

Cluster Analysis applied to the evaluation of urban landscape quality

A. Campos & R. C. Oliveira
*Ceris, Cesur, Instituto Superior Técnico,
Universidade de Lisboa, Portugal*

Abstract

The quality of the urban landscape is a complex concept which is difficult to define and synthesize. The tri-dimensional characteristics of this quality – aesthetic, cultural and functional features – can only be improved through ensuring that some key elements (such as harmony, equilibrium, coherence, diversity, complexity, management and the sustainability of the urban landscape) are taken into account. Therefore, a deeper analysis of those aspects of the urban environment and public space which devalue the quality of the urban landscape is required, and this is a new approach that needs to be adopted in order to consider all of these aspects.

The methodology adopted in this paper is based on the development of a set of criteria and indicators (variables) which not only enables the creation of a ranking of the quality of the urban landscape, but also permits deeper insights into those aspects that need to be specifically addressed in policies and actions, making it possible to map out Lisbon's best and worst urban landscapes, thus establishing a basis for comparison, based on experts' evaluation and Cluster Analysis.

Cluster Analysis is a statistical classification technique that aims to group entities (geographical units) in clusters which are internally homogeneous and heterogeneous among themselves, e.g. as different as possible between groups and as similar as possible within each group. The results obtained show that Cluster Analysis gives a solid response in grouping geographical areas, and the several clustering methods (hierarchical and non-hierarchical) used point to the same groups constitution, in terms of geographical units, which confirms the high consistency and robustness of the formed clusters.

Keywords: urban landscape, quality, indicators and indexes, Cluster Analysis.

1 Introduction

The quality of the urban landscape has both objective and subjective components. On one hand, the former are related to physical characteristics, the quality of the materials that constitute the public space, urban design, urban furnishings, air quality and noise levels. On the other hand, the subjective components refer to aesthetics, pleasure, experience and individual preferences (even smell) [1].

A direct relation between the objective and subjective components of quality do not always exist and, in fact, an objective assessment of quality does not necessarily imply an identical subjective assessment. Most times the positive assessment of quality is made more according to the subjective component, rather than the objective one.

The perception of quality, and the way that it is assessed, is reflected by self-experience and by the appropriation of space, and also by the way that one reviews the urban landscape. Apparently, some areas have a lot of quality but are not appreciated. It is common to hear the statement “I do not like it, but I recognize that it has quality”, which is the same as saying that there is no correlation between quality and the individual subjective assessment of what is quality. This reaction to the urban landscape also has to do with how the space is appropriated for its function, which is not always as suitable as it was imagined. Two spaces with similar characteristics and design may have different ratings, depending on how much they are enjoyed. If two equal spaces in terms of design and materials are considered, one of which is central and accessible, and is full of life, whilst the other, although similar, has an eccentric location or difficult access conditions and is empty, then the assessment of the quality of each of these two spaces is carried out differently, albeit from the point of view of design and material it is strictly the same.

The sum of the quality of urban elements, which range from the artificial to the natural, from buildings to the public space and includes the satisfaction of basic socio-environmental needs, a clean environment and access to water supply, socio-economic and cultural needs – in short a qualified urban landscape, inevitably all translate into an improved quality of life.

The quality of the urban landscape is the result of combining and weighing up of all the visible and invisible factors, objectives and subjective aesthetics and also the appropriation of space. Despite relying on the criteria of the observer, the intrinsic values of each user, individual patterns and time and the type of requirements and function that is aimed for, there are factors and features which together can be identified in order to create a standard for the assessment of quality of the urban landscape.

