PMKT
BRAZILIAN JOURNAL OF MARKETING, OPINION, AND MEDIA RESEARCH

Chief Editor: Fauze Najib Mattar | Evaluation System: Triple Blind Review
Published by: ABEP - Brazilian Association of Research Companies
Languages: Portuguese and English | ISSN: 2317-0123 (online)

Glocal Segmentation[1]

Glocal Segmentation

Luiz Sá Lucas*
MC15 Consultoria, Rio de Janeiro, RJ, Brazil

Leonardo Soares
Instituto COPPEAD da Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

Wagner Esteves
IBOPE Inteligência, Rio de Janeiro, RJ, Brazil

ABSTRACT

The article presents an algorithm to create a Glocal Segmentation, that targets not only the whole set, but also its parts, all in a coordinated way and not with independent segmentations. The technique is applied to synthetic data for five Latin American countries. Another original aspect of the study is the fact that all data analysis was performed using data visualization.

KEYWORDS: Segmentation; Cluster analysis; Data visualization; Genetic algorithms.

RESUMO

O artigo apresenta um algoritmo para criar uma Glocal Segmentation, que segmenta não só o conjunto, mas também suas partes, isso tudo de uma forma coordenada e não com segmentações independentes. A técnica é aplicada a dados fictícios relativos a cinco países da América Latina. Outro aspecto original do estudo é o fato de toda a análise dos dados ser realizada por meio de visualização de dados.

PALAVRAS-CHAVE: Segmentação; Análise de *cluster*; Visualização de dados; Algoritmos genéticos.

Submission: 07 June 2016
Approval: 09 December 2016

Luiz Sá Lucas
Master in Systems Optimization by Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE) at Universidade Federal do Rio de Janeiro. Consultant at MC15 Consultoria.
(CEP 22451-070, Rio de Janeiro, RJ, Brazil).
E-mail: luizsa.lucas@mc15.com.br
Endereço: Rua Vice-Governador Rúbens Berardo, 65, b1oco A 406 –22451-070 - Rio de Janeiro, RJ, Brazil.

Leonardo Soares
Master in Business Administration from the Federal University of Rio de Janeiro. Consultant of K2 Achievements.
E-mail: leonardosoares@k2a.com.br

Wagner Esteves
Master's degree in Production Engineering from the Universidade Federal Fluminense. Manager of Technical Assistance and Planning of IBOPE Inteligência.
E-mail: wagner.esteves@ibopeinteligencia.com.br

INTRODUCTION

1.1 A GLOBAL COMMUNITY?

Levitt (1983) has been cited as the creator of the term "globalization", presented in the article in question. However, the term globalization was already in general use since the 1940s and by economists since the 1980s. This does not change the fact that Levitt is probably the term's originator in marketing. In the cited article he also described what he understood as a Company acting in a Global Community. According to him, in an international economy, as traditional markets became saturated, it was necessary to meet the demand of the new ones. But how to do that? If the goal was to serve each market separately, then "the game was not worth the effort". However, according to Levitt (1983), this was not necessary: well-managed companies should emphasize the provision of globally standardized products that are advanced, functional, reliable and low-priced. According to Levitt (1983), a powerful force caused convergence: technology. "Almost everyone wants the things they have heard about, seen or experienced through new technologies." It should be noted here that Levitt (1983) showed that each market had its characteristics:

> In Brazil, thousands of people are swarming in the darkness of pre-industrial Bahia (sic) in an explosion toward coastal cities, to quickly install television sets in batches of corrugated boxes (sic) and, alongside damaged Volkswagen cars, making ritual sacrifices of fruit and chickens just sacrificed to macumba spirits by candlelight.

During the fray fight in Biafra, still according to our master Levitt (1983), "daily reports on TV showed soldiers carrying bloody colleagues while drinking Coca-Cola."

In addition to his deep ignorance about overseas, Theodore Levitt wanted to emphasize that whatever culture, a set of standardized, internationalized commodities would be consumed without any need for adaptation.

