



Conference on “Visualising and Presenting Indicator Systems” 14 – 16 March 2005

Socio-economic typology of the Metropolitan Area of Lisbon Ana Alexandrino da Silva

My presentation is based on the project called “**Socioeconomic Typology of the Metropolitan Area of Lisbon**”. In this territory, that represents only 7% of the Portugal’s area, lives more than one quarter of the Portuguese population.

The indicators were built from Census micro-data.

This meant, in certain cases, to work with more than 2 and a half million records, one for each person that lives in the Metropolitan area of Lisbon.

The syntax of SPSS provided a useful way of programming the different parts of the project, namely:

- Retrieving data from this huge census database;
- Building indicators taking into account their concepts and metadata;
- Validating the information;
- Testing which indicators were relevant to factor analysis.

Only after having a stable set of indicators one could test the univariate and multivariate analysis that is shown on the poster.

Sixty (60) indicators were built using the census information for **People, Families, Dwellings and Buildings**. As we know, there isn’t an optimal set of indicators to make such a portrait. One indicator that is relevant for one metropolis can be inappropriate to another.

The choice of the indicators took into account different socioeconomic dimensions that characterize the Metropolitan territory of Lisbon. These dimensions and their indicators are shown in this poster: Mobility, Qualification, Urbanization, Precariousness, Ageing, family structure, etc]. For example, to measure **urbanisation** we kept:

- The Percentage of Buildings with 7 or more flats
- The Percentage of Not exclusively residential buildings
- The Percentage of rented dwellings
- The Percentage of owner occupied dwellings without mortgage and
- The Social diversification index

Sometimes we had more than one indicator to describe the same idea. For example to characterize **Urbanization** we chose *Percentage of Buildings with 7 more flats*, but we also had *Average number of floors per building*, or *Number of flats per floor*, among others. Factor analysis decided which one was more adequate.

Please before we proceed I would like to make 2 remarks as these units were defined mostly for collecting data purposes:

- 1) Non residential areas are shaded in light grey and are not included in this study
- 2) The territorial division used for this study takes into account the resident population: bigger areas have less people and smaller areas have more people - Therefore the visual effect given by the maps may be a problem to an innocent viewer, but it is clear once you know how to look.

In the publication of the results, that you can look afterwards, we had to make two maps for each indicator:

Why 2 maps? To have both an overview of the area of study and a zoom of the area around Lisbon, where most of the population lives.

As you know, in a choropleth thematic map like these, the image you give depends on how data is classified. There is no best image for a certain indicator, but you can choose a method of classification that gives you an approximate image of the reality as you cognitively think the reality is. This means that we shouldn't classify data at our own pleasure but need to take into account at least the data distribution.

Two different methods were chosen: for symmetrical distributions the average and standard error classification method, that gave us the possibility to say that the 2 first classes were below the average and the 2 last ones were above the average. For asymmetrical distributions we use Jenks method (called 'natural breaks' in Arcview software) that minimizes between class variance and maximizes interclass variance.

Factor analysis attempts to identify components that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of components that explain most of the variance observed in a much larger number of variables.

In this case the data matrix was reduced in 6 components that explained 67,6% of the total variance.

Which components describe the data? The most important is "renewal", followed by "qualification", "urbanisation" that is shown here, "mobility", "migration" and finally "precariousness".

To have an overview of the distribution of each phenomenon on the territory we made maps for each component.

Why 3 classes? An odd number allows an intermediate class ("neutral") and focus on the positive, or negative, importance of the phenomenon.

Which classification method? Since the components have a symmetrical gaussian distribution, the component data results were divided in four groups by the quartile method.

In the "urbanisation" component, for instance: The 1st quartile represents the rurality (blue fill) The 4th quartile refers to the presence of urbanisation (red fill) and the 2nd and 3rd quartile represents the absence of the phenomenon (white fill).

This procedure allowed us to have a bipolar legend - positive, neutral, negative meaning - with the same legend titles; easy and meaningful for those using these results.

Finally, the cluster analysis gave us the socioeconomic typology that we were looking for.

The dendrogram, this graph that looks like a tree on reverse, helps us to choose the number of clusters and shows where are the breaks. We can see 3 main groups: the old violet, on one hand; the unqualified Yellow, orange and red on the other hand, but also the qualified light and strong blue.

The violet is called **Urban consolidated** and we can find it in the city centres. Old people living alone, old and rented buildings and tertiary activity are the main characteristics. Please look at this dot chart where is represented the percentage population living in this cluster. Lisbon and Almada (Like Buda and Pest...) are the only municipalities above the average.

Yellow represents **Rural**. This cluster is found in the area apart from the Lisbon and the main roads. Small buildings, bigger families, primary and secondary activity and low education level characterize it.

Orange represents **unqualified urban and suburban** territories found mostly on the first ring around Lisbon. Long time-distance and commuting public transportation, big density buildings with small, not paid houses where lives married couples with children are the main characteristics of this cluster.

Red areas are located in the municipalities around Lisbon or in its boundaries. They correspond to **precariousness** and social exclusion. In fact, we find there big families living in low-cost little rented

houses or shanties; people with low academic qualification and low qualified professionals or unemployed; also foreigners and people with consensual unions...

Qualified areas are shaded in light blue and are located in noble territories from the social point of view, in central territories from the time-distance point of view and also in periphery green areas. We can find here high-qualified professionals' with high education, using motorcar in their commuting journey. People that lives in big and comfortable and already paid dwellings.

Finally the **new suburban** shaded in strong blue, mostly found in the second suburban ring around Lisbon. It's characterization is shown in the second dot chart.

We decided to call it NEW because low resident population and buildings average age. High and crowded buildings are predominant and most dwellings are still being paid by residents that didn't live in that municipality before. In these typical suburban areas lives couples with children that subsist exclusively from its salary and uses motorcar to go to work in another municipality.

To characterize each cluster we also made a map where all the other clusters were shaded in white. The colour choice was warm colours for less integrated territories, like red for precariousness, and cold colours for more integrated territories.

The legend ordering is different from the dendogram. We put urban before rural and qualification before exclusion.

Between 1991 and 2001, the metropolitan area of Lisbon had a strong socioeconomic change, especially in the first and second suburban ring around Lisbon.

After applying the same methodology described before we arrived approximately to the same clusters. Because we considered for the same little areas both 1991 and 2001 data, after the cluster classification we could say that one area changed from rural to qualified suburban. In general, we can say which areas changed their socioeconomic characteristics – here shaded in black – and which remained the same. If The cd-rom shows us the changes at neighbourhood level, something impossible with this maps...

On the other hand we can see to which socioeconomic cluster each area changed to. Half of the territory changed it's socioeconomic characteristics.

Qualified suburban, new suburban and urban consolidated increased population. Unqualified suburban, rural and precariousness lost.

Integration and qualification of this territory are the key-words for this study.

This project took 9 months, it was made by two persons and has been used by municipalities, university researchers and to construct samples for surveys purposes... The next study only can be done in 10 years because it was based and census data.