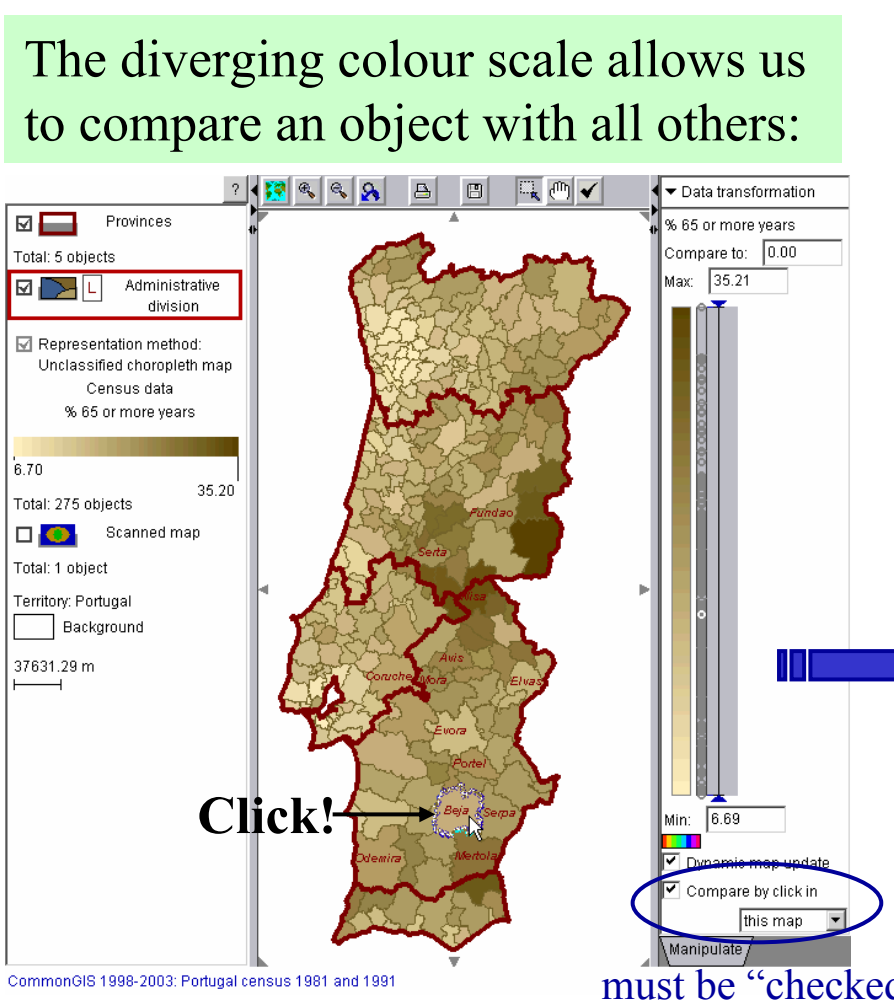
Removing Outliers (2)



After the removal of two outliers, the differences are better seen

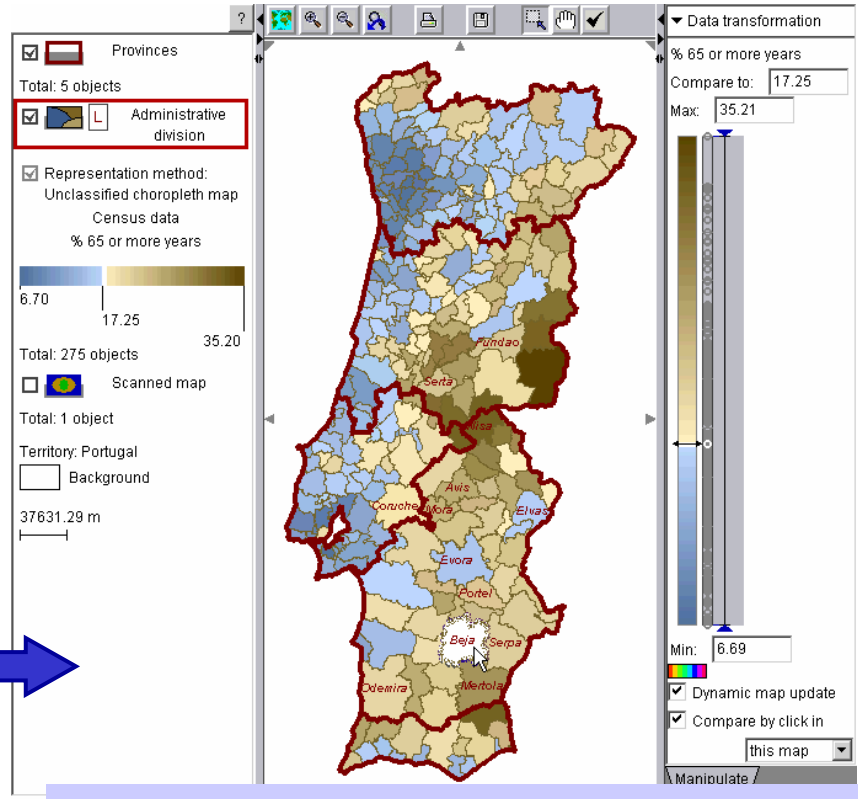
Object Comparison

The diverging colour scale allows us to compare an object with all others:



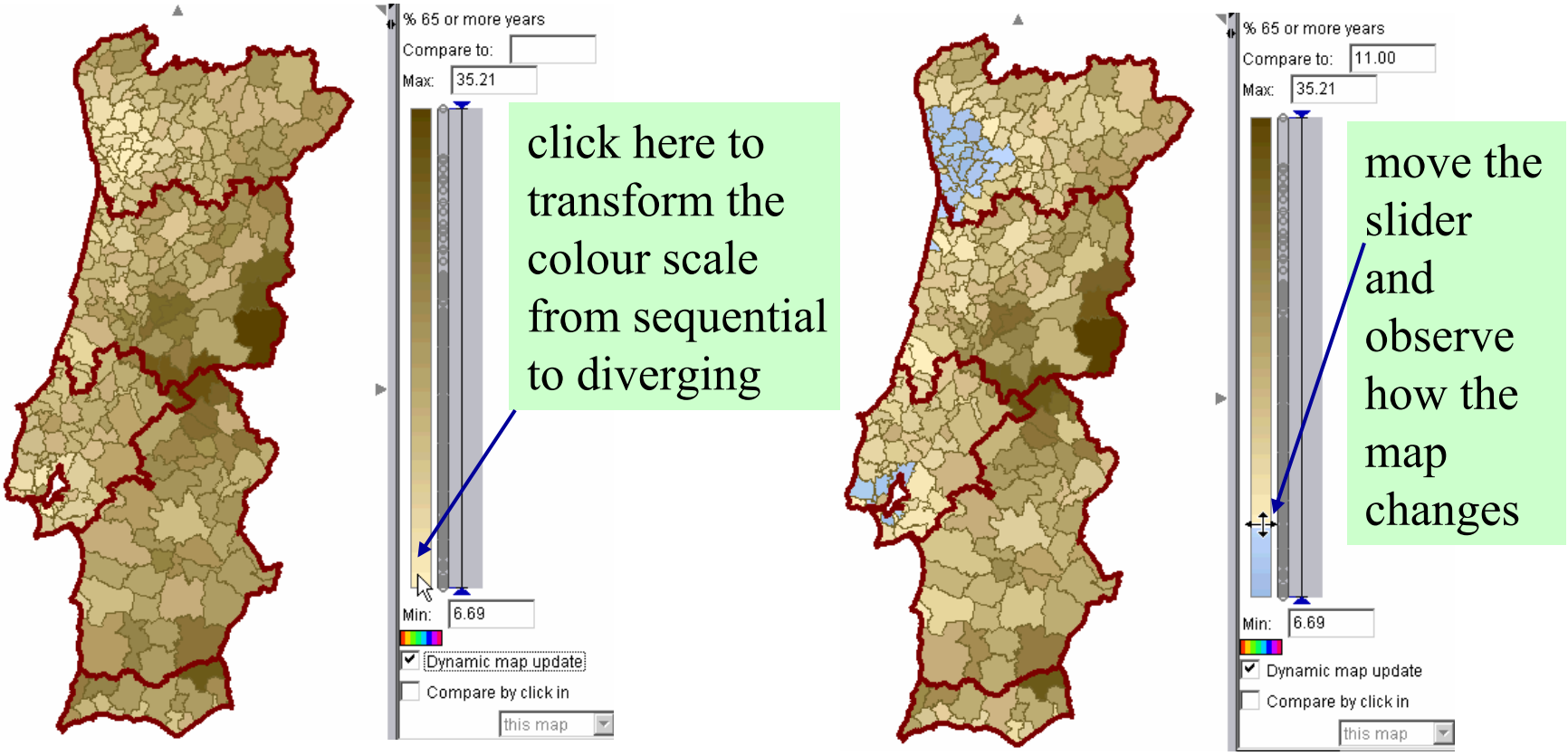
CommonGIS 1998-2003: Portugal census 1981 and 1991

must be "checked"



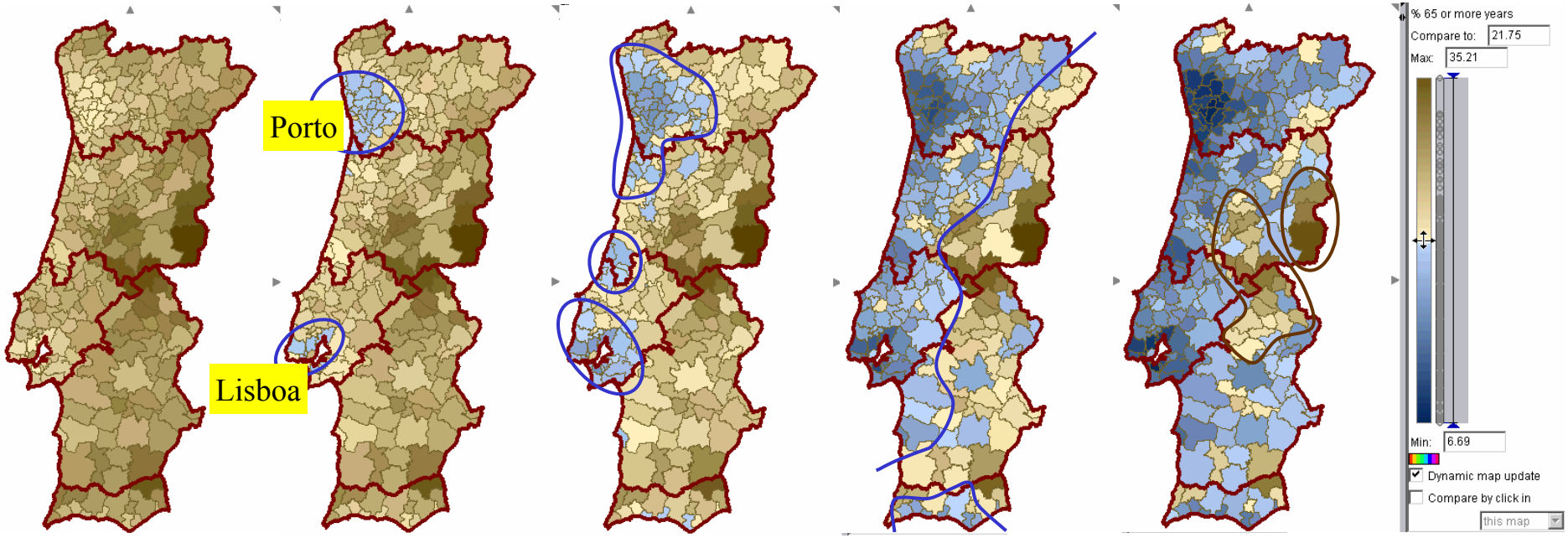
After clicking on Beja:
lower values than in Beja → blue
higher values than in Beja → brown

Pattern Investigation (1)



Pattern Investigation (2)

By moving the slider, we see more patterns and gain more understanding of value distribution