1.2 A COMUNIDADE GLOCAL

Although the expression, as known today, is not explicit in the book Cities in Evolution (Gueddes, 1915), the idea, applied to urban planning, is there clearly expressed. Global brands should be unique, meaningful and long lasting in all markets in which they operate. However, an action that takes in account, as is almost always the case, a segmentation, must be based on the needs and benefits perceived in each target / geography, not only within a country, but also within the group of them. The site BBA-MBA.net presents an expression related to the rather interesting case: "Think local. Learn global. Act glocal "(BBA-MBA.net, 2016). The idea is that companies of all sizes, present in large and small markets and perceived as universal by their consumers as competing in a global market, should focus their resources on a successful strategy across their geography. In other words, its global future will only be achieved if the company is, firstly, successful locally.

However, it does not make sense and neither does the quoted text affirms that a different segmentation should be done for each specific area or target / target. Then we would fall back on what Levitt (1983) commented on as "the game is not worth the effort". On the other hand, a single segmentation that addresses only the total will not consider specific aspects of geographic areas and/or targets that will be selected.

The concept of marketing segmentation was created in the mid-1950s (Smith, 1956). Excellent texts on the subject have been there for some time. About segmentation and marketing strategy is worth mentioning Myers (1996). On segmentation techniques, one of the best, if not the best, to our knowledge, is the work of Wedel and Kamakura (2000). There are more recent texts, but these are works that stand up to time.

The existing literature deals with the general problem: segmenting the whole. But how to segment a whole and, at the same time, segmenting its parts so that all results are coherent? This is what our article is about.

2 METHODOLOGY

Imagine a company that operates primarily in five countries within Latin America: Argentina, Brazil, Colombia, Mexico and Chile, and that, in its work in the area, it wants to take in account the attitudes of mothers about their children's upbringing up to 4 years.

Let us also assume that a study applies to a sample of 4,000 cases (800 in each country) a battery of 30 attributes with respect to this education: traditional or more modern methods, division of labor with husbands, access to the Web, time dedicated to children, working versus dedicating oneself to children, putting up early in daycare or equivalent, allowing the consumption of soft drinks, etc. The data in our example are artificial and we do not know of any study in these five countries that deals with the subject. Thus, the results of this work are **synthetic** and **cannot be considered necessarily valid** in the real case.

This method is a meta-method, since it can be implemented with several different algorithms, and is outlined in Figure 1. The idea of accomplishing this type of segmentation is not completely new: a few years ago it was presented at the Advanced Techniques Forum - ART Forum, from the American Marketing Association, an article on the subject (Lee & Kelley, 2010). The methodology described there is quite different from ours and is close to a *gibbs sampler*. In that case, the R software *package clue* (which deals with cluster ensembles) can be used, the method consumes a lot of RAM and the required computing time can be quite large, up to four days for processing for around 600 iterations.

By perfecting the algorithm and using only part of the information, Lee and Kelley (2010) have achieved a running time of up to 23 hours for 2000 iterations, depending on the number of observations (the authors did not report this number).

As we said, the idea is to segment the whole without neglecting the parties (the countries). Here we want to use all the information throughout the process. The idea, in our case, is to perform six segmentations (clusterings), but somehow coordinating these processes. We would then have six clustering problems, with a coordinating level, as in the scheme shown in Figure 1.