2 Methods overview

2.1 Interviews with a panel of experts

This study aims at the evaluation of the quality of the urban landscape, which is a complex theme for which no instruments exists that are capable of presenting



unique and credible answers. It was decided to survey the opinion of experts on urbanism on the topic of the urban landscape, through interviews with a panel of experts made up of 22 well-recognized scholars and decision-makers, in order to synthesize information from a variety of sources and present a set of views to arrive at generalized conclusions. The experts, who were all connected to universities, public authorities and research laboratories, contributed with their experience and knowledge in providing a value judgment on the topic proposed. The interviews, which lasted about an hour and a half, were on the theme of the urban landscape, focusing on four specific questions: “what do you consider that defines the quality of the urban landscape?”; “what do you consider that devalues it?”; “which urban landscapes in Lisbon have more and less quality, and why?; “how do you rate the quality value of all areas in a scale of 0 to 10 (10 being the highest value of quality)?” From the experts’ answers, an analysis of the content of the interviews was made of the number of occurrences, extracting the most important aspects, which allowed for establishing common denominators for the vectors of “quality” and “factors of devaluation of the urban landscape”, and “ranking of Lisbon urban landscapes of higher and lower quality”. The experts’ opinion was decisive for designing a proposal for a consistent assessment and criticism about the various realities surveyed. The panel of experts identified and grouped *eight* criteria of the greatest relevance and presented their respective justifications. These were: harmony (balance and coherence), the presence of natural elements, complexity/diversity, management and maintenance, panoramic views, the quality of the elements, security/safety, and accessibility. The theoretical definition of the quality of the urban landscape and public spaces, as reported by the opinion of experts, provided a general analysis that defined the aspects of quality which permitted the construction of a set of indicators for measuring the quality of the urban landscape, which is needed to provide technical and political decision-makers with the tools for introducing improvements in the urban landscape. The results of the interviews of the panel of experts identified also some general characteristics that cause the disqualification of the urban landscape, and also the most relevant aspects of the factors of devaluation, which are largely associated with the negative elements which provide quality, giving rise to the *six* most important criteria like chaotic parking, dirt, presence of specific social groups, visual pollution and absence of urban design.

2.2 Geographical units, data collection and presentation of variables

The 16 geographical units selected in this study (8 of higher, and 8 of lower quality) are part of a set of 26 districts (or areas) previously qualified by the experts in terms of quality of the urban landscape. Through observations made and registered directly on site, data were collected in order to calculate the relative variables built to characterize the urban landscape.

It is presented below the list of data collected at various geographical units, according to the quality criteria of the urban landscape to which each refers, and which form the basis for the construction of the variables used in this study:



a) Equilibrium/coherence

HE1 – Sedimentation of various eras – represents the ratio of the number of buildings that maintain the features of the original building in relation to the total number of buildings.

HE2 – Ruptures in building – represents the absence of architectural ruptures inserted into the built in relation to the total number of buildings.

HE4 – Natural elements – represents the ratio of the number of types of natural elements (example: water lines, lakes, trees, flowers) in a given territorial unit.

HE5 – Demographic balance – represents the ratio of the number of age groups (1, 2 or 3 groups; Group of 0–14 years; of 15–64 years, and more than 65 years) in the total number of age groups considered in this study.

HC1 – Volumetric disparity – represents the ratio of the number of buildings that differ volumetrically of the surrounding.

HC2 – Alignment of heights – represents the heights misalignment, in total of buildings.

HC3 – Alignment of facades – represents the misalignment of facades of buildings in total.

HC5 – Architectural styles – represents the variety of architectural styles, for a total of buildings.

b) Natural elements

EN1 – Trees – represents the presence of trees in a particular territorial unit.

EN2 – Water plans – represents the presence of water lines, water mirrors or any water plan in the neighborhood.

EN3 – Flower beds with flowers on streets and squares – represents the presence of flower beds with flowers.

c) Diversity/complexity

D2 – Variety of uses – represents the variety of urban activities, for a total of activities (example: housing, commerce, services, school and health equipment).

D3 – Variety of commercial units at the ground floor – represents the variety of commercial units and different services.

