Exploratory Spatial Data Analysis

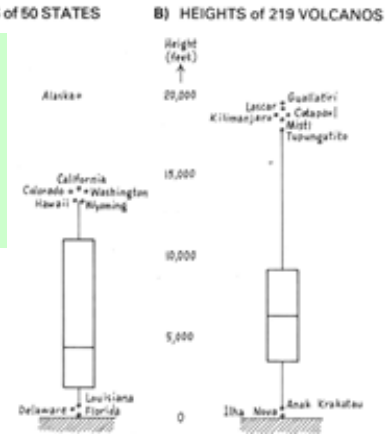
Part I

Interactive Maps

Contents

- Introduction
  - What is Exploratory Data Analysis
  - Spatially referenced data and cartographic visualisation
  - Major techniques for thematic mapping
- Interactive maps
  - Manipulation of unclassified choropleth maps
  - Manipulation of chart maps
  - Dynamic classification and cross-classification

Exploratory Data Analysis

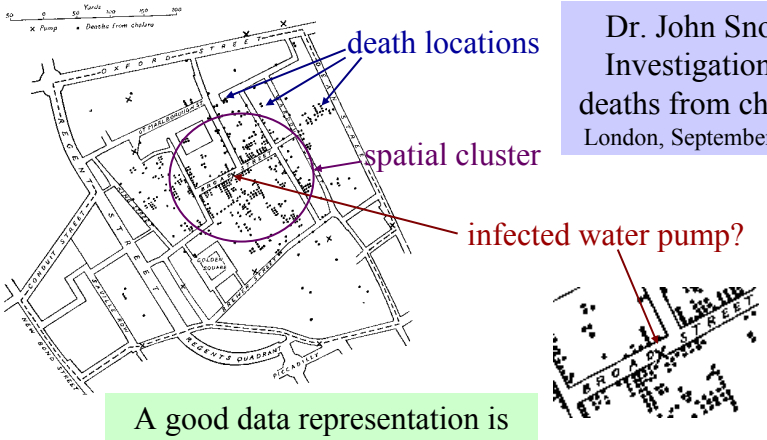


Originator: John W.Tukey
Main idea: represent data so as to facilitate understanding and prompt hypotheses

The greatest value of a picture is when it *forces* us to notice what we never expected to see.

John W. Tukey

Exploratory Spatial Analysis (example)



Dr. John Snow: Investigation of deaths from cholera
London, September 1854

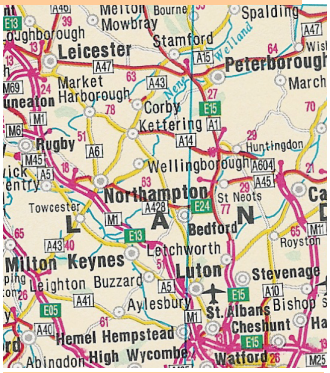
A good data representation is the key to solving the problem

Typical Tasks in Exploratory Spatial Analysis

- Identify spatial patterns
- Identify information relevant to explaining the patterns
- Identify relationships between spatial phenomena

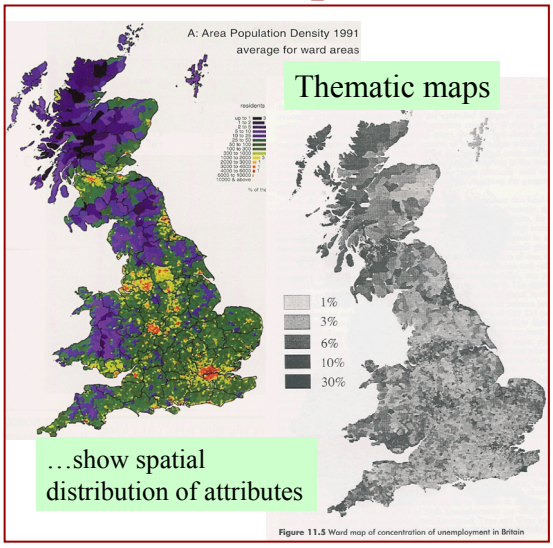
Tools for Exploratory Spatial Analysis: Thematic Maps

General reference maps



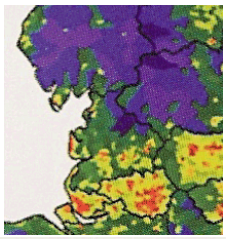
…show locations of objects

Thematic maps

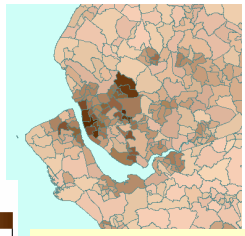


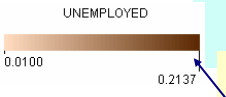
…show spatial distribution of attributes

Some Techniques for Thematic Mapping (1)

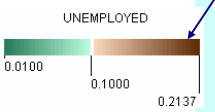


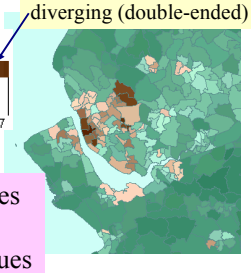
Choropleth maps: enumeration units coloured or shaded to represent different magnitudes of an attribute