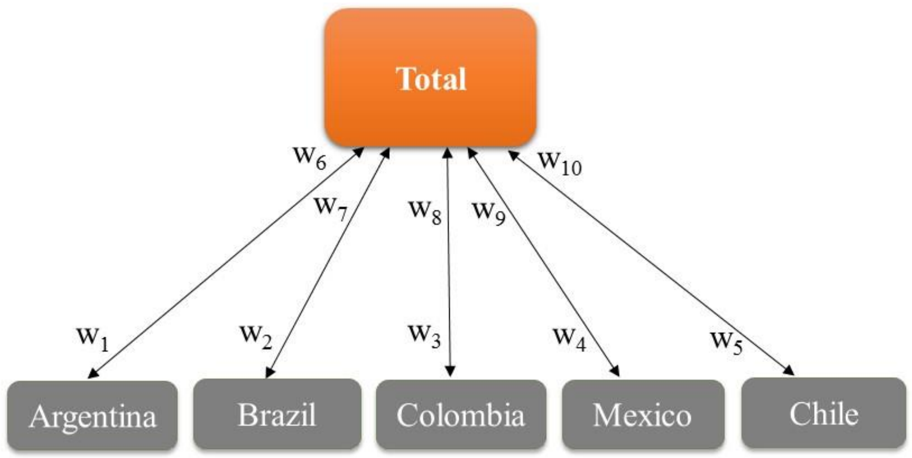

Figure 1 – Meta-heuristics

At the coordinating level we have a clustering of Total. At the lower level we have the clusterings of each of the five countries. Coordination is performed through the values of the *w*.

For a better understanding of the algorithm we will introduce here the concept of *median* a group of multivariate observations. The concept is illustrated by Figure 2. In that figure we have eight elements. The median is element k that minimizes the sum of the distances (euclidian, manhattan, etc.) between *k* and the various group elements *i*, *j*, etc.

Let then be *m* groups, where each one is similar to the one in Figure 2. Let *n* be the number of observations as well, in which, in each of them, there are *p* characteristic measurements. These characteristics can be quantitative (as a scale of agreement), or categorical (such as sex and marital status, for example). We can then define the *k-median* problem. The objective of this problem is to obtain the set of *m* medians such that the sum of the distances of each observation to its median (the median of its group) is minimal. The problem is described in detail, for example, in Sá Lucas (1983).