D4 – Number of visible breaks – represents the ratio of the number of break points of visibility along the street (example: number of convex curves in the longitudinal profile, number of curves in plan, arches, roundabouts or other elements that create obstruction or loss of visibility to the end of the street).

d) Management

G1 – Maintenance of public space – represents the management of local authority as regards the maintenance of public space (conservation of materials of sidewalks, decks and street furniture).

G2 – Maintenance of private facades – represents the management of local authority as regards the application of particular care (or legislation and its application) in maintenance of private buildings.

e) Views

V – Panoramic views – represents the rate of panoramic views from a particular street or territorial unit.

f) Quality of the elements

Q1 – Quality of the materials of the buildings – represents the quality of the materials of the buildings (which evaluates the use of more or less rich materials).

Q2 – Quality of materials of public space – represents the quality of the materials of the public space (use of more or less rich materials).

Q3 – Air quality – represents air quality in a given territorial unit.

Q4 – Quality of noise levels – represents noise levels.

g) Safety

S1 – Lighting – represents the ratio of the number of lamps on a street length.

S2 – Policing – represents the presence of police officers on a street.

S3 – Presence of passers-by – represents the average number of passers-by (analyzed on a weekday, in the working period).

h) Accessibility

A1 – Continuity of the urban fabric – represents the ratio of the number of streets with exit, in the total road network.

i) Presence of specific social groups

PSSG – Presence of specific social groups (homeless, gangs, ethnic, gangs and prostitution) – represents the use by a certain type of population.

j) Parking

CA1 – Chaotic Parking – represents the ratio of the number of sidewalks with car parking fully occupied.

k) Cleanness

SJ1 – Washing and sweeping – represents the washing and sweeping of the public space.

SJ2 – Garbage disposals – represents the ratio of the number of litter bins.

SJ3 – Graffiti – represents the ratio of the number of buildings with graffiti.

2.3 Values of the variables

The variables obtained for this study have different units (absolute values and percentages) and numerically translate objective realities, directly measurable, and some other are qualitative, for which a 3 classification levels scale was adopted. The latter were converted to numerical value equal to “0”, “0.75” or “1.0”. These discrete values were associated to the qualitative variables to reflect the perception of quality. Thus, to the “average/good” level it was assigned a numeric value “0.75”, while to the “average/bad” and “bad” levels were assigned the numeric value of “0”.



Since not all the variables are expressed in the same units, they were standardized, making them “dimensionless”. All the standardized variables have a variance equal to unity (and a mean equal to “0”, so they can take positive and negative values), thus avoiding the effects of dispersion in the construction of the clusters. The standardized variables now have, from the outset, identical “weight” for the construction of factors and clusters.

3 Results and discussion

3.1 Cluster Analysis

Cluster Analysis is a statistical technique that aims to group entities (geographical units, in our study) in clusters which are internally homogeneous and heterogeneous among themselves, e.g. as different as possible between groups, and as similar as possible within the same group. Clustering methods are classified in hierarchical and non-hierarchical methods, and the former can be agglomerative or divisive [2].

In the first case, one begins with “n” clusters with a single entity that will be successively merged, until it is achieved a single cluster grouping all entities. In the second case, one starts with a single cluster with all entities that will be subdivided in groups until obtaining “n” clusters with one entity each. Non-hierarchical methods are iterative procedures where entities are aggregated in a pre-specified number of clusters. The agglomerative hierarchical methods applied here used various measures of distance or (dis) similarity between clusters already formed by various entities or by a single entity.

3.2 Hierarchical methods

The hierarchical aggregation methods used here, were the “Complete linkage” method (largest distance between pairs of entities, one for each cluster), and “Ward” method (minimization of internal variances of clusters). The methods were applied with two types of distances: Euclidean distances and “Cityblock distances”. The results for the “Single Linkage” method (smallest distance between pairs of entities, one for each cluster) and “Chebychev” distances, did not present results as “clear” as the other used here.