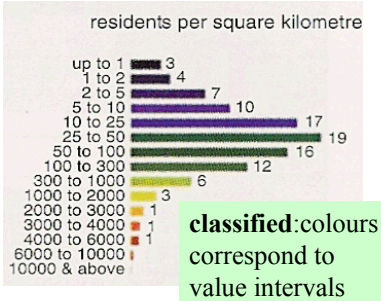




colour scales:
sequential (gradient)
diverging (double-ended)







classified:colours correspond to value intervals

unclassified:degrees of darkness are proportional to values

Some Techniques for Thematic Mapping (2)

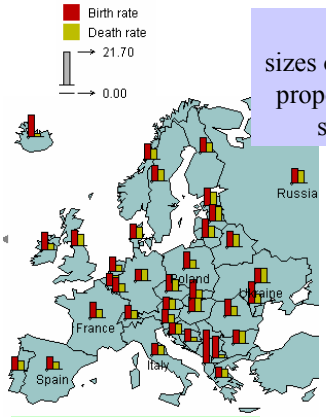
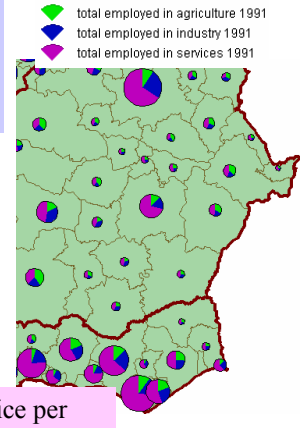


Chart maps: sizes of chart segments are proportional to values of several attributes



Bar charts: one bar per attribute, height proportional to value

Pie charts: one slice per attribute, angle proportional to value; pie size (area) proportional to sum of all

Current Exploratory Tools

High interactivity

Due to *direct manipulation* computer screens will play no less revolutionary role for data exploration than the invention of Cartesian coordinates

W.Cleveland 1993

Enabling multiple complementary views

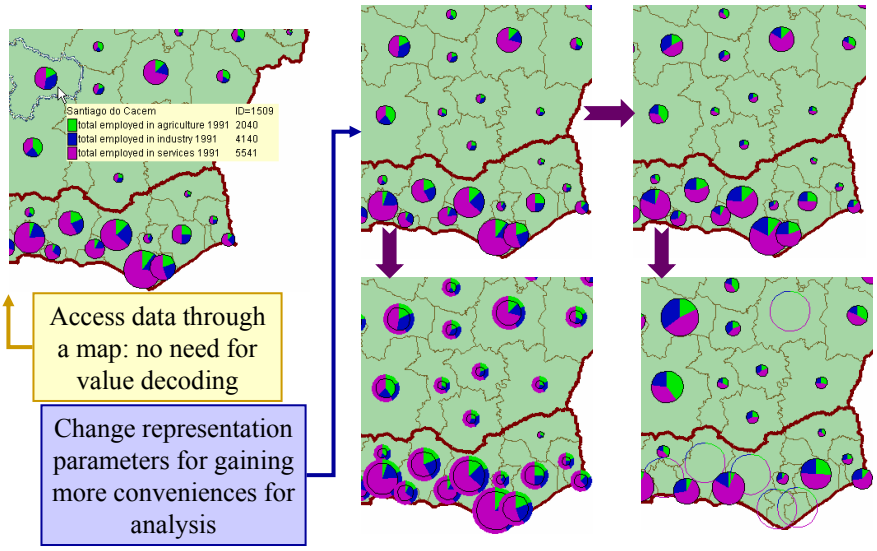
allow the user ... to "see" data from multiple perspectives

A.MacEachren and *M.-J. Kraak* 1997

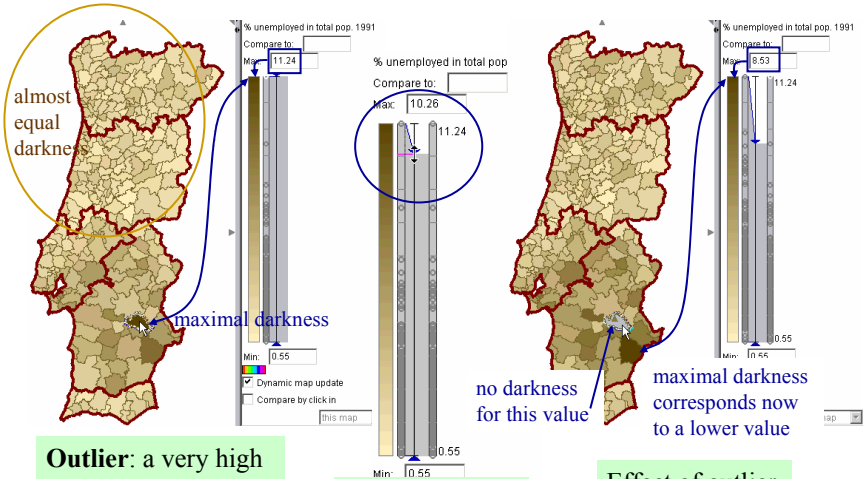
Interactive Maps

- Introduction
  - What is Exploratory Data Analysis
  - Spatially referenced data and cartographic visualisation
  - Major techniques for thematic mapping
- Interactive maps
  - Manipulation of unclassified choropleth maps
  - Manipulation of chart maps
  - Dynamic classification and cross-classification

Types of Map Interactivity



Unclassified Choropleth Maps: Removing Outliers (1)