West-to-east increase

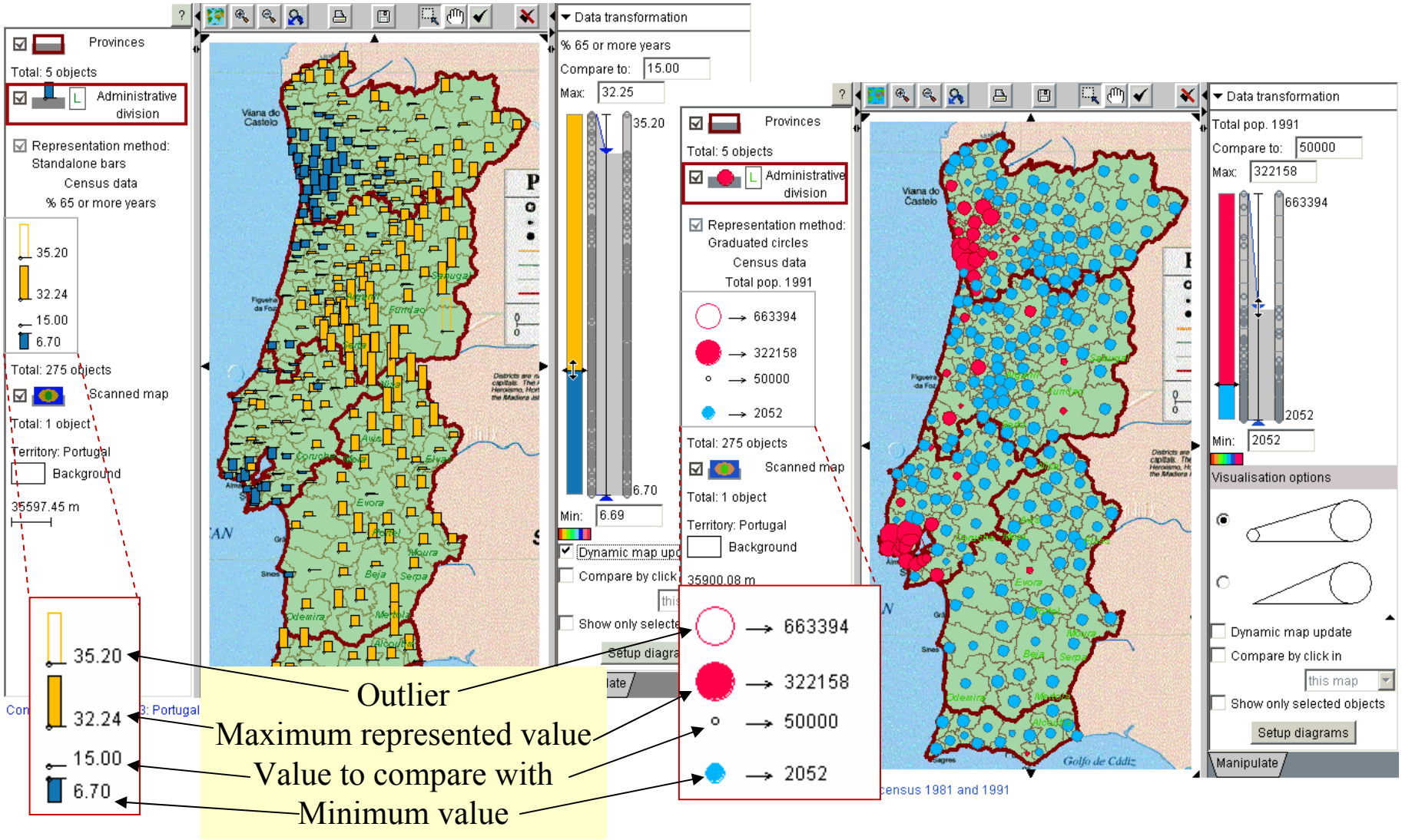
Clusters of low values around Porto and Lisboa

One more cluster of low values

Coast-inland contrast

Clusters of high values in central-east

Focusing and Visual Comparison on Other Map Types

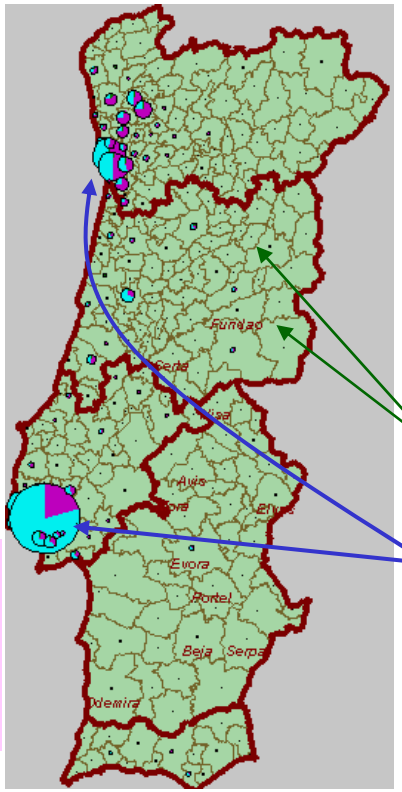


Piechart Map

Applicable to several attributes that together give some meaningful whole

"Pie" size is proportional to the total (sum of the attribute values)

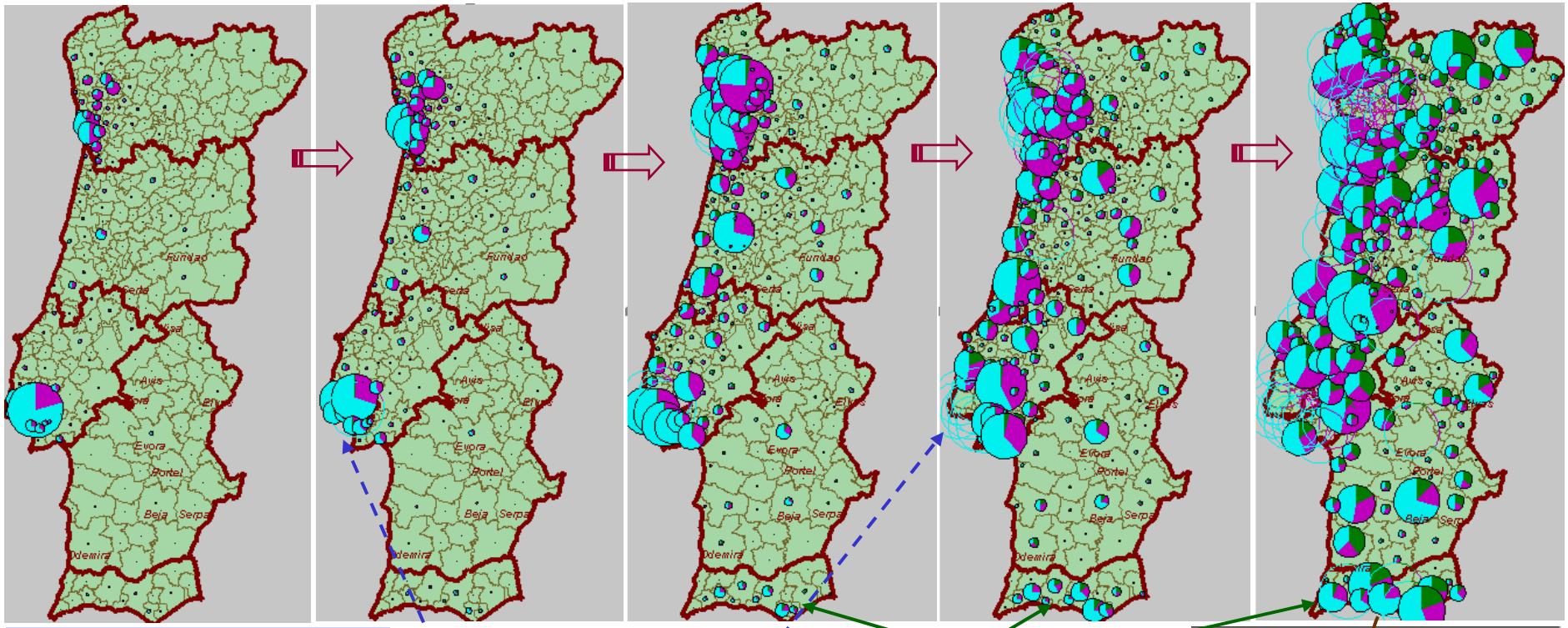
The division into slices shows proportion of each attribute in the total



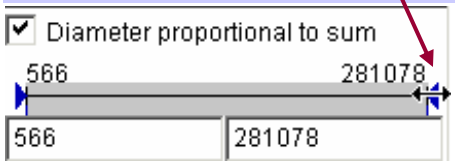
Here the population is very small in comparison to the large cities. Therefore, the pies are too small to be seen

However, the map often looks like this:

Piechart Map: Focusing

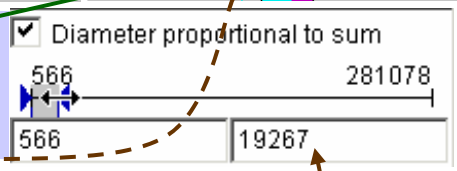


Move this delimiter to the left



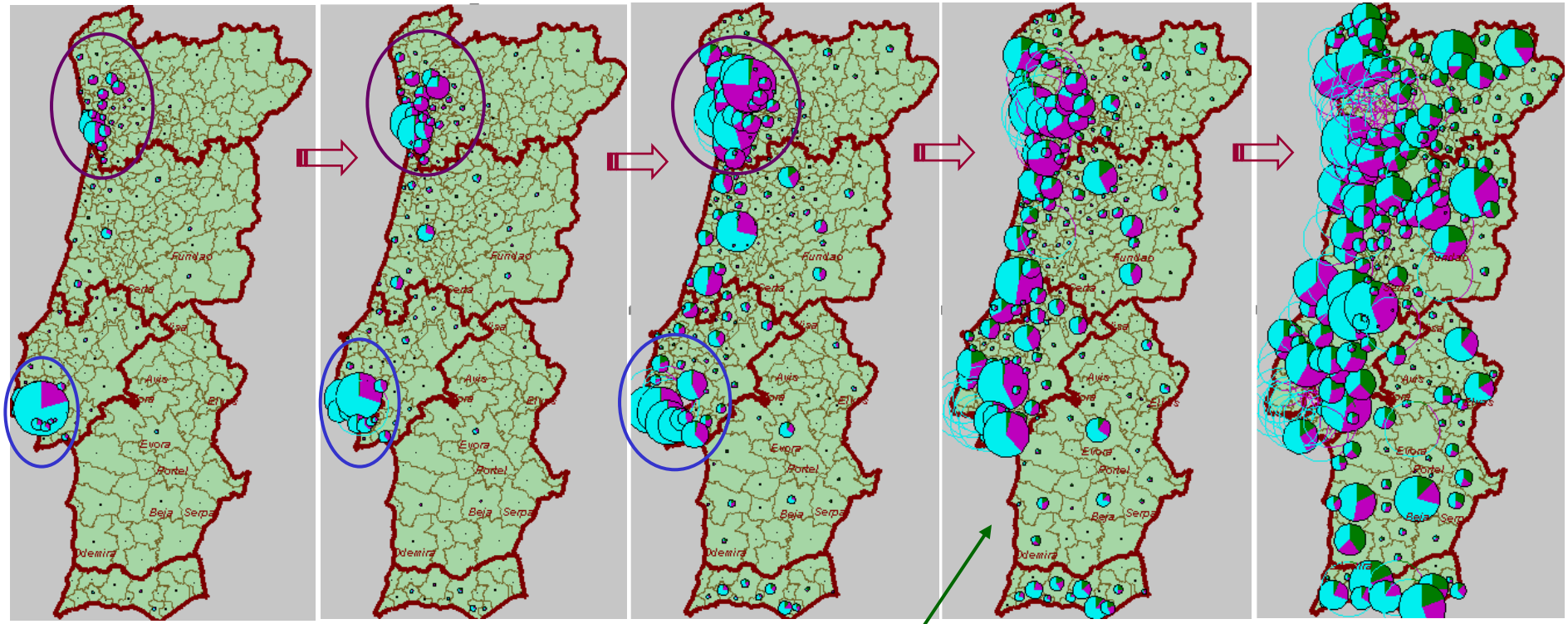
The largest pies are gradually removed (replaced by hollow circles)

The remaining pies become larger



Now the maximum pie size corresponds to this value

Focusing and Data Investigation



In districts with much population people work in industry (magenta) and services (cyan).
Northwest: more industry
Centre-west: more services

At this stage, the agricultural part (green) becomes visible

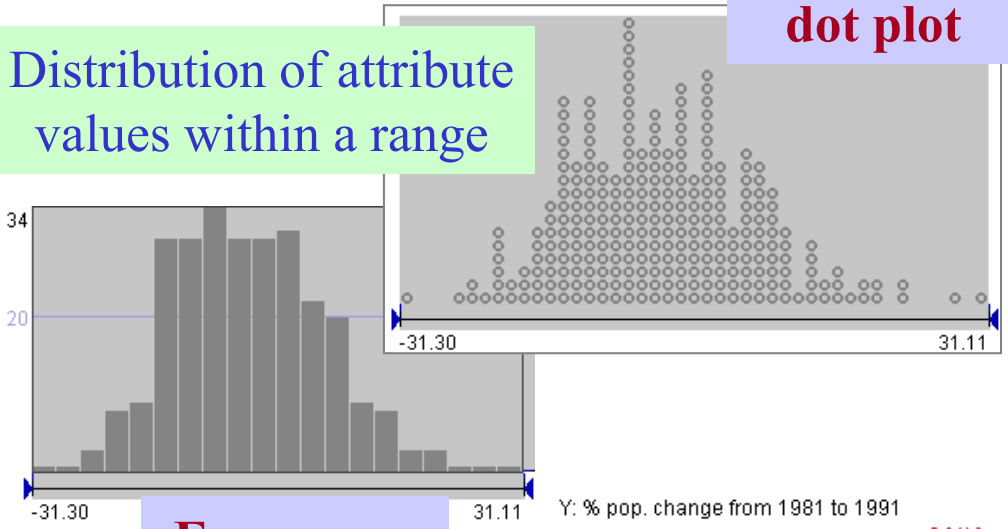
In districts with little population considerable proportion of people works in agriculture, but services still prevail

Why to Use Multiple Views?