Figure 1 shows the tree diagram (dendrogram) obtained with the “Complete Linkage” method and it is possible to identify *five clusters*:

Cluster 1 : *Vale Alcântara (isolated);*

Cluster 2 : *Musgueira, Galinheiras, Chelas;*

Cluster 3 : *Bairro da Liberdade (isolated);*

Cluster 4 : *Arroios, Bairro Alto, Xabregas, Benfica, Carnide, Lapa, Avenidas Novas Alvalade, Baixa/Chiado, Campo de Ourique;*

Cluster 5 : *Parque das Nações (isolated).*

As it can be observed in Figure 1, if *three clusters* were considered, “Vale de Alcântara” (Cluster 1) would be merged with Cluster 2 and “Bairro da Liberdade”

(Cluster 3) would be merged with Cluster 4. “Parque das Nações” would remain isolated, merging with Cluster 4 only for a two cluster solution.

The same results are obtained by applying the same method (“Complete Linkage”) but using Cityblock Distances.

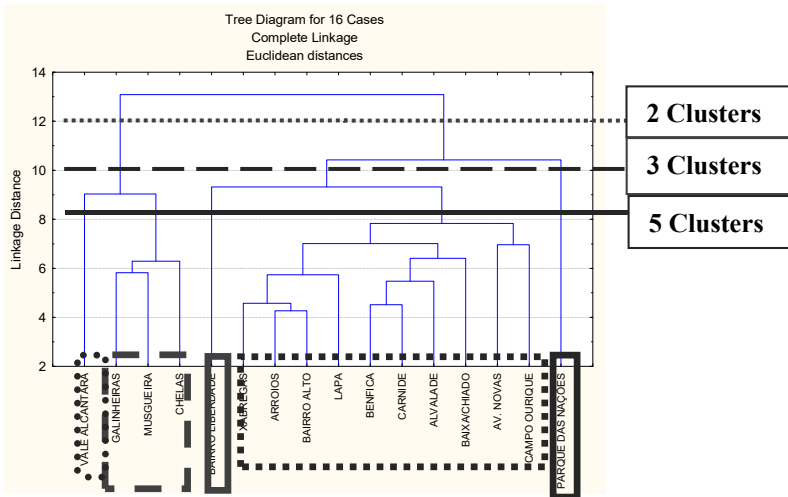


Figure 1: Tree diagram “complete linkage” (“Euclidean distances”).

By applying the “Ward method” (Figure 2), the very same *five clusters* are obtained. This illustrates the robustness of this classification in 5 Clusters.

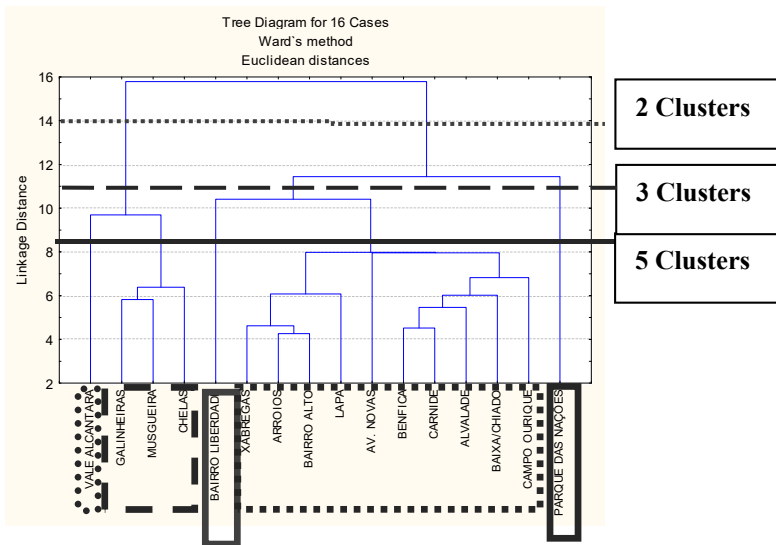


Figure 2: Tree diagram “Ward’s Method” (Euclidean distances).