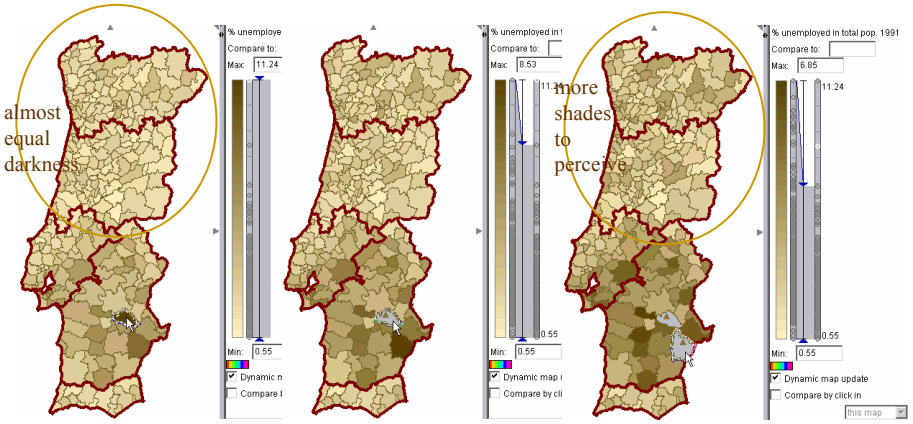


Outlier: a very high (or very low) value, far apart from others

Interactive outlier removal

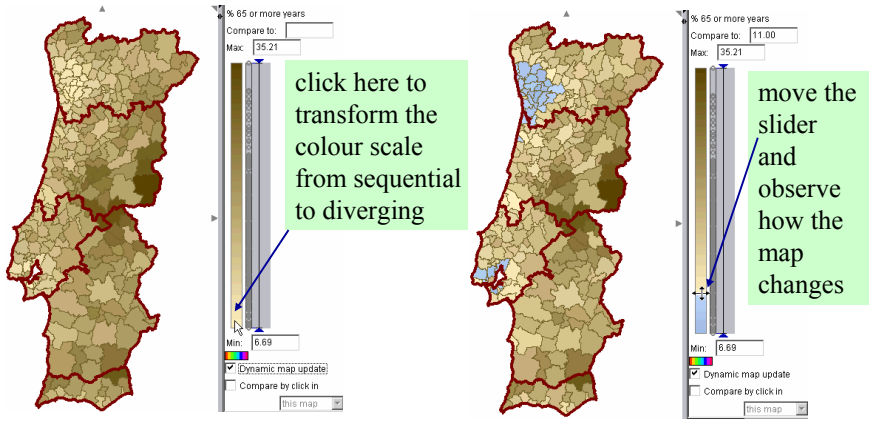
Effect of outlier removal

Unclassified Choropleth Maps: Removing Outliers (2)



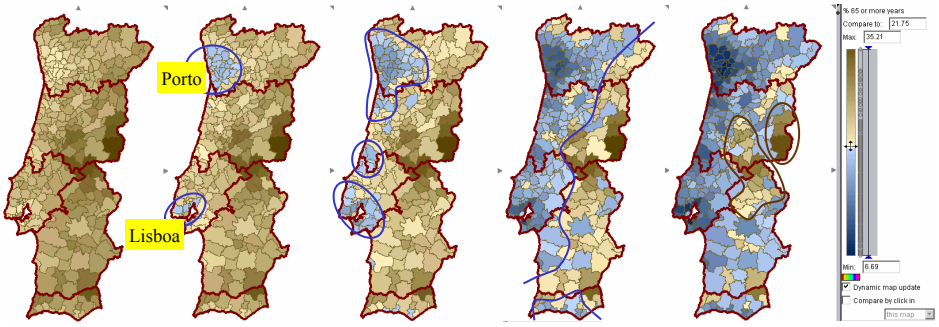
After the removal of two outliers, the differences are better seen

Unclassified Choropleth Maps: Pattern Investigation (1)



Unclassified Choropleth Maps: 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