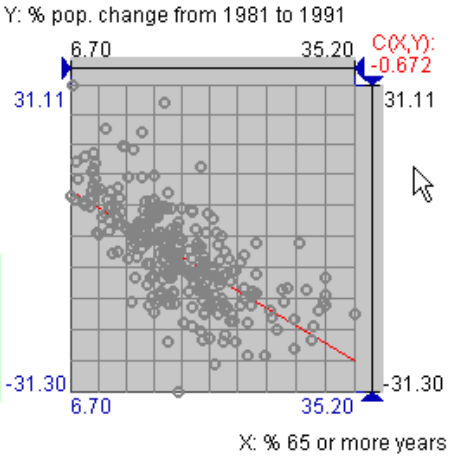
Dot plot

Stacked dot plot

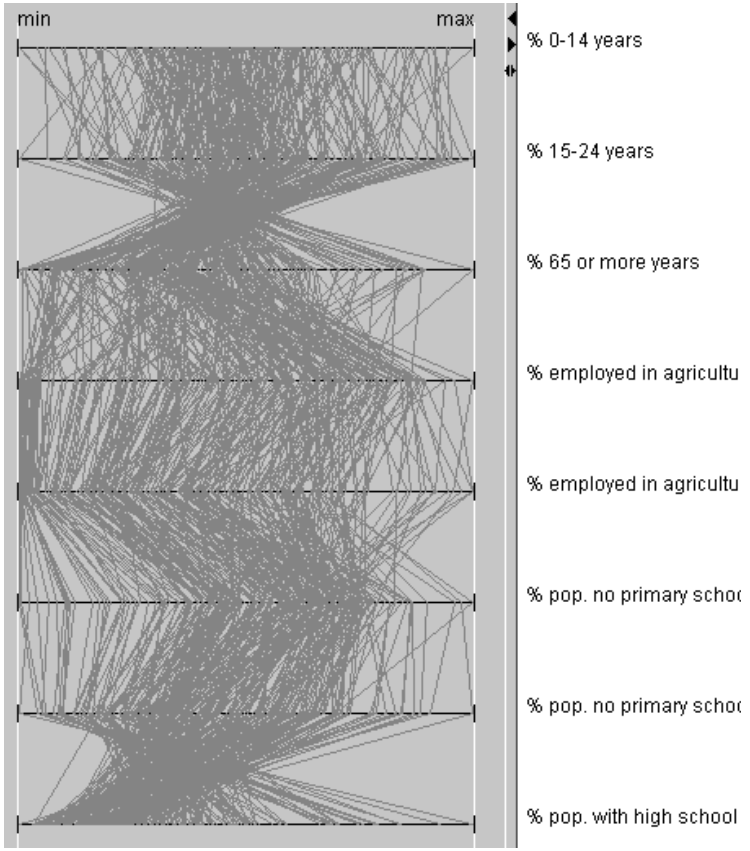
Distribution of attribute values within a range



Frequency histogram

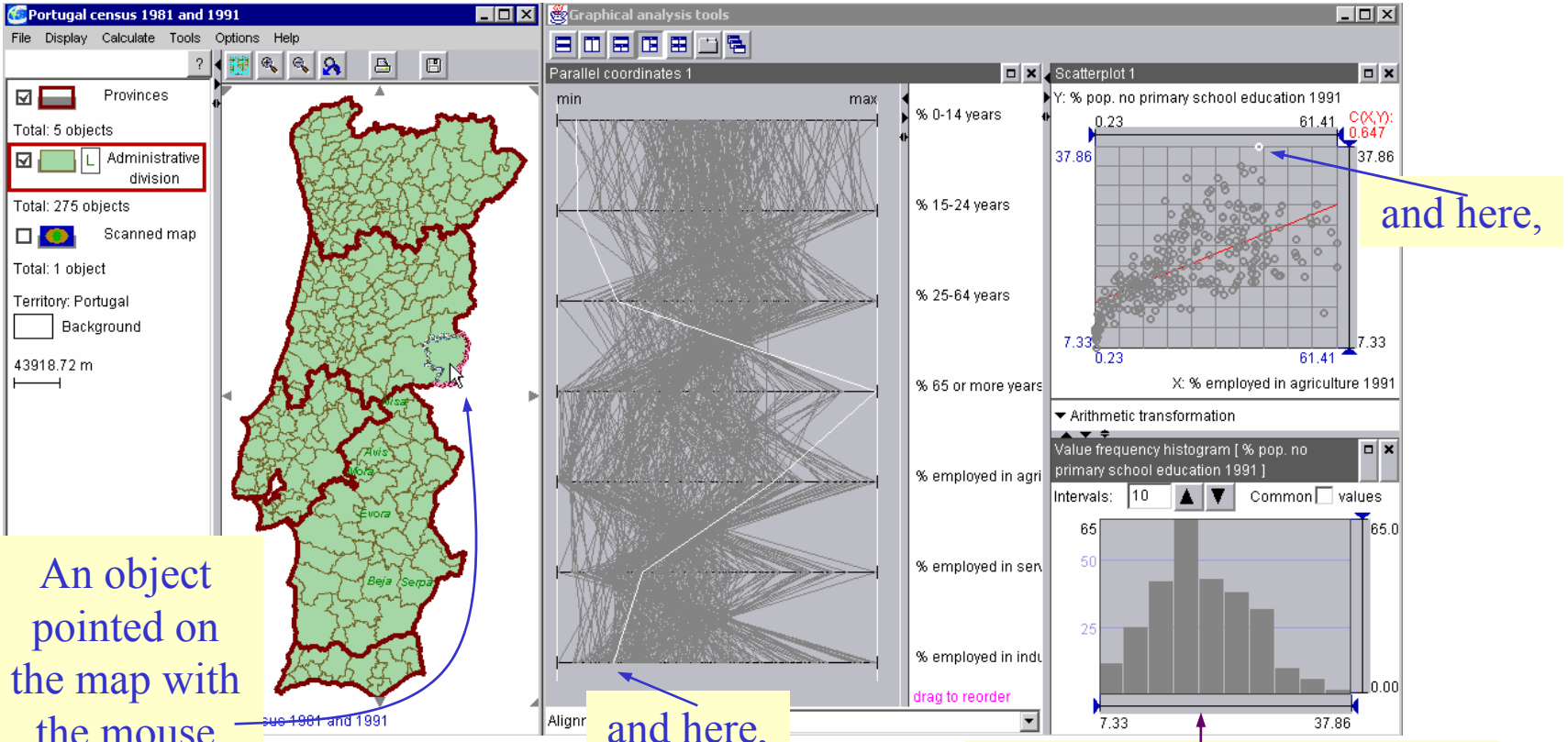


Scatter plot: shows how two attributes are related



Parallel coordinates: object characteristics profiles; relationships between attributes (look at line slopes)

Display Linking by Highlighting



An object pointed on the map with the mouse

is simultaneously highlighted here,

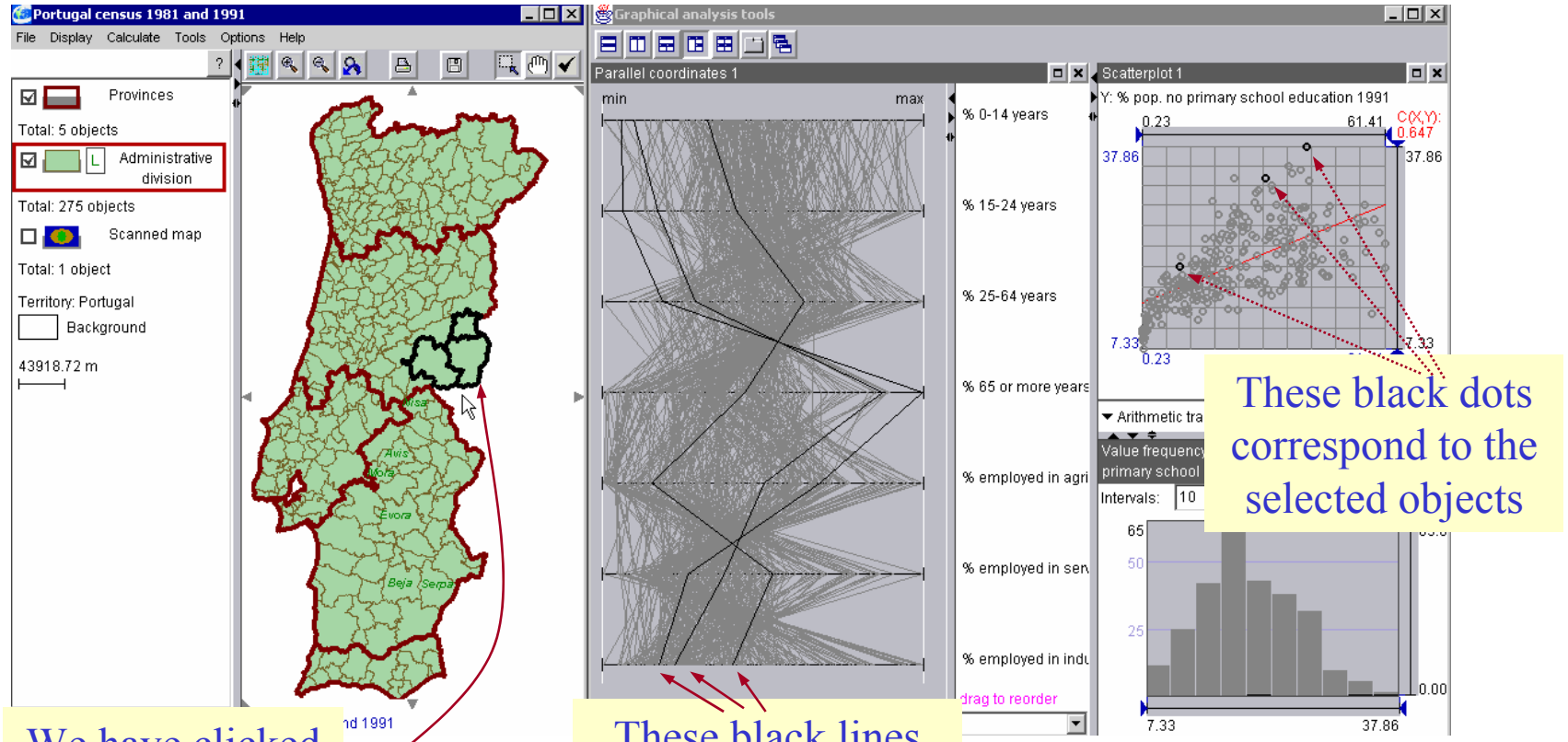
and here,

and here,

but not here: this is an aggregated view that does not show individual objects

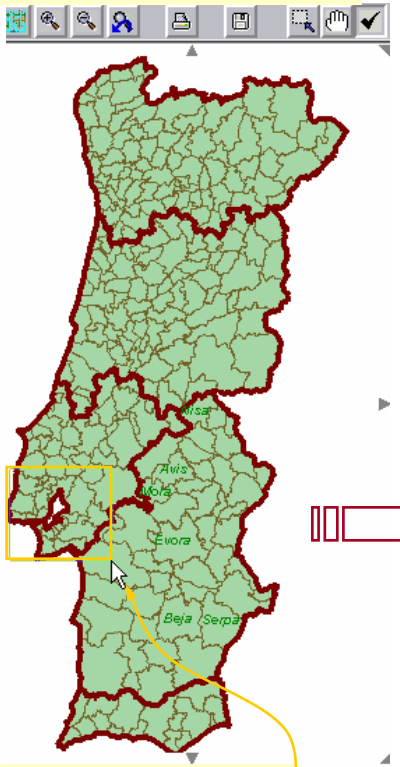
Display Linking by Selection

Selection (durable highlighting) does not disappear after the mouse is moved away. One or more objects may be selected e.g. by mouse-clicking on them.

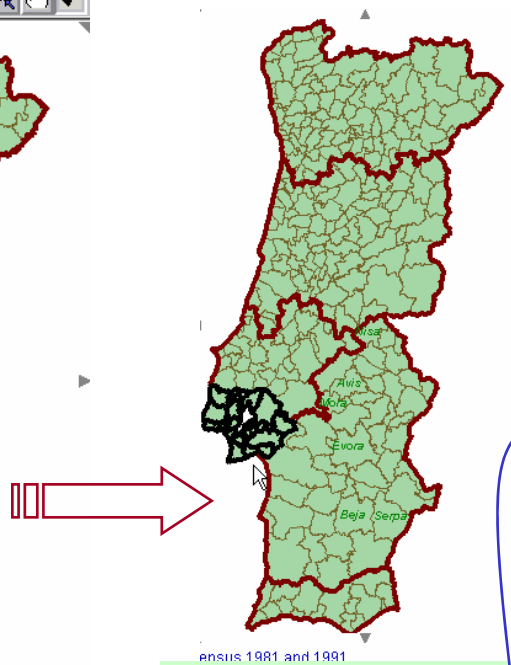


Using Display Linking (1)

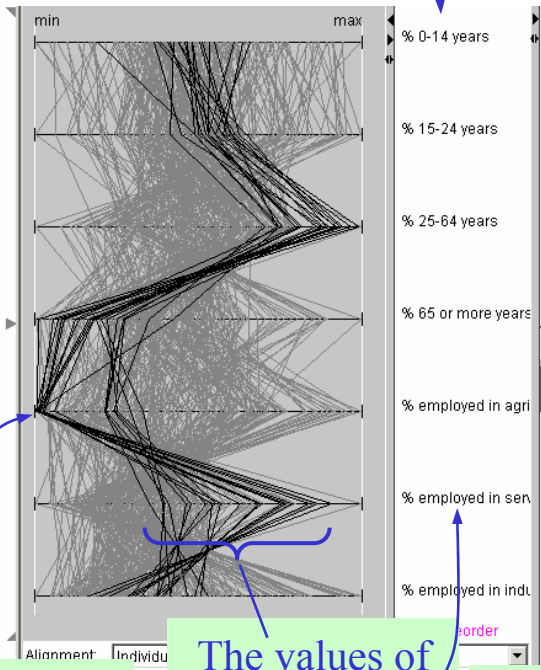
Let us examine characteristics of districts in this area



Enclose the area in a frame



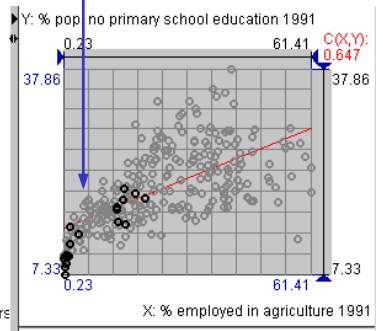
The characteristics in terms of the upper 4 attributes are rather coherent