When applying again the “Ward method”, but now with City-Block Distances, the same five Clusters are obtained (Figure 3). However, for a *three-cluster* aggregation, “Bairro da Liberdade” joins “Parque das Nações” instead of merging with Cluster 4. This is a surprising result that appears to make little sense in face of the dissimilarities between these two units in terms of the quality of the urban landscape.

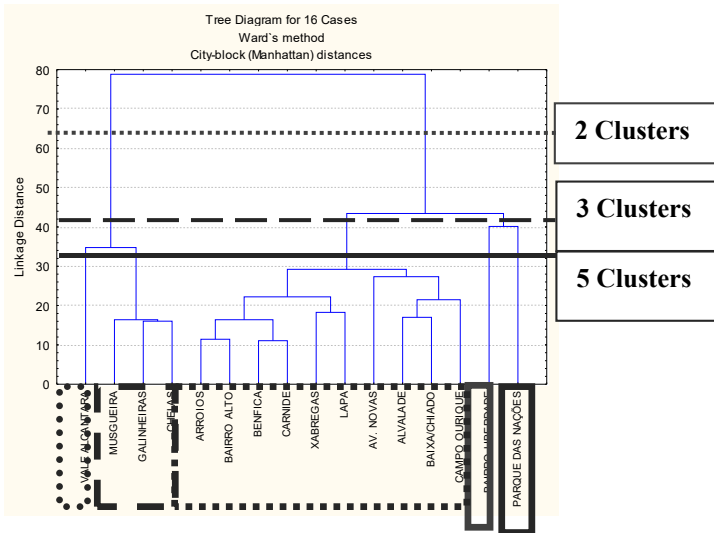


Figure 3: Tree diagram “Ward’s Method” (“Cityblock Distances”).

In the case of *two clusters*, we get the following groups for all clustering methods and distances:

- Cluster 1:** Vale de Alcântara, Musgueira, Galinheiras, Chelas;
- Cluster 2:** Parque das Nações; Bairro da Liberdade, Arroios, Bairro Alto, Benfica, Carnide, Xabregas, Lapa, Avenidas Novas, Alvalade, Baixa/Chiado, Campos de Ourique.

This classification in *two groups* of the geographical units clearly divides the areas of higher urban landscape quality and lower quality, making sense the composition generated with the selected variables.

Additionally, this classification matches perfectly the expert’s opinion.

3.3 Non-hierarchical methods

The application of “software” STATISTICA using the “K-means method” for a pre-defined number of *five clusters*, made possible to identify the following composition:



- Cluster 1:** *Campo de Ourique, Baixa/Chiado, Alvalade Lapa, Carnide, Avenidas Novas, Bairro Alto, Xabregas, Benfica, Arroios;*
- Cluster 2:** *Parque das Nações (isolated);*
- Cluster 3:** *Vale de Alcântara (isolated);*
- Cluster 4:** *Bairro da Liberdade (isolated);*
- Cluster 5:** *Galinheiras, Musgueira, Chelas.*

This again coincides with the results obtained with hierarchical methods and reinforces the robustness of this classification.

The application of “software” STATISTICA using the “K-means method” for *three clusters* identified the following composition:

- Cluster 1:** *Galinheiras, Musgueira, Chelas.*
- Cluster 2:** *Vale de Alcântara (isolado)*
- Cluster 3:** *Parque das Nações, Campo de Ourique, Baixa/Chiado, Alvalade, Lapa, Carnide, Av. Novas, Bairro Alto, Xabregas, Bairro da Liberdade, Benfica, Arroios.*

In this case, “Vale de Alcântara” appears isolated and “Bairro da Liberdade” joins the cluster of units with higher urban landscapes quality.

The application of “software” STATISTICA using the “K-means” for the pre-defined *two clusters* made possible to identify the following composition:

- Cluster 1:** *Vale de Alcântara, Galinheiras, Musgueira, Chelas;*
- Cluster 2:** *Parque das Nações, Campo de Ourique, Baixa/Chiado, Alvalade, Lapa, Carnide, Av. Novas, Bairro Alto, Xabregas, Bairro da Liberdade, Benfica, Arroios.*

This composition makes much sense by joining “Parque das Nações” with other units with admittedly more quality of the urban landscape, and aggregates “Vale de Alcântara” with lower quality units.