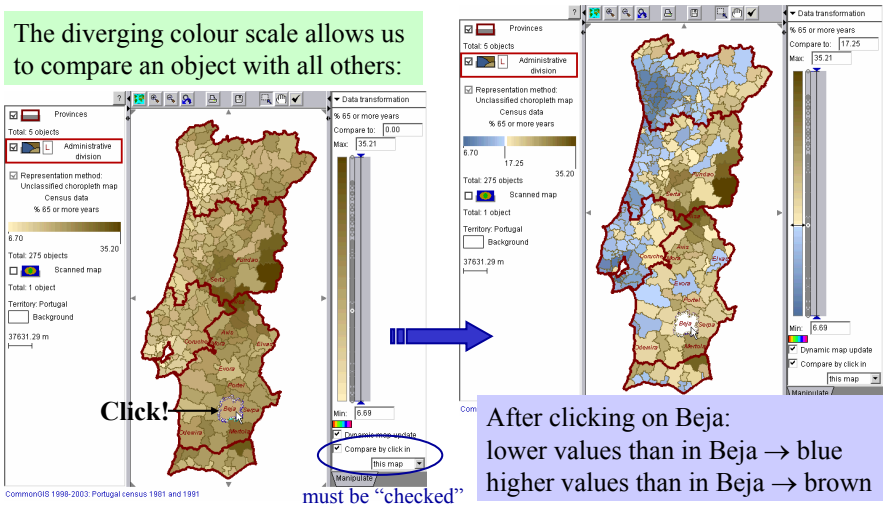
One more cluster of low values

Coast-inland contrast

Clusters of high values in central-east

Unclassified Choropleth Maps: Object Comparison

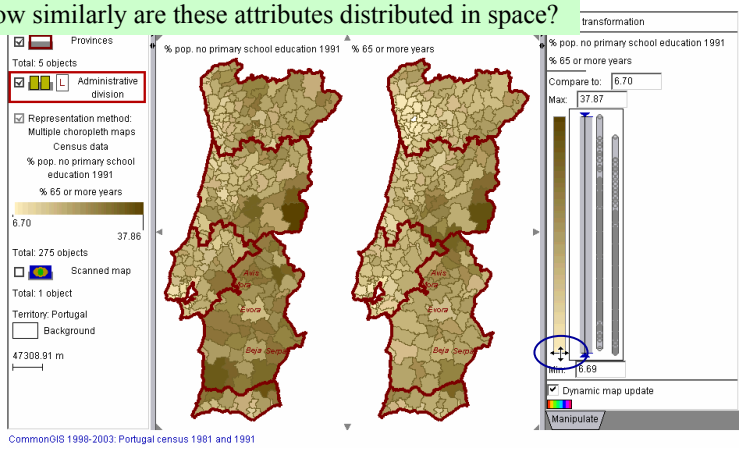
The diverging colour scale allows us to compare an object with all others:



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

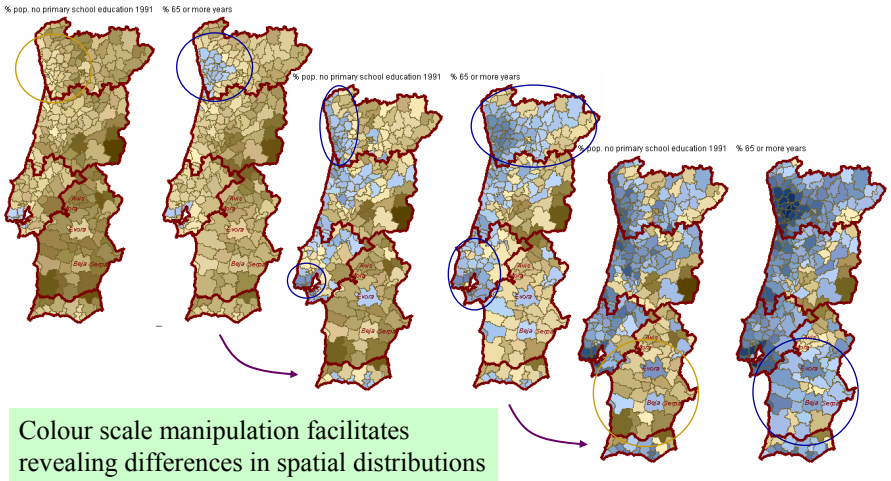
Unclassified Choropleth Maps: Pattern Comparison (1)

How similarly are these attributes distributed in space?



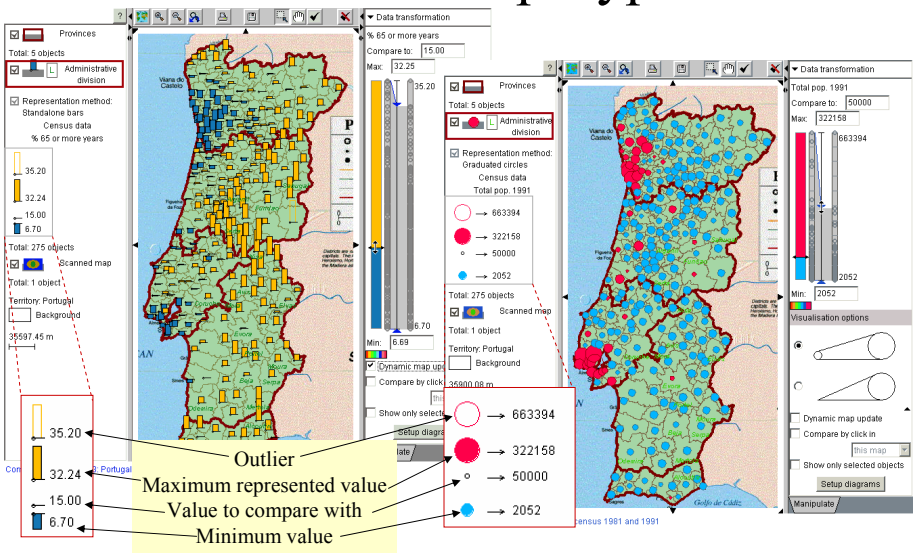
Let us synchronously manipulate the colour scales of the two maps…

Unclassified Choropleth Maps: Pattern Comparison (2)



Colour scale manipulation facilitates revealing differences in spatial distributions

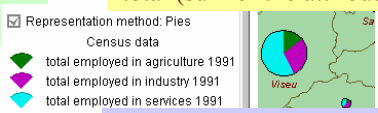
Focusing and Visual Comparison on Other Map Types