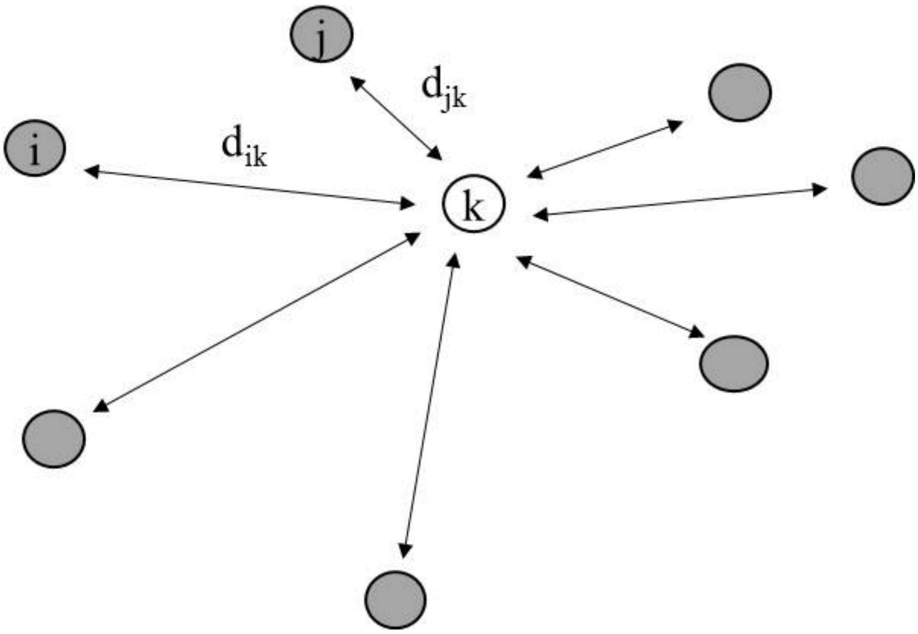

Figure 2 – Group Median

Thus, it may become clearer what each *w* means. If we have *m* groups, then we have *m* medians. Each of those medians has *p* observations (the characteristics based on which we are doing the segmentation/clustering). The median *m* of each problem (in fact of each of the six sub-problems in our case) can be considered as a set of *m* observations that can be aggregated w times to the corresponding sub-problem. Correspondence is given by the arrows in Figure 1. The intensity of coordination is given by the values of *w* which, of course, are integers. Suppose *m* is equal to five (5) groups for the master program (the Total problem). If, for example, w_1 is equal to 3 (three), we will be adding the five medians of the Total problem to the Argentina sub-problem three times ($w_1 = 3$). On the other hand, if w_6 equals 2 (two), and if the sub-problem Argentina has 6 (six) groups, we will be adding the six medians from Argentina to the Total twice.

Note that Total and country targets do not need to have the same number of groups. It is worthwhile, however, to make the Total number of groups equal to or greater than the number of groups for the sub-problems which, in turn, may have different number of groups among them.

The values of w indicate the degree of influence of one problem on the other, that is, the emphasis of coordination between them. It should be remembered that the data of Brazil (800 observations) are the same in the Total problem and in the Brazil problem. The exception, of course, is the data derived (added) from the activity of *w*.

The final goal of the metaheuristic is that, for example, the affinity between the Brazil segmentation derived from the Brazil problem and the Brazil segmentation obtained in the Total problem is as large as possible, although these segmentations are not necessarily the same. The measure of affinity between the segmentations of each country in the sub-problem and in the Total was the measure by the Kappa coefficient (Viera & Garrett, 2005). This coefficient ranges from 0 (zero) - concordance due to chance - to 1 (one) - the segmentions are the same. The geometric mean of the five Kappas, one for each country, will also be between zero and one, and the closer the value is to 1 (one) the better the overall solution obtained.

To sum up: we have a rather complex optimization problem that involves defining the coordination parameters *w*, integers that, by solving in each iteration six clustering problems, maximize the geometric mean of the Kappas for each country. The final solution of the problem is given by the segmentation of the Total problem. In our case, the solution of the problem, that is, the set of values of w_i, was obtained by means of a genetic algorithm.

As is the case with genetic algorithms, the obtained solution w is not necessarily optimal, but the global agreement, being a geometric mean of Kappas, can be evaluated by its final value: the closer to 1 (one) the better.

3 RESULTS

The results will be presented here also in an original way. A set of techniques, some very recent some other secular, called Data Visualization (DV), have been increasingly used in the literature. An interesting presentation of this approach is given in Knaflic (2015). Our presentation will address the data in this way, which will give the reader the opportunity to examine its practical utility.

However, although the DV is normally used to present the data to an audience (company management, clients, etc.), we will adopt a different strategy here. The visualization will be used for the analysis of the data, not necessarily involving the presentation to any audience. This will allow the analyst to use the same more complex graphs for analysis, as will be seen below.

3.1 RESULTS - GENETIC ALGORITHM

Figure 3 shows the evolution of the genetic algorithm. This algorithm works, in each iteration, with several solutions (here called a population of solutions). In each iteration we have, for each population:

- The best value obtained until the iteration (the maximum value);
- The average value;
- The third quartile;
- The median value;
- The first quartile;
- The minimum value.

In the graph, the solutions in each iteration are presented in the order described above. In the algorithm, we used a population of 50 solutions in each iteration (maximum iterations was equal to 30), as shown in Figure 2.