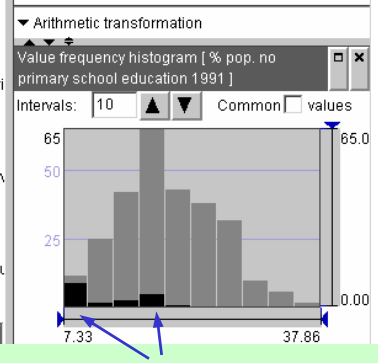


The values of this attribute are split in two groups with a gap between

The values of this attribute greatly vary

Two distinct clusters in the value space of these two attributes

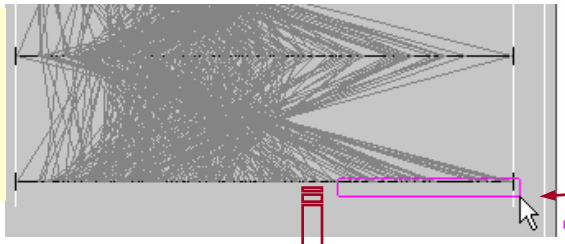




The districts fit in the left half of the histogram, mostly in bars 1 and 4

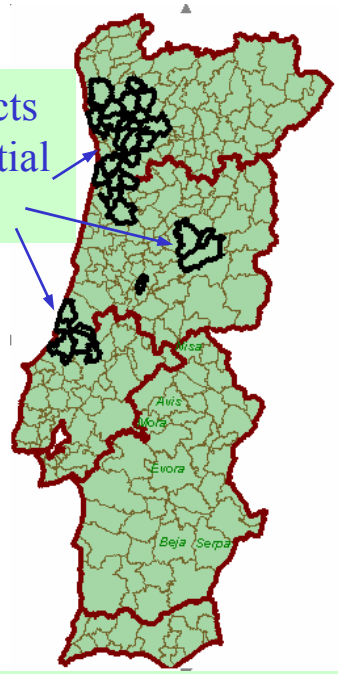
Using Display Linking (2)

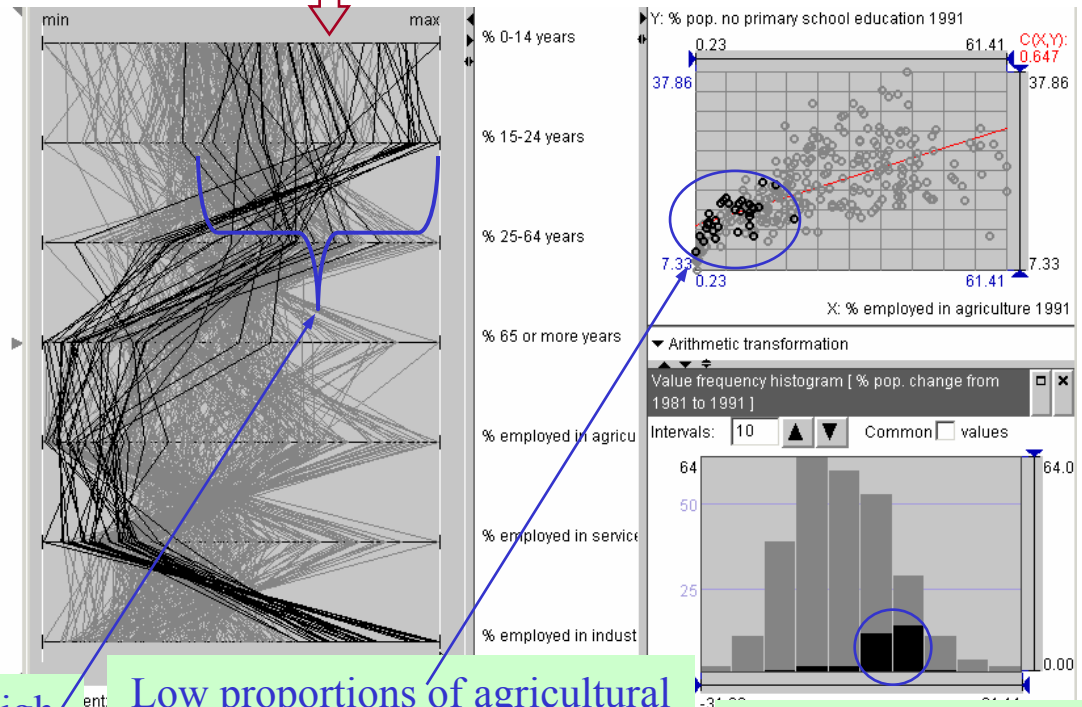
Let us look at the districts with high % employed in industry:



Select high values by drawing a frame

The districts form 3 spatial clusters





The districts have average or high proportions of children and young

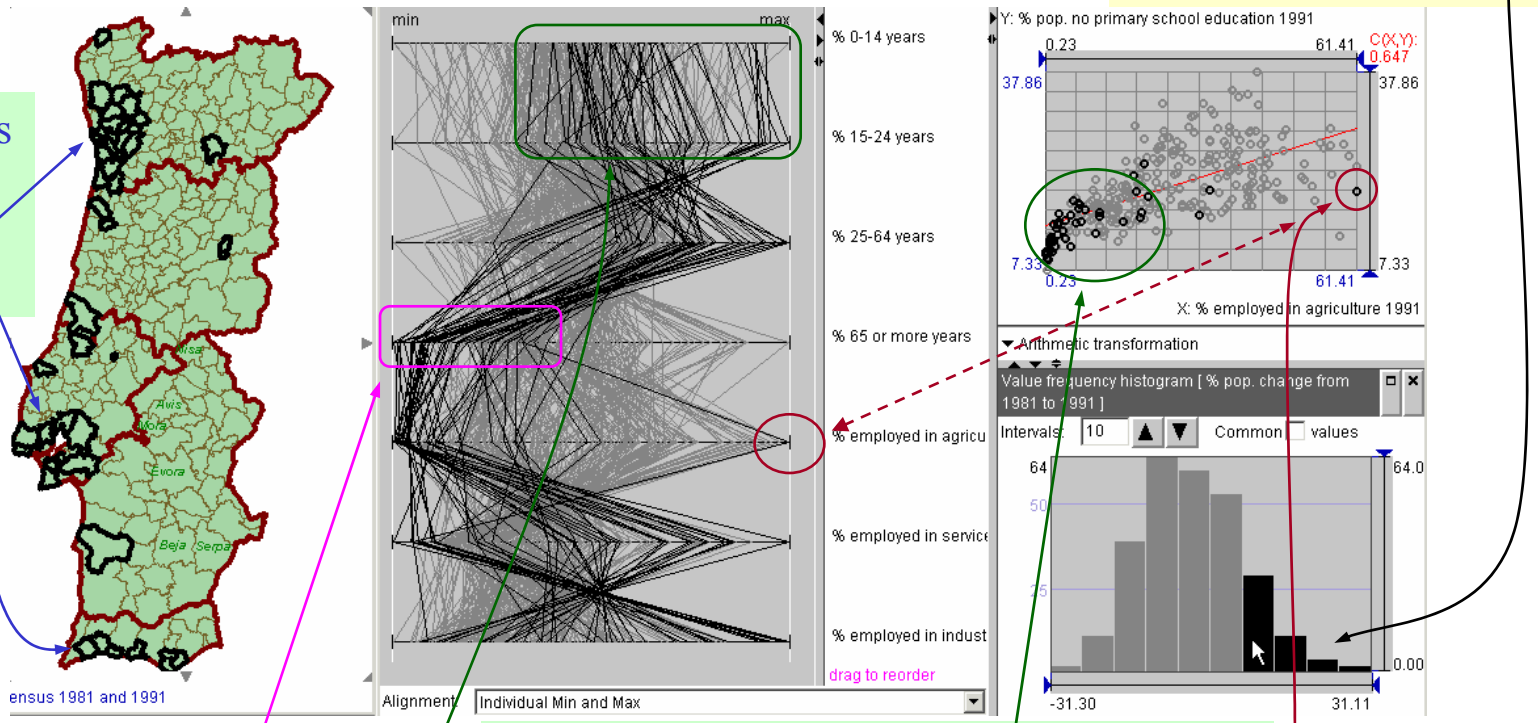
Low proportions of agricultural workers and people without primary school education

Population change: mostly between –0.1% and 12.4%

Using Display Linking (3)

Let us look at the districts with the highest population growth:

Click on the rightmost bars in the histogram



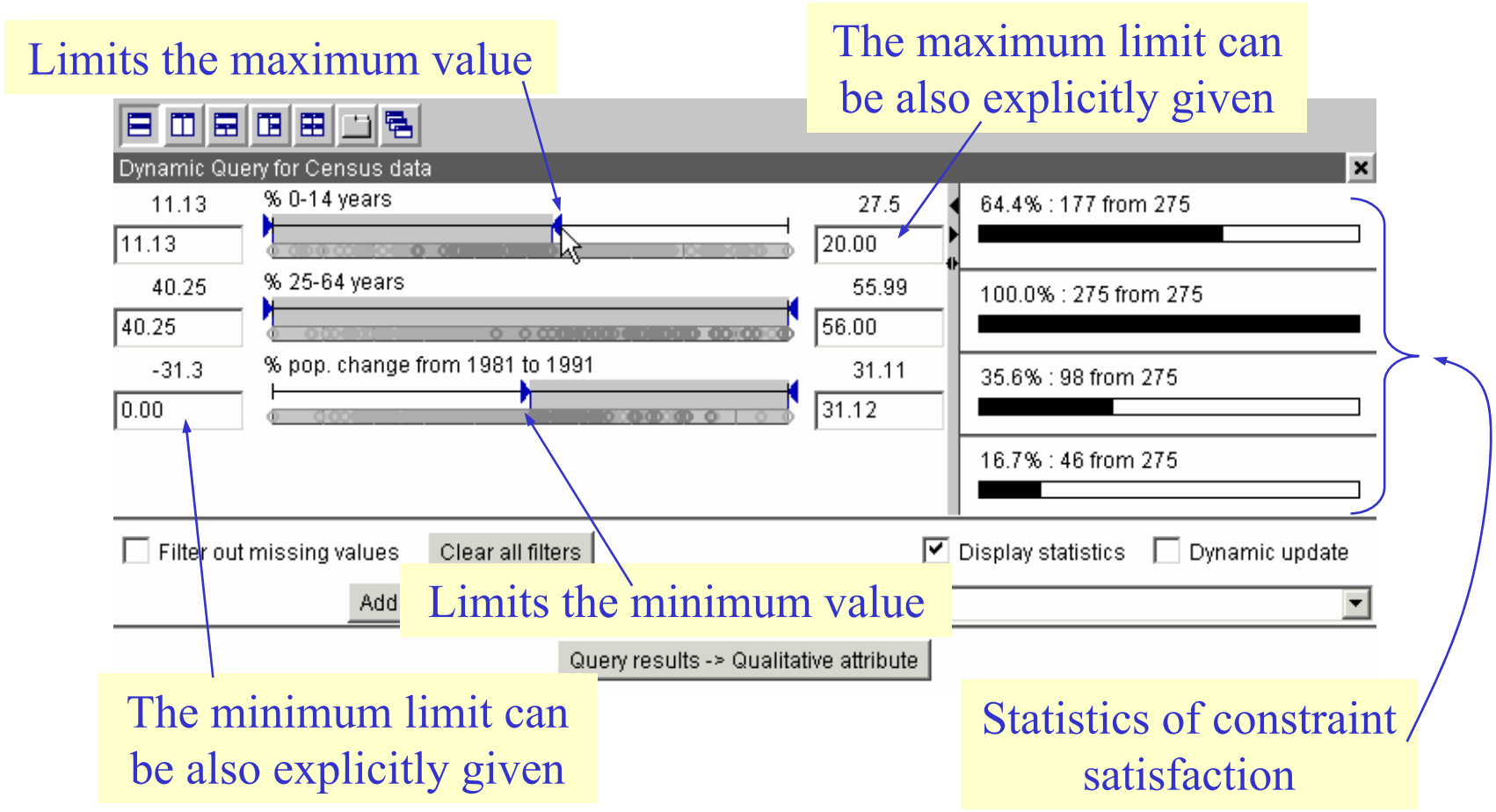
The districts form some spatial clusters

The districts have average or high proportions of children and young and low proportions of old people

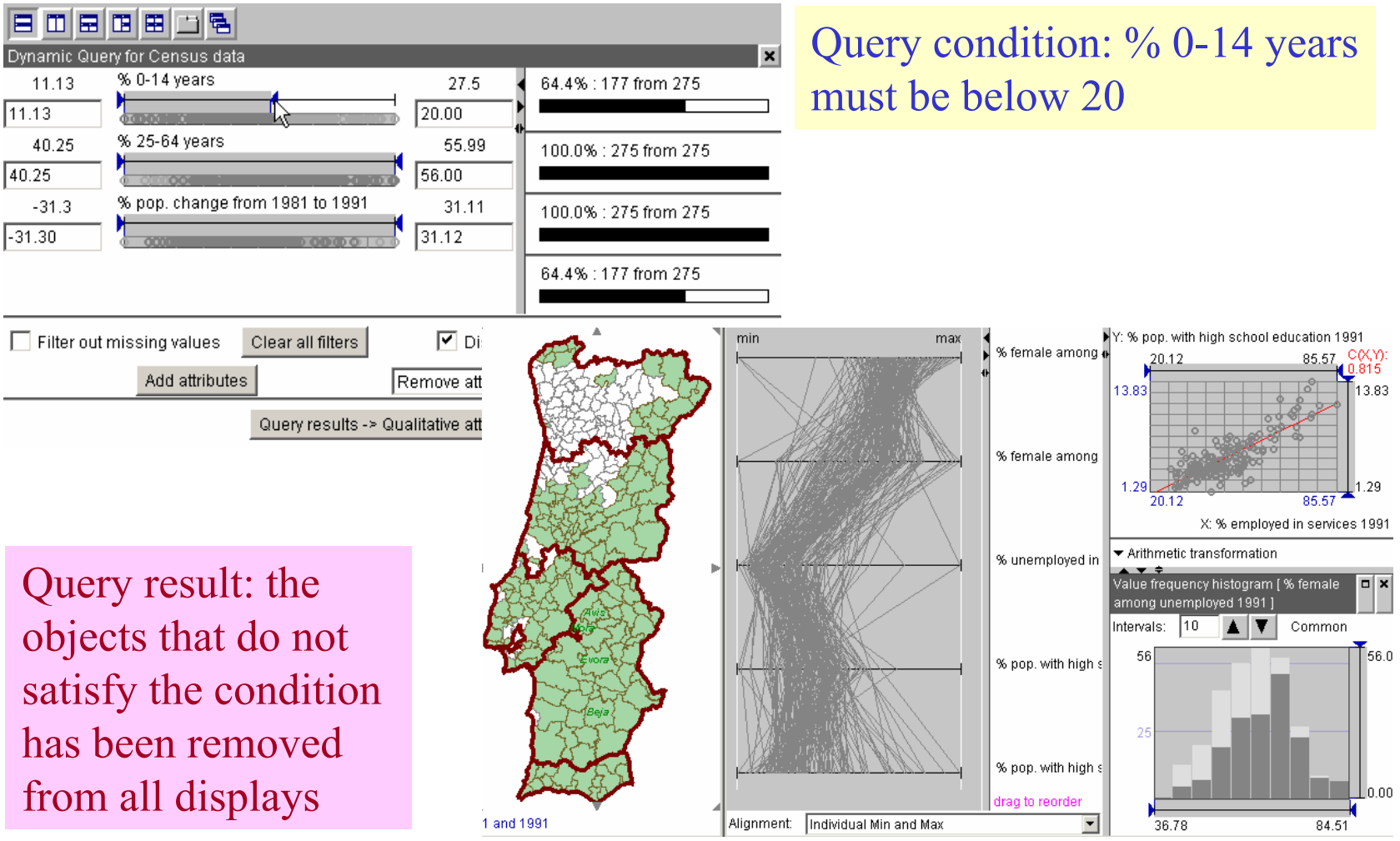
The proportions of agricultural workers and people without primary school education are mostly low, but there is an outlier