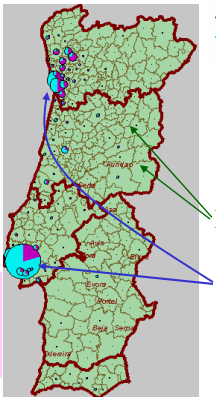
Piechart Map

Applicable to several attributes that together give some meaningful whole

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



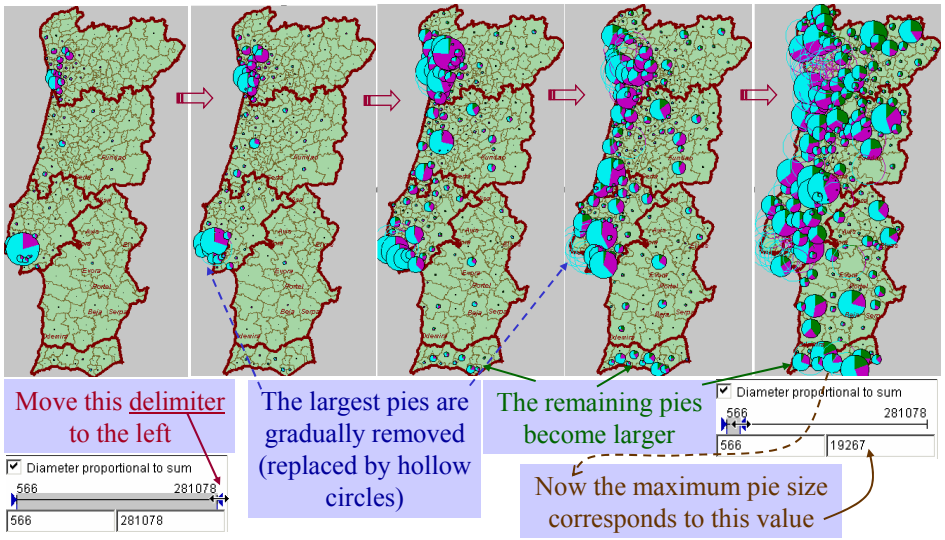
The division into slices shows proportion of each attribute in the total



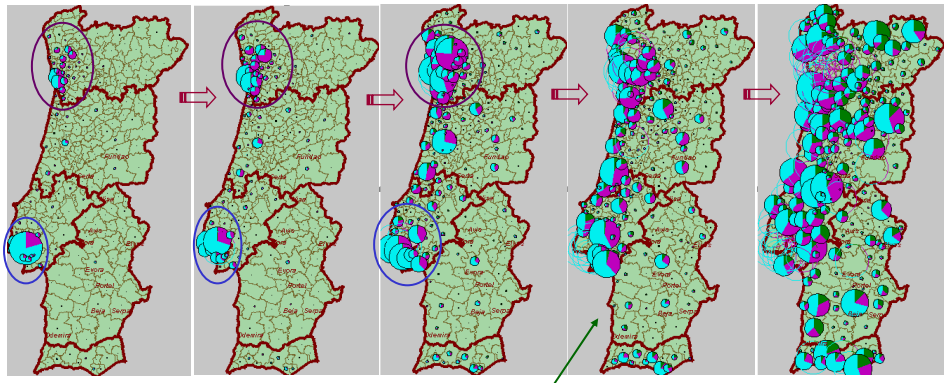
However, the map often looks like this:

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

Piechart Map: Focusing



Focusing and Data Investigation

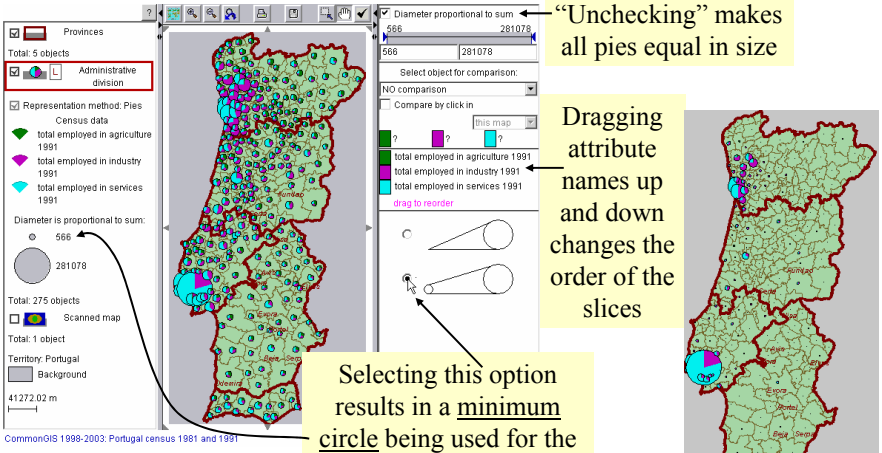


In districts with much population people work in industry (magenta) and services (cyan). Northwest: more industry Centre-west: more services

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

Piechart Map: More Facilities (1)



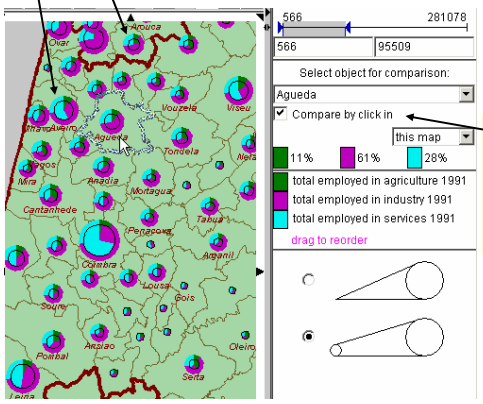
"Unchecking" makes all pies equal in size

Dragging attribute names up and down changes the order of the slices

Selecting this option results in a minimum circle being used for the minimum total sum (compare to the "strictly-proportional" variant)

Piechart Map: More Facilities (2)

In the "visual comparison" mode, each pie is surrounded by an outer circle showing the proportions in a selected reference object

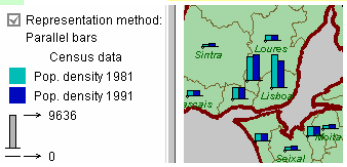


"Checking" this box allows us to compare any district to all others by just clicking on it on the map