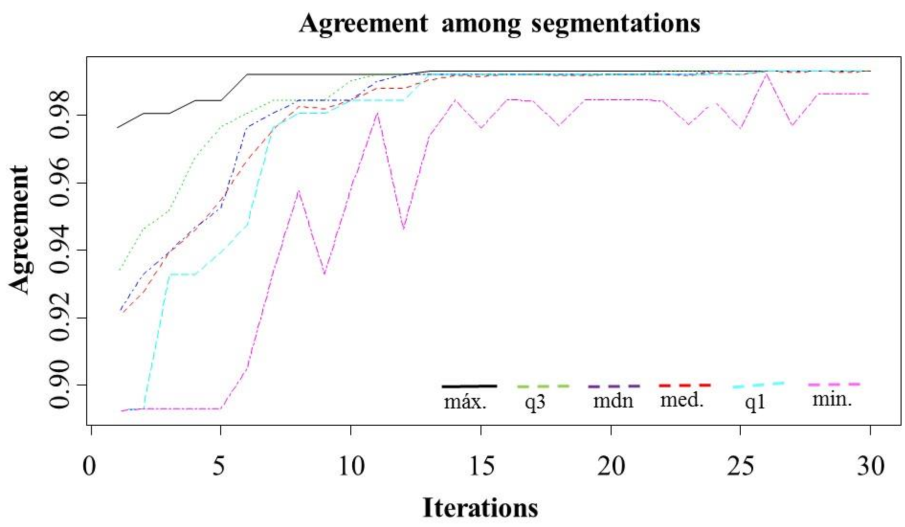

Figure 3 - Evolution of the geometric mean of the Kappas

It is possible to notice that, since the seventh iteration, a value very close to the best one had already been obtained, and this best value is very close to 0,99. This number corresponds to the geometric mean of the Kappas of each country, and as it is seen, is practically equal to 1 (optimal value).

3.2 RESULTS - GROUPS OBTAINED

The solution obtained for Total and each sub-solution for each country, in five groups, are shown in Figure 4.

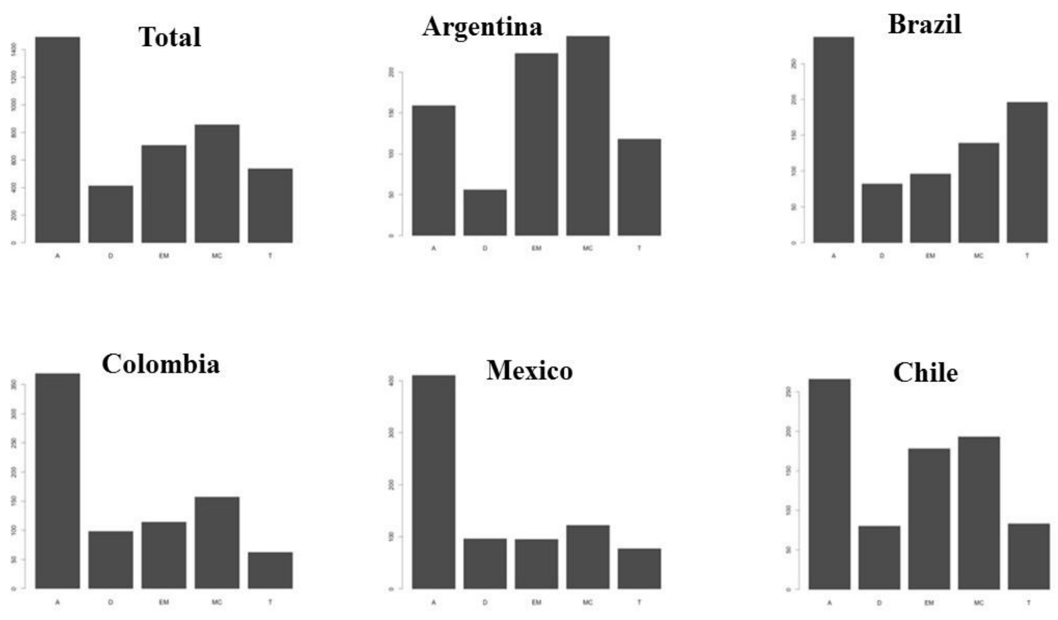

Figure 4 – Segmentations with five groups

It is noted that the total global solution, above left, has different observation numbers in each group from each individual country. Thus, the incidence of the same group in each country is different: a global strategy, unique, can be adapted to each country, emphasizing