Dynamic Query

Dynamic query allows us to set constraints on attribute values



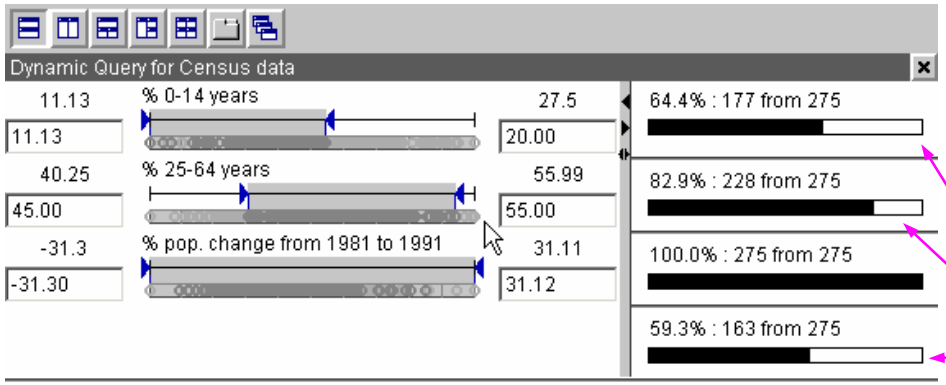
Dynamic Query in Action (1)



Query condition: % 0-14 years must be below 20

Query result: the objects that do not satisfy the condition has been removed from all displays

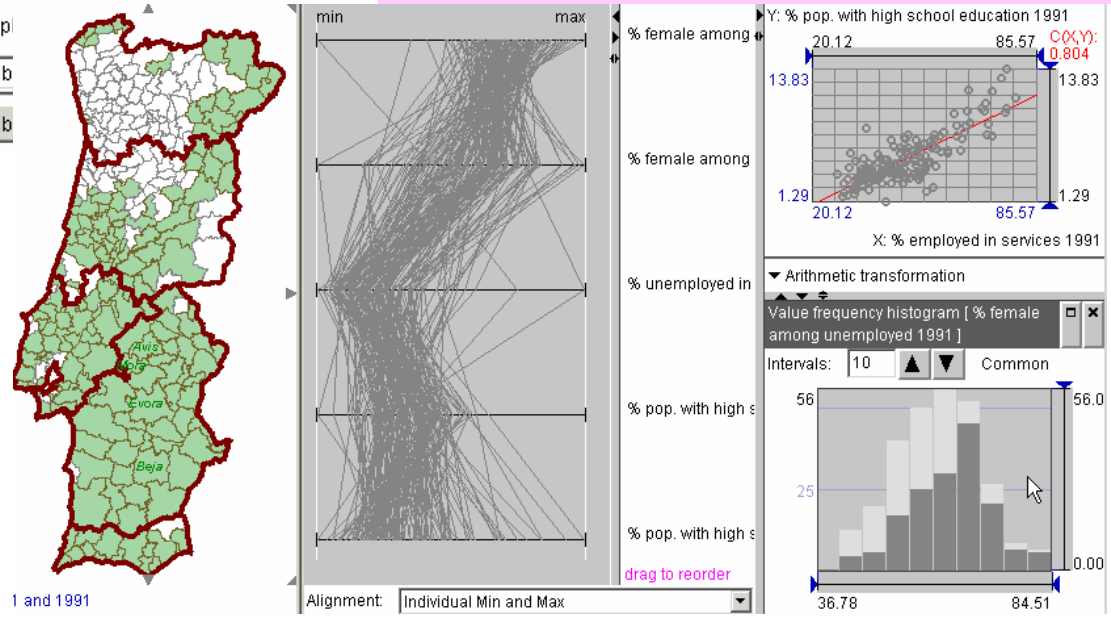
Dynamic Query in Action (2)



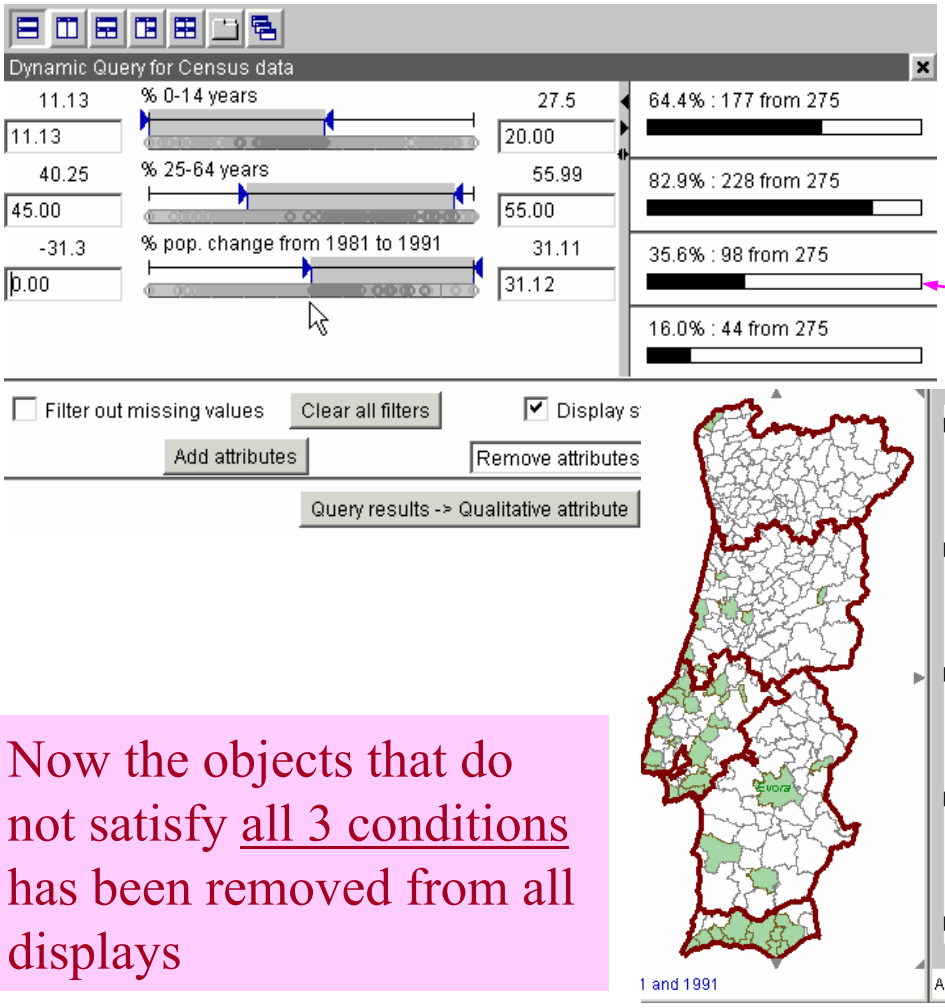
A second query condition added: % 25-64 years must be between 45 and 65

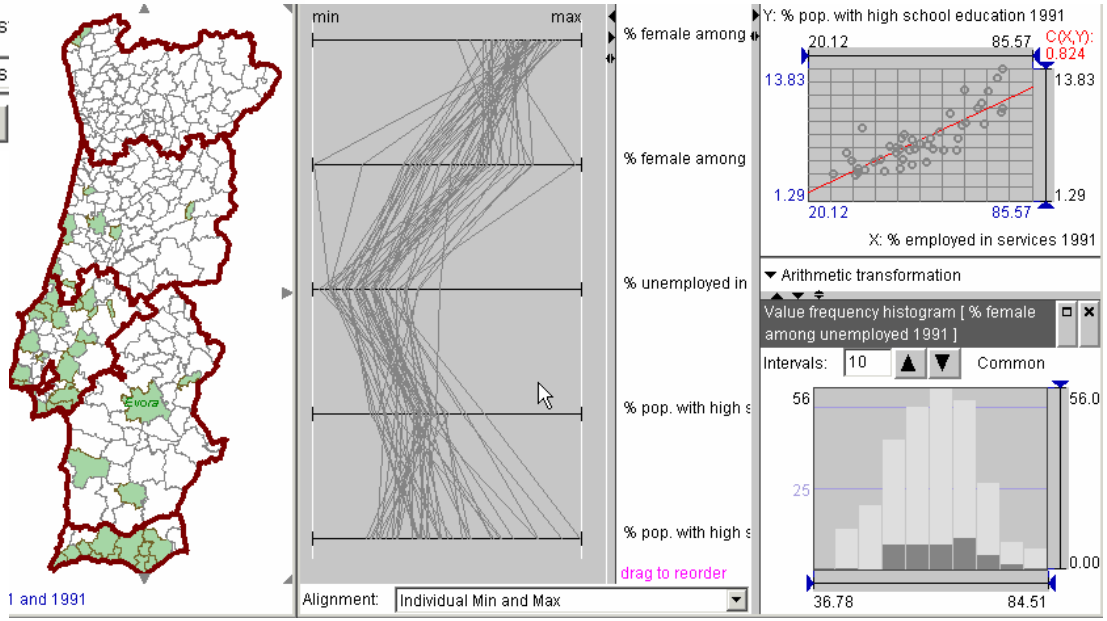
177 objects (64%) satisfy the 1st condition
228 objects (83%) satisfy the 2nd condition
163 objects (59%) satisfy both conditions

Query result changed: the objects that do not satisfy <u>both conditions</u> has been removed from all displays



Dynamic Query in Action (3)





One more query condition added: % pop. change from 1981 to 1991 must be positive

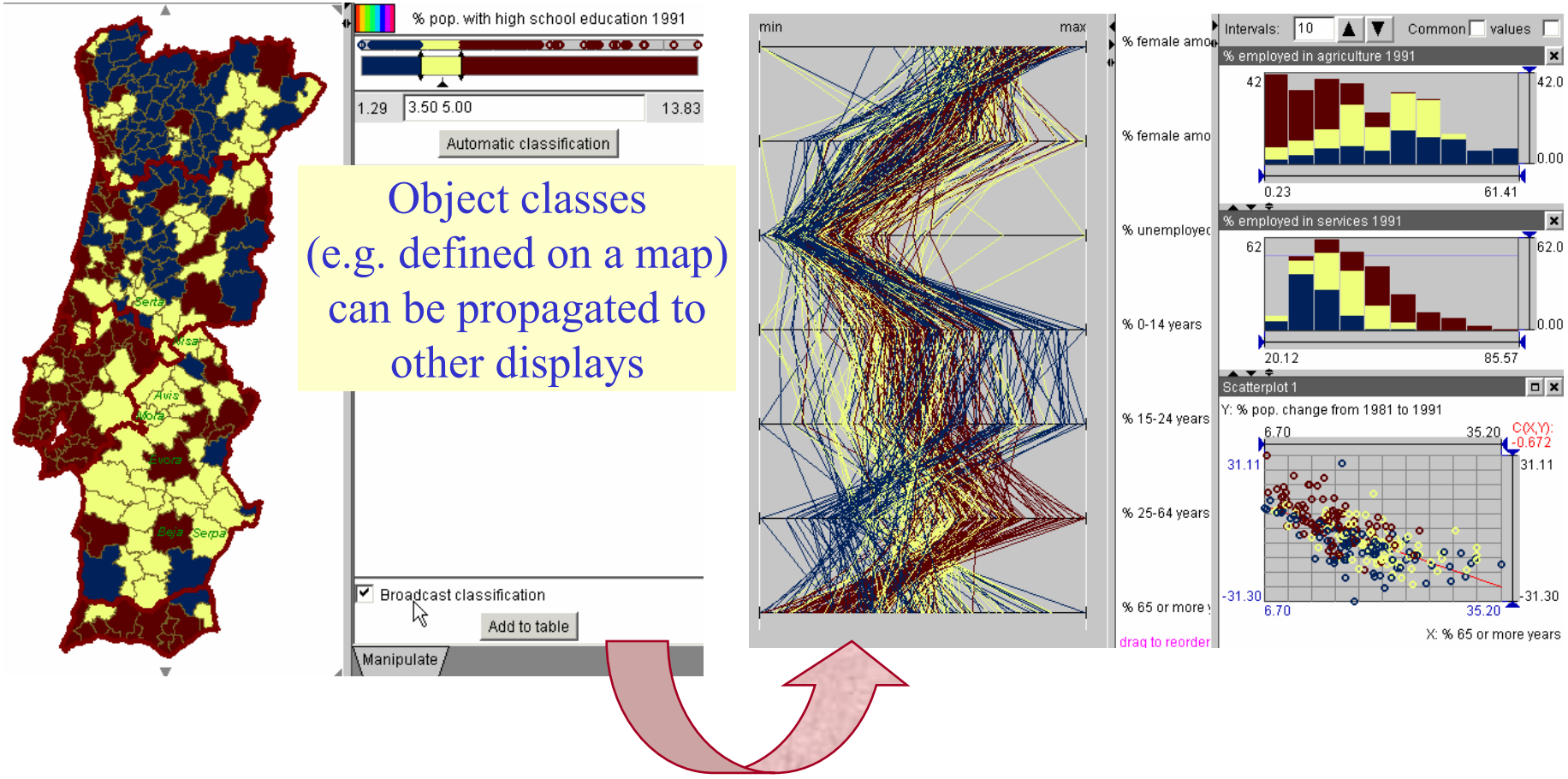
…

98 objects (36%) satisfy the 3rd condition

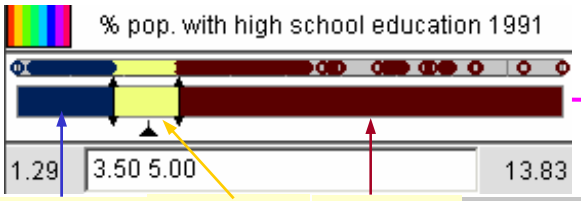
44 objects (16%) satisfy all 3 conditions