Barchart Map

Applicable to several comparable attributes; one bar per attribute

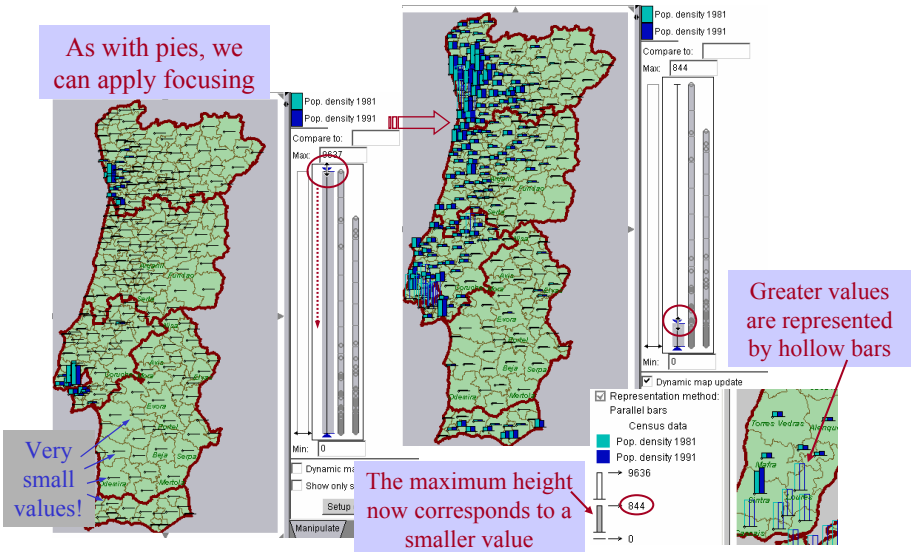
Bar height is proportional to the value of the corresponding attribute



Convenient for local comparisons, e.g. values in different years

Barchart Map: Focusing

As with pies, we can apply focusing



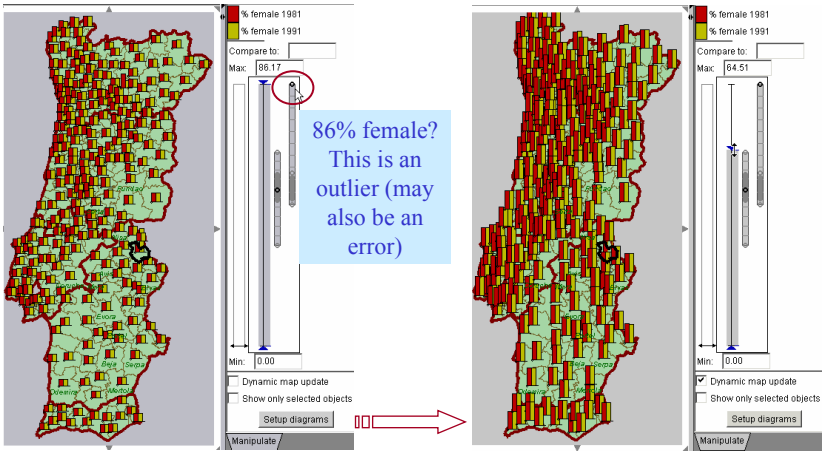
Very small values!

The maximum height now corresponds to a smaller value

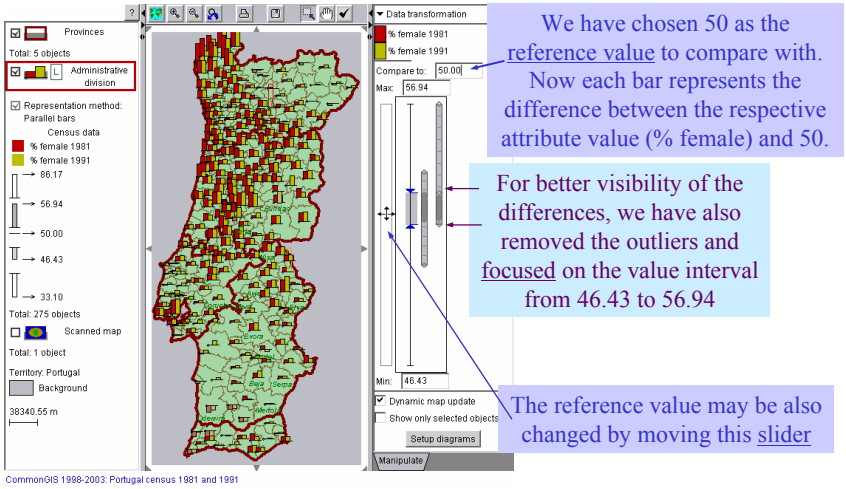
Greater values are represented by hollow bars

Barchart Map: Outlier Removal

The same technique is suitable for removing outliers:



Barchart Map: Comparison to a Number (1)



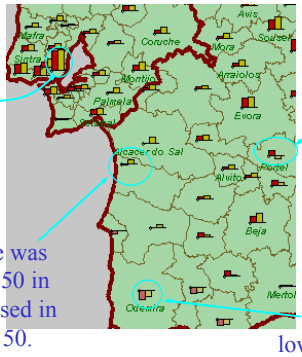
We have chosen 50 as the reference value to compare with. Now each bar represents the difference between the respective attribute value (% female) and 50.