However, “Bairro da Liberdade” is also aggregated to higher quality units of urban landscape which, not making much sense may indicate some “excessive” weight given to some variables, such as the “views” factor that can “pull” this geographical unit into a group of quality landscapes (which goes against the pre-conceived idea about this neighborhood).

In any case, this group in two clusters allows highlighting features that divide the geographical units of higher and lower quality of the urban landscape.

In Table 1 are presented comparisons between clusters obtained with hierarchical and non-hierarchical methods, and a comparison between experts’ evaluation and cluster analysis. Cluster 1 represents lower quality of the urban landscape while Cluster 2 represents higher quality of the urban landscape.

Table 1: Cluster analysis – comparing the clusters of hierarchical and non-hierarchical methods (*for two and five clusters*).

	Hierarchical methods and non-hierarchical methods		Expert's perception
<i>C1</i>	<i>Vale de Alcântara, Galinheiras, Musgueira e Chelas</i>	<i>C1</i>	<i>Vale de Alcântara, Galinheiras, Musgueira, Chelas, Xabregas, Bairro da Liberdade, Benfica e Arroios</i>
<i>C2</i>	<i>Parque das Nações, Campo de Ourique, Baixa/Chiado, Alvalade, Lapa, Carnide, Av. Novas, Bairro Alto, Xabregas, Bairro da Liberdade, Benfica e Arroios</i>	<i>C2</i>	<i>Parque das Nações, Campo de Ourique, Baixa/Chiado, Alvalade, Lapa, Carnide, Av. Novas, Bairro Alto.</i>
<i>C1</i>	<i>Vale de Alcântara</i>		-----
<i>C2</i>	<i>Bairro da Liberdade</i>		-----
<i>C3</i>	<i>Musgueira, Galinheiras, Chelas</i>		-----
<i>C4</i>	<i>Arroios, Bairro Alto, Benfica, Carnide, Xabregas, Lapa, Av. Novas, Alvalade, Baixa/Chiado, Campo de Ourique</i>		-----
<i>C5</i>	<i>Parque das Nações</i>		-----

4 Conclusions

The main objective of this study was to determine the quality criteria of the urban landscape, to identify the factors that devalue it and to apply Cluster Analysis in some geographical Lisbon areas. To achieve these objectives, one fundamental instrument had a major contributor – the interviews with the panel of 22 experts in the subject concerned. This method of interviews was very relevant, as it enabled to determine the quality criteria and those devaluation factors, which led to the construction of indicators (variables), as well as determining the geographical areas with the best and worst urban landscape quality in Lisbon.

On the other hand, Cluster Analysis, both by hierarchical and non-hierarchical methods, produced sensible results with regard to the grouping of geographical units. The five identified clusters “make sense” in this study, since they grouped, as intended, entities with homogeneous characteristics among themselves in terms of quality of urban landscape and, on the other hand, separates those that clearly are distinct from other clusters. Such conclusions were not evident at the beginning of this study.

It should be noted that both hierarchical and non-hierarchical methods point to classifying geographical units, either isolated or in groups, which confirmed, and

highlight, some relevant aspects for the analysis of urban landscapes of higher and lower quality.

It is believed that a larger sample size, with a larger number of observations per geographical unit, or even a greater number of them, could enrich the analysis of clusters, and bring greater clarity in the interpretation of results.

References

- [1] Campos, A., *Contribution to the urban landscape evaluation*, Dissertação de mestrado, Instituto Superior Técnico, Universidade Técnica de Lisboa, pp. 7–10, 2015.
- [2] Reis, E., *Estatística Multivariada*, Edições Sílado (2^a Edição), Lisboa, pp. 287–328, 2001.