Now the objects that do not satisfy <u>all 3 conditions</u> has been removed from all displays

Propagation of Object Classes



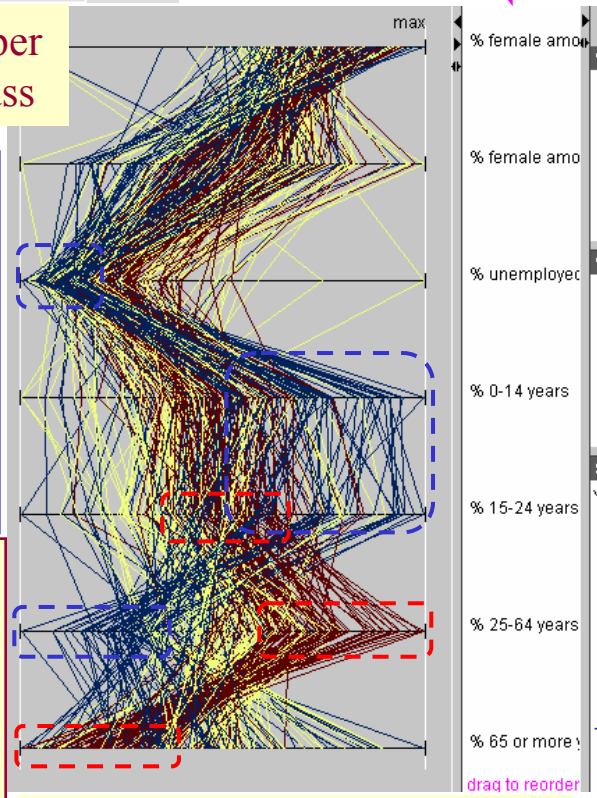
Propagation of Object Classes: Use Example



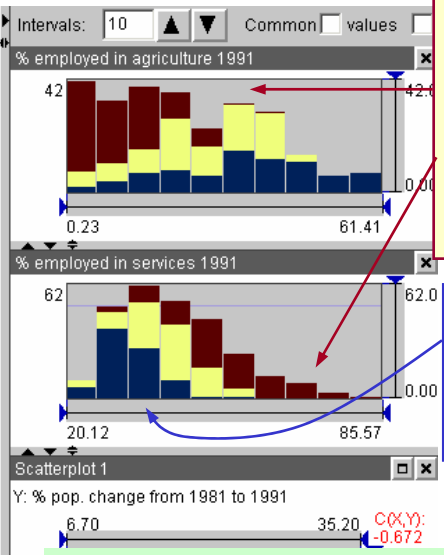
lower class | middle class | upper class

Lower class prevails among districts with low % unemployed, average or high % 0-14 years and % 15-24 years, and low % 25-64 years

Upper class prevails where % 25-64 years is high and % 65 or more years is low. It occupies mostly the middle part of the axis % 15-24 years.



Characteristics of the middle class are highly variant



Upper class co-occurs with low % employed in agriculture and average to high % employed in services

Lower class co-occurs with low % employed in services

In districts with high % employed in agriculture the proportion of people with high school education is low (mostly below 3.5%). For % employed in services we see the opposite relationship

Table View and Table Lens (1)

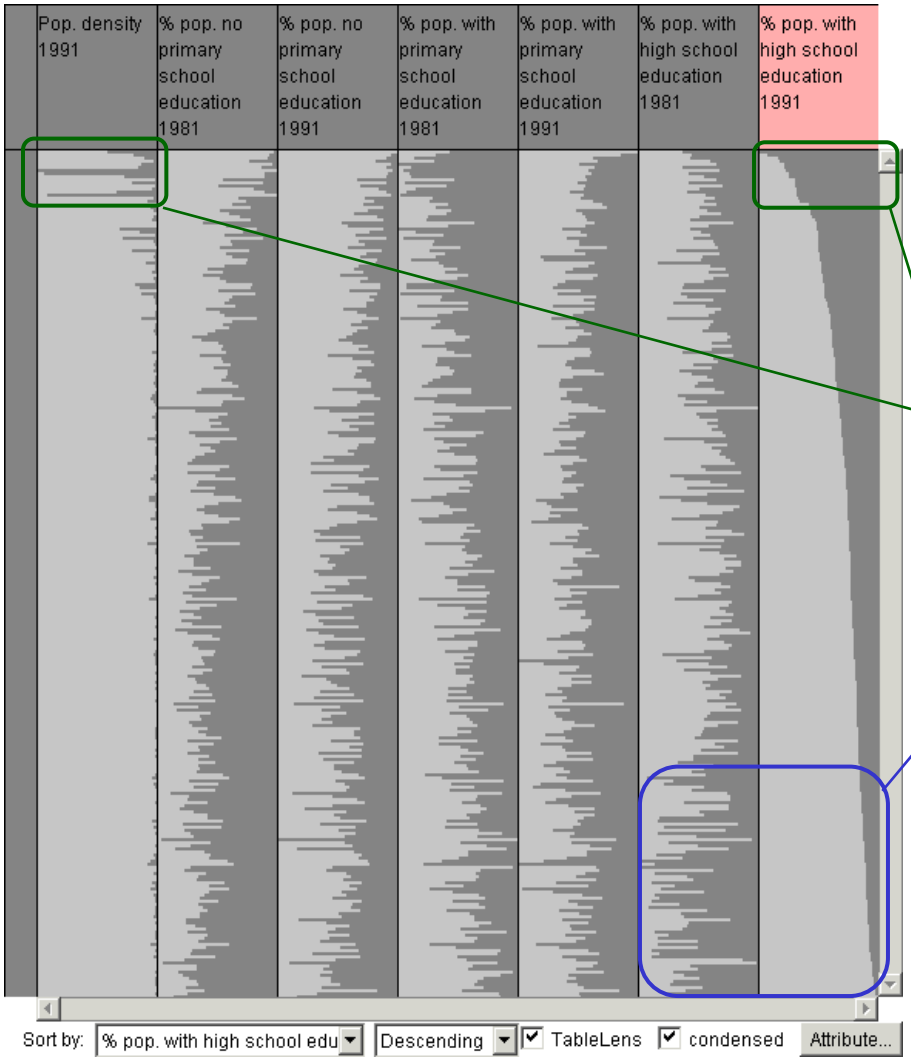
Table cell shading shows the relative position of the values between the minimum and maximum values of the respective attributes

Click for sorting

identifiers	Pop. density 1981	Pop. density 1991	% pop. no primary school education 1981	% pop. no primary school education 1991	% pop. with primary school education 1981	% pop. with primary school education 1991	% pop. with high school education 1981	% pop. with high school education 1991
Lisboa	9636.423	7912.43	37.26	8.51	39.49	25.45	7.675	11.11
Porto	7858.089	7260.49	37.32	8.83	39.97	26.22	8.756	9.84
Amadora	6895.481	7454.641	40.41	9.68	38.66	25.18	9.941	10.80
Oeiras	3257.592	3301.527	37.89	7.33	32.80	18.69	10.444	13.83
Barreiro	2796.012	2723.485	40.64	10.01	39.09	28.16	10.089	9.85
Matosinhos	2190.979	2434.703	42.45	11.01	40.12	29.08	9.857	7.78
Almada	2110.581	2169.072	40.58	10.26	38.49	25.99	8.813	10.66
Sao Joao da Madeira	2027.62	2275.216	41.12	10.21	39.62	28.59	9.658	7.68
Espinho	1513.025	1631.933	44.68	11.37	35.82	27.50	10.919	7.56
Cascais	1457.75	1579.276	39.37	8.92	34.97	21.13	9.718	12.95
Loures	1419.716	1654.349	40.84	10.29	39.32	27.12	9.950	10.09
Vila Nova de Gaia	1324.968	1455.128	43.66	11.61	39.18	29.86	10.352	7.07
Gondomar	981.172	1074.426	43.20	12.31	39.93	29.72	10.493	6.93
Maia	975.854	1112.915	44.45	11.71	39.08	30.47	10.406	7.07
Moita	966.594	1181.663	46.27	14.50	37.64	26.61	9.123	7.71
Seixal	952.844	1249.3	41.18	10.15	38.33	24.60	10.375	11.15
Valongo	880.038	1016.194	42.80	11.90	39.94	30.70	11.492	6.40
Entroncamento	874.161	1038.394	38.51	9.10	40.18	26.73	7.952	11.22

Sort by: Pop. density 1981 | Descending | TableLens | condensed | Attribute...

Table View and Table Lens (2)

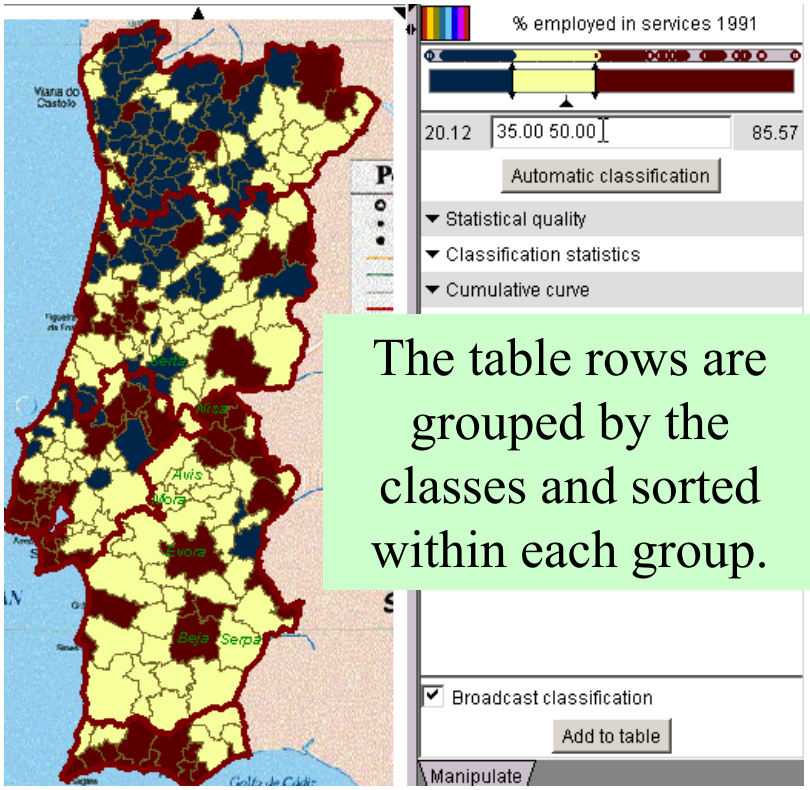


The same information can be represented in a "condensed" form. We do not see the details about particular objects but get an overall impression about value variation and relationships between attributes.