For better visibility of the differences, we have also removed the outliers and focused on the value interval from 46.43 to 56.94

The reference value may be also changed by moving this slider

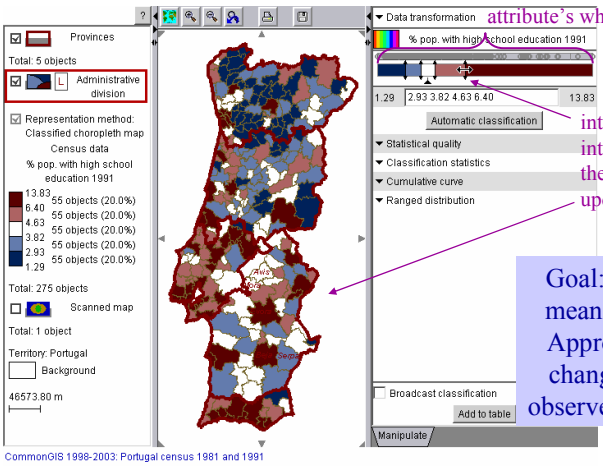
Barchart Map: Comparison to a Number (2)

Comparison of % female in 1981 and 1991 to 50%:



Dynamic Classification

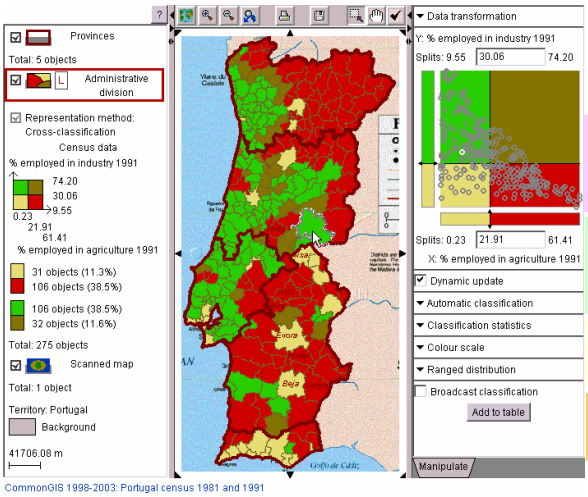
(by values of a single numeric attribute)



Goal: detect interesting and meaningful spatial patterns. Approach: be able to easily change classes and quickly observe the results on the map.

Cross-classification (1)

Classification by values of 2 numeric attributes



The default class breaks are the median values of the attributes

Red: low employment in industry and high employment in agriculture

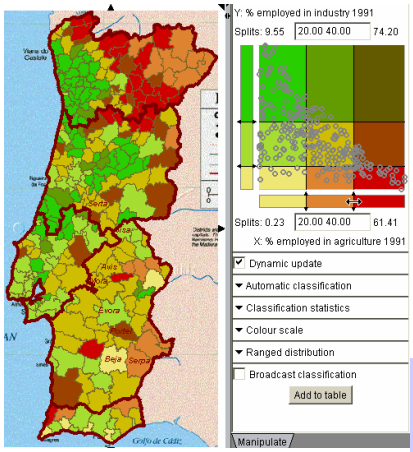
Green: high employment in industry and low employment in agriculture

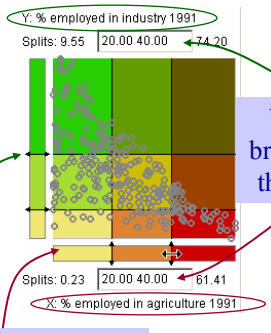
Yellow: low employment in both industry and agriculture

Brown: high employment in both industry and agriculture

Cross-classification (2)

The classes can be changed:



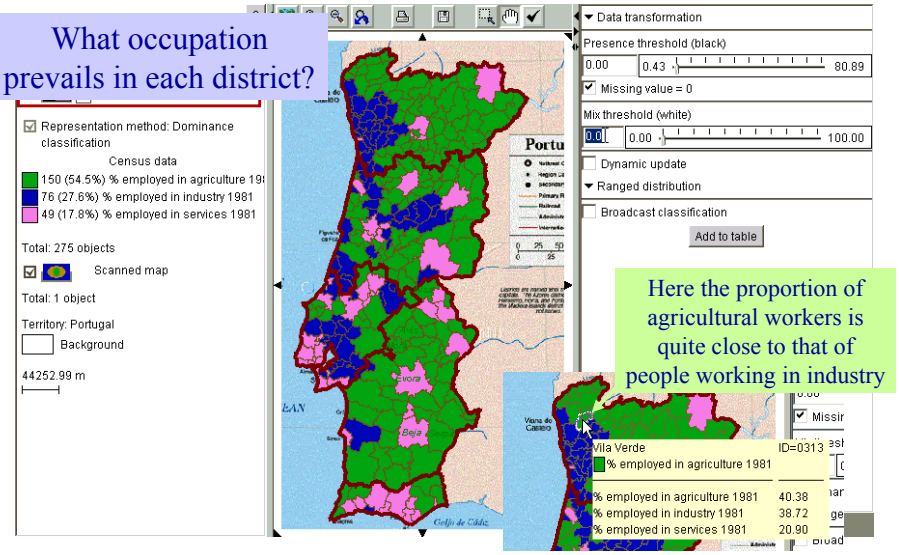


by entering break values in the text fields

by clicking on the color bars and moving the sliders

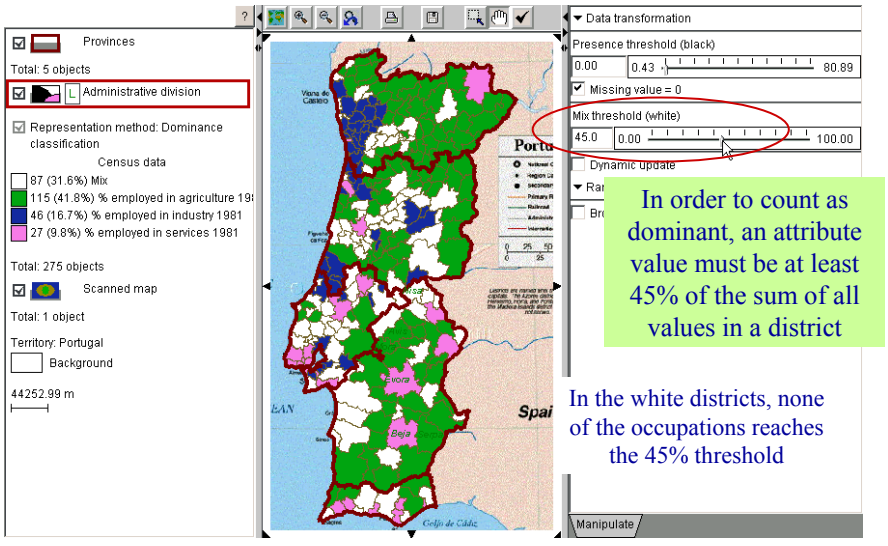
Classification by Dominance (1)

What occupation prevails in each district?



Here the proportion of agricultural workers is quite close to that of people working in industry

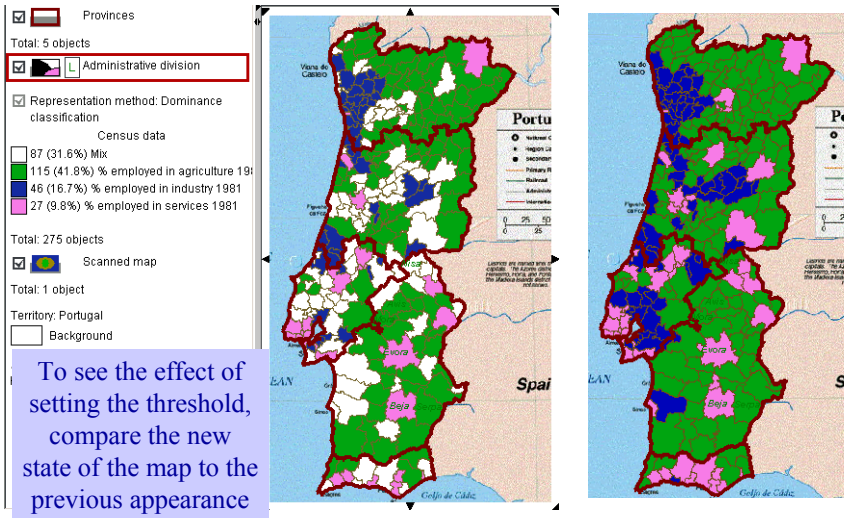
Classification by Dominance (2)



In order to count as dominant, an attribute value must be at least 45% of the sum of all values in a district

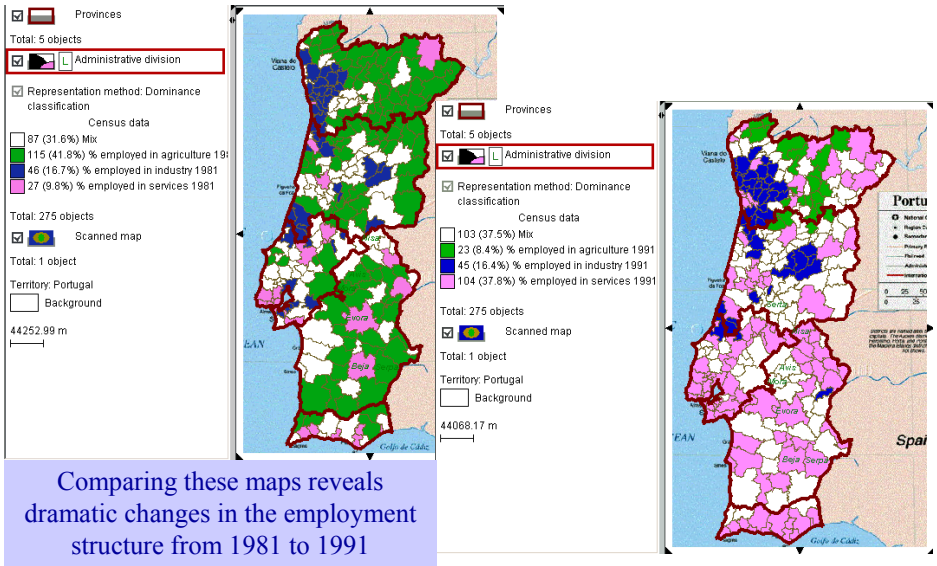
In the white districts, none of the occupations reaches the 45% threshold

Classification by Dominance (3)



To see the effect of setting the threshold, compare the new state of the map to the previous appearance

Classification by Dominance (4)



Comparing these maps reveals dramatic changes in the employment structure from 1981 to 1991

Summary

This lecture was supposed to

- introduce the ideas of exploratory data analysis
- explain what an interactive map is
- demonstrate various methods of cartographic visualisation and user-map interaction
- show how to use these tools for data investigation