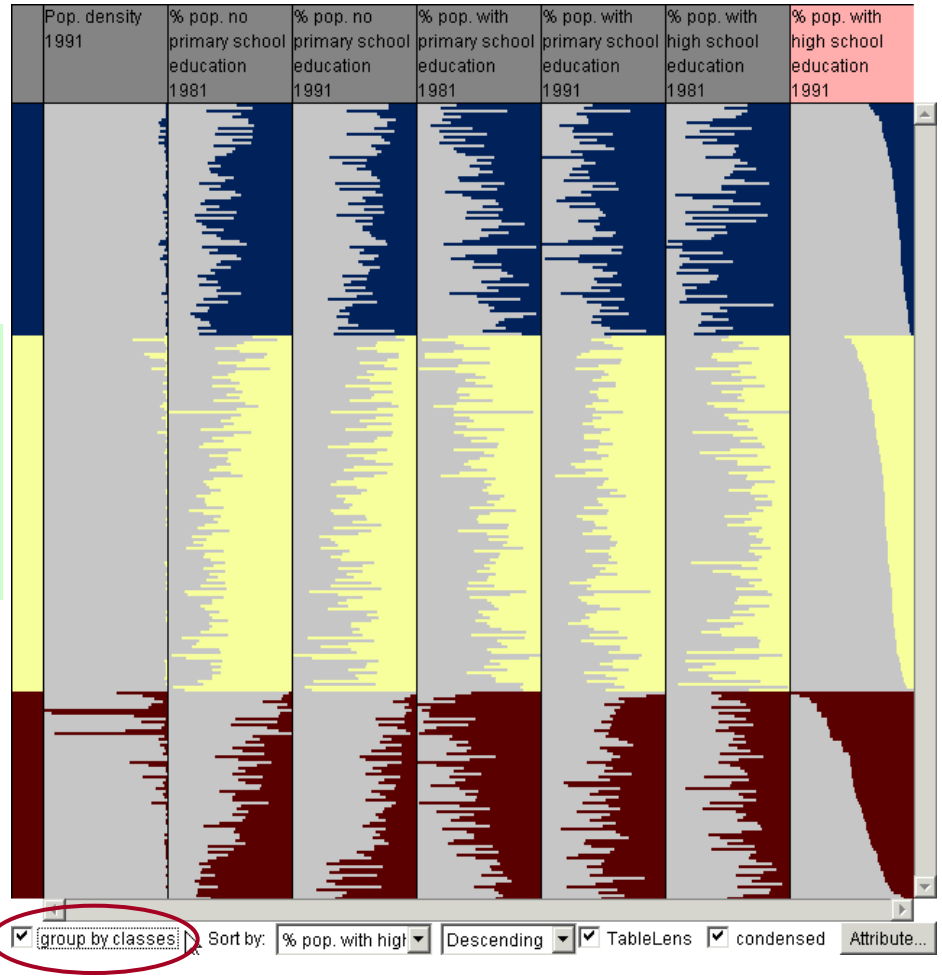
High proportions of people having high school education often co-occur with high population density

Surprisingly, the districts with the lowest proportions of people having high school education in 1991 had much higher proportion of such people in 1981

Class Propagation to Table View (1)

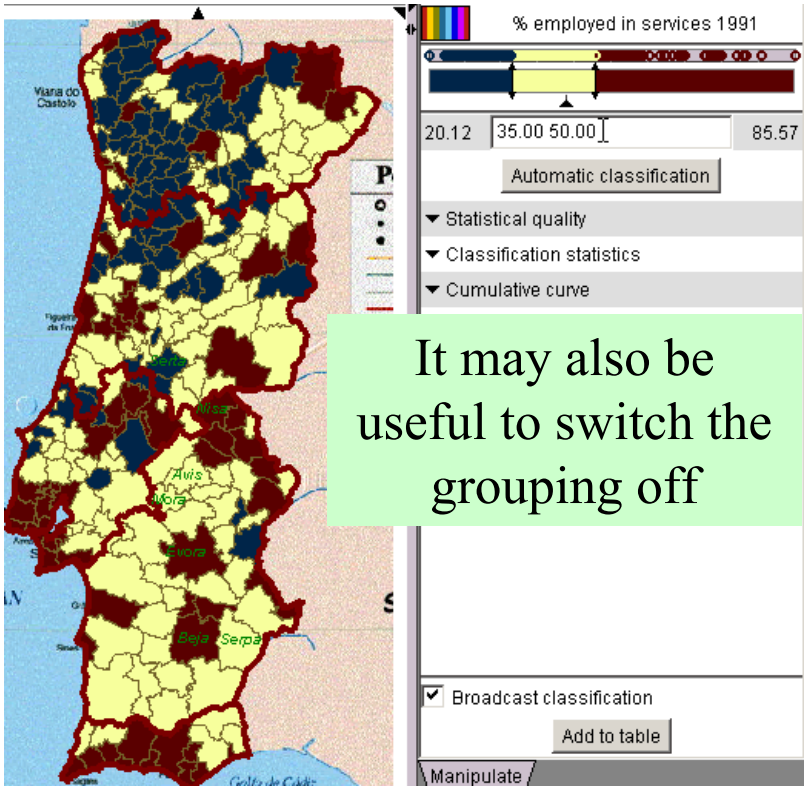


The table rows are grouped by the classes and sorted within each group.

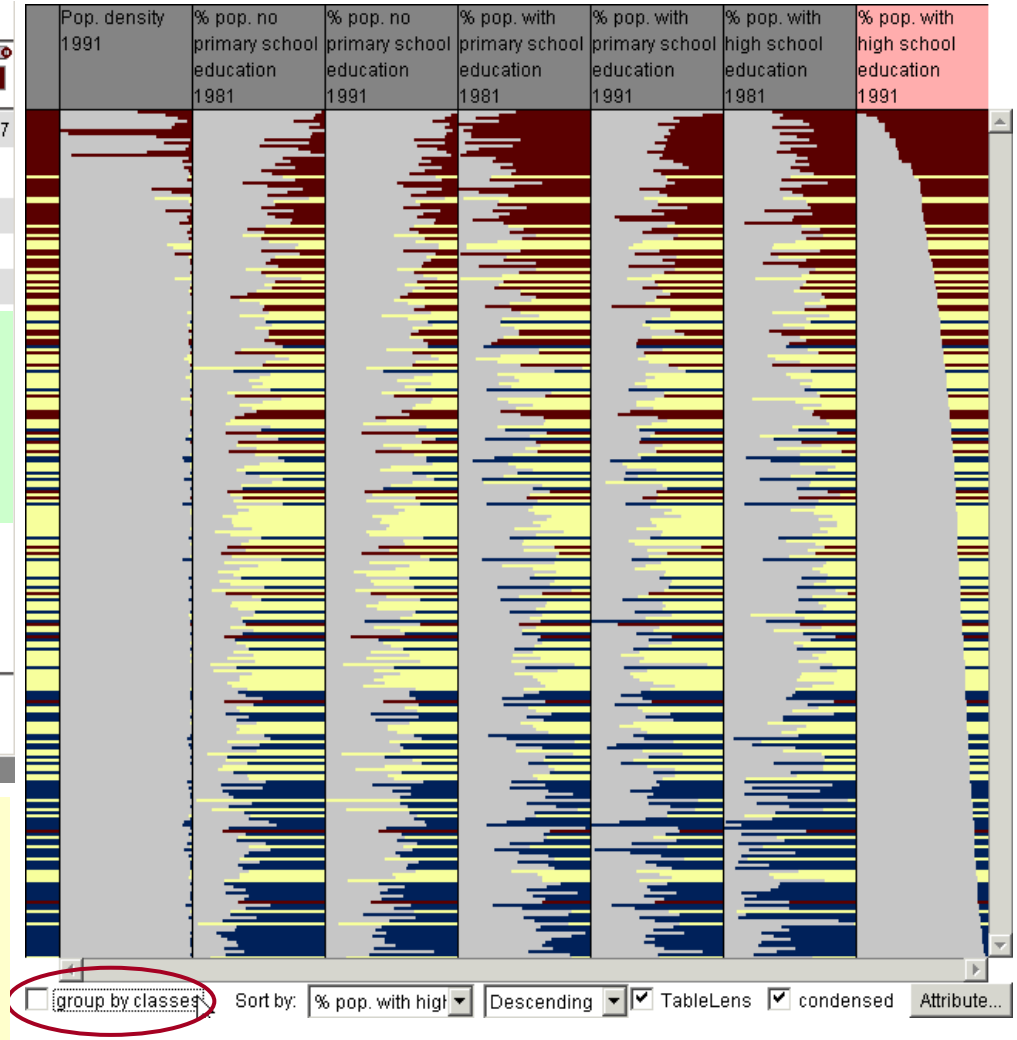


These linked views show us, for example, that the general educational level tends to be higher in districts with high proportion of people employed in services

Class Propagation to Table View (2)



It may also be useful to switch the grouping off

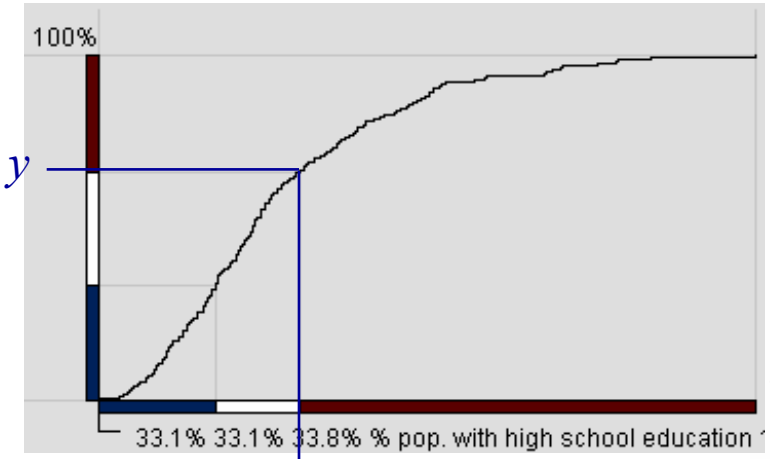


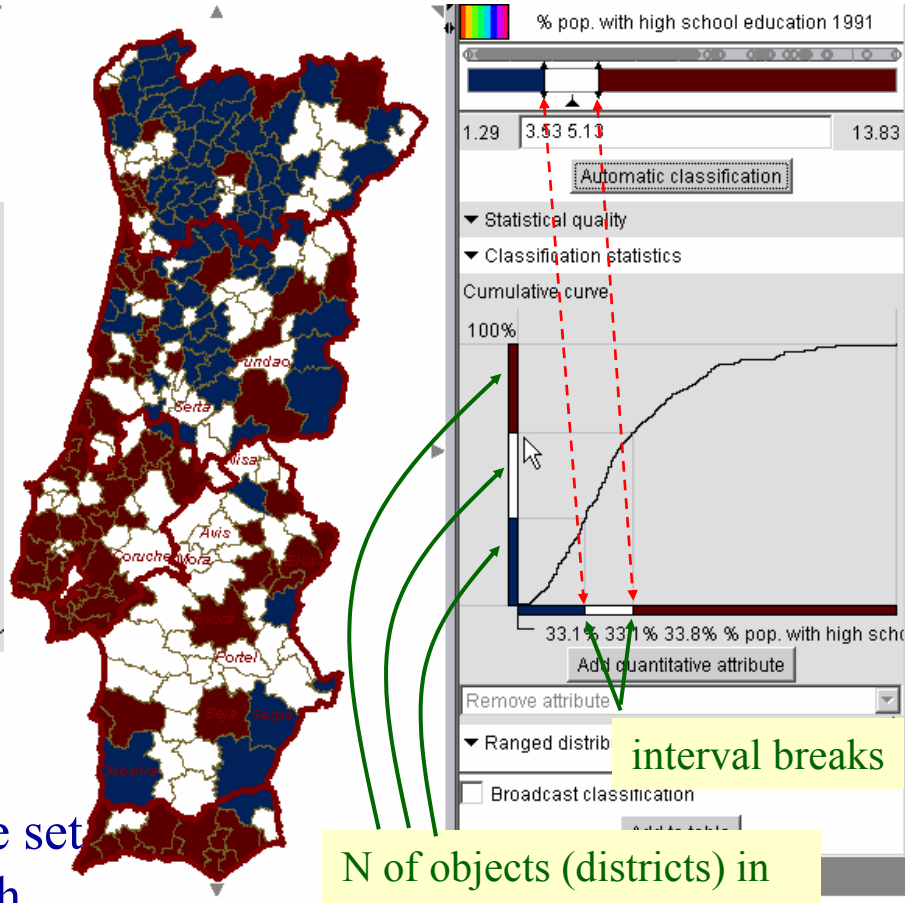
We see that the red rows occur mostly at the top of the table and blue ones at the bottom. Note that the rows are sorted according to % people with high school education in 1991.

Dynamic Classification: Additional Analytical Facilities

Cumulative Frequency Curve

How it is built:





X-axis: attribute's value range

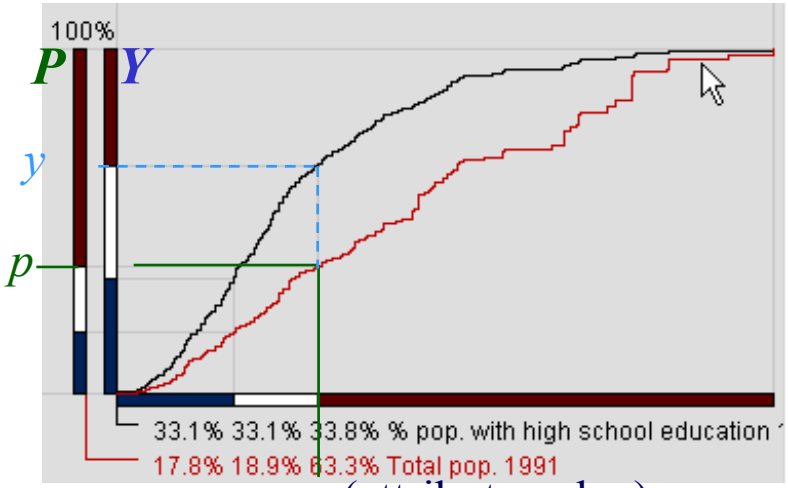
Y-axis: object number or % of the whole set

y is the number of objects (districts) with values less than or equal to *x*

Cumulative Curve: an Extension

Cumulative Population Curve

How it is built:

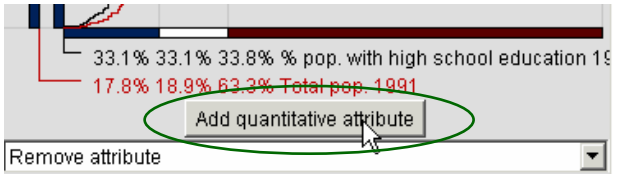


x (attribute value)

X-axis: attribute's value range
Y-axis: object number or % of the whole set
P-axis: population number or % of the whole country's population
p is the **aggregate population** of districts with values less than or equal to *x*

What we can learn about the distribution of the population over the classes:

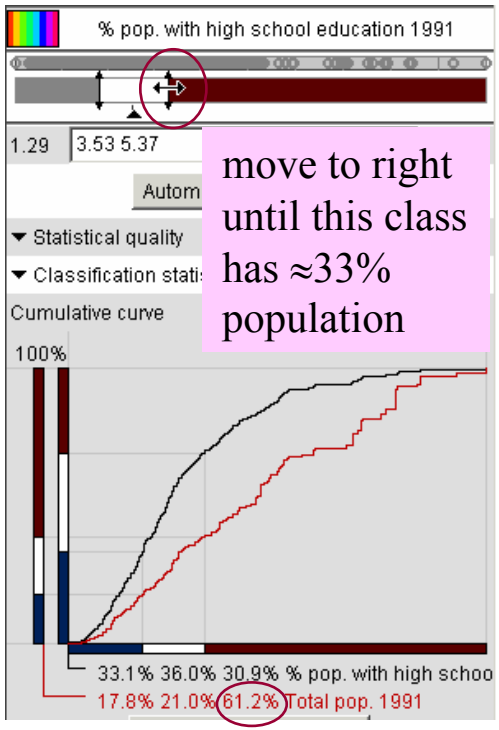
% pop. with high school education (classes)	N of districts	Aggregate population (% of country's total)
up to 3.53	91 (33.1%)	17.8%
over 3.53 up to 5.13	91 (33.1%)	18.9%
over 5.13	93 (33.8%)	63.3%



…and any other quantitative (summable) attribute can be analogously considered

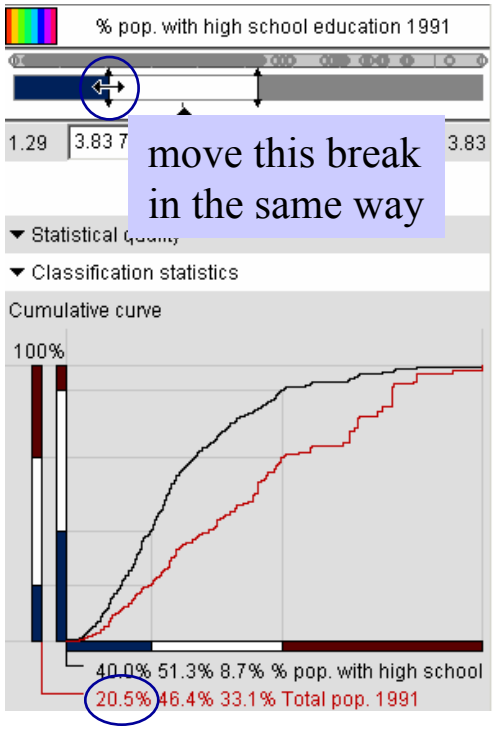
Using Cumulative Curves (1)

Let us move the breaks so that the classes have approx. equal population



move to right until this class has ≈33% population

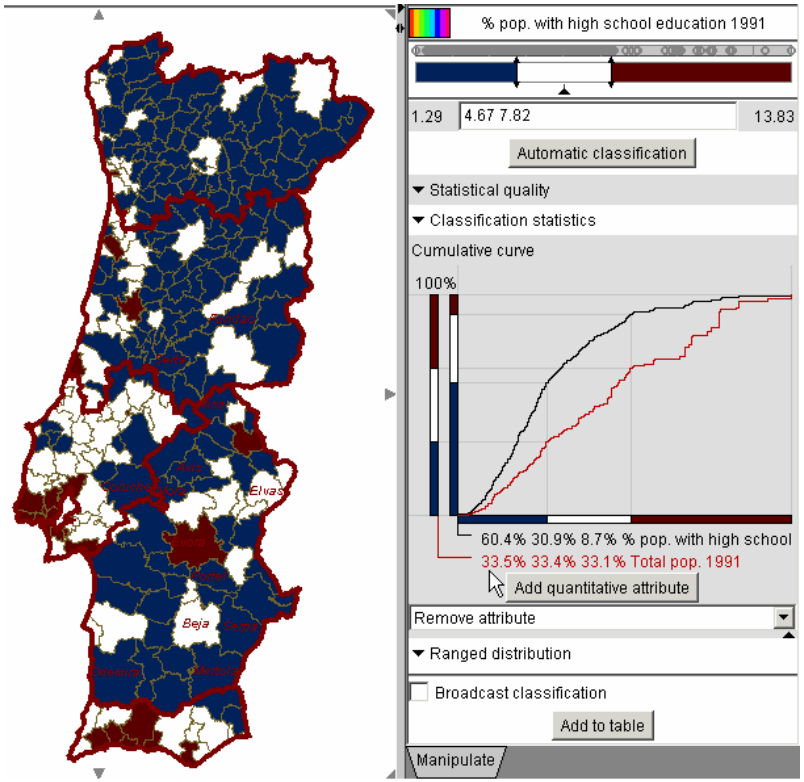
Look here!



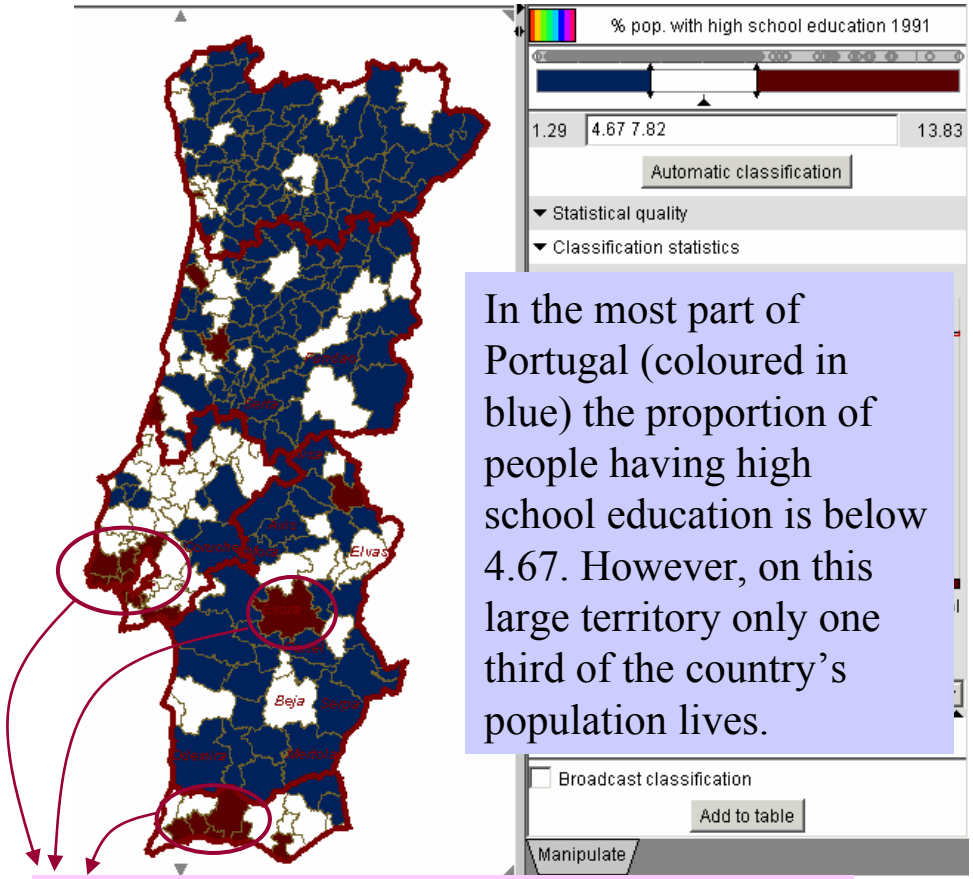
move this break in the same way

Look here!

And here is the result:



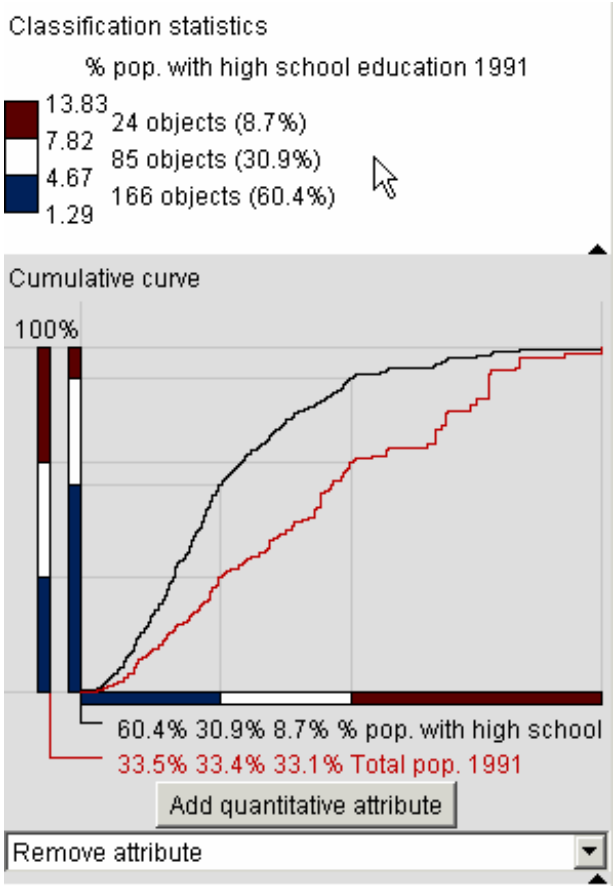
Using Cumulative Curves (2)



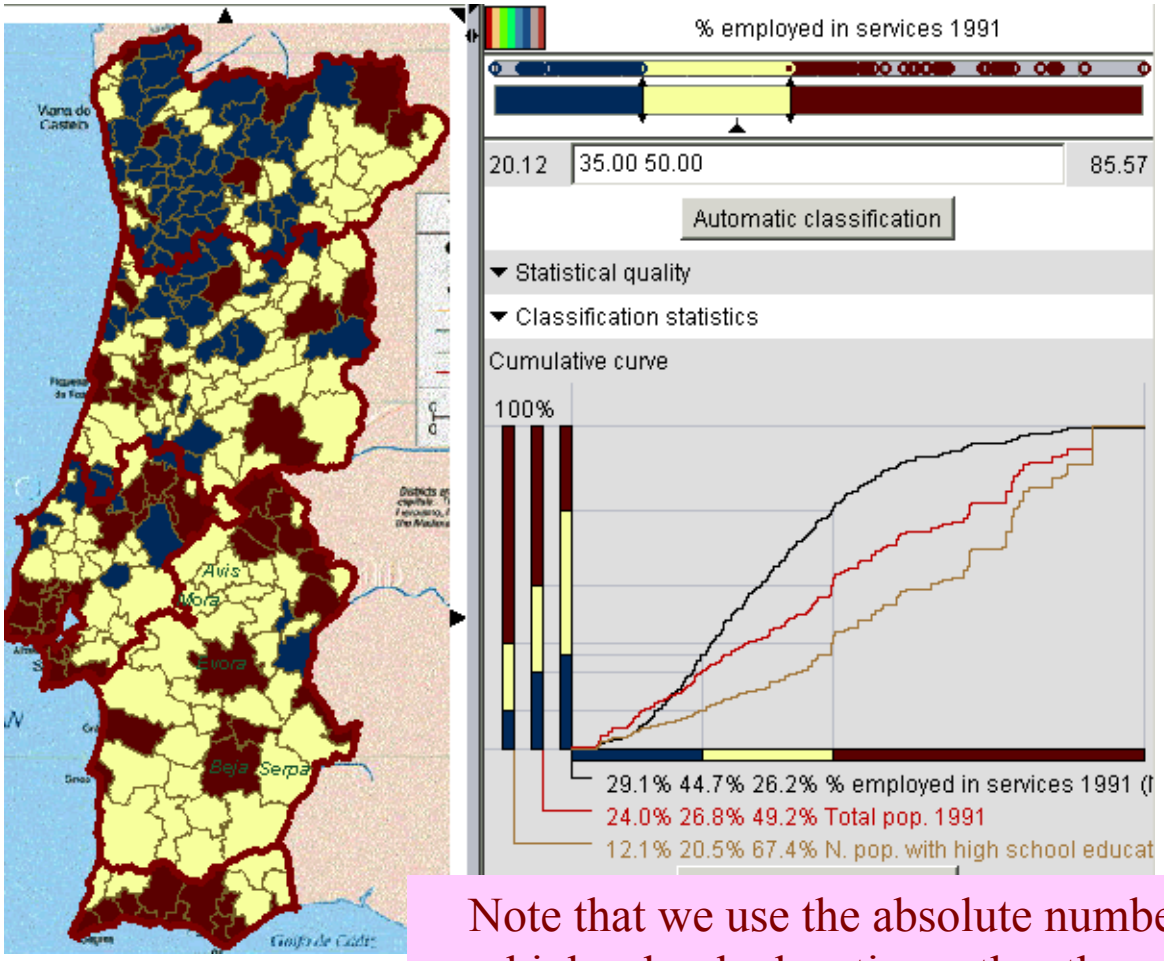
In the most part of Portugal (coloured in blue) the proportion of people having high school education is below 4.67. However, on this large territory only one third of the country's population lives.

In these areas over 7.82% people have high school education. Here lives 33.1% of the total country's population.

Some statistics about the result:



Some More Discoveries with Cumulative Curves



Cumulative curves can also tell us about the relationship between employment in services and educational level.

In 26.2% districts of Portugal 50% or more working people are employed in services. In these districts, almost a half (49.2%) of the total country's population lives and 67.4% of all people having high school education.

Note that we use the absolute number of people with high school education rather than the proportion.

Summary

This lection was supposed to

- introduce the concept of analytical interactive maps
- stress the importance of exploring various aspects of data using multiple views
- demonstrate some types of non-cartographic displays useful in analysis of geodata
- demonstrate various techniques of display linking
- show how to use this in data analysis

See also

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

  section 4.8, pp.428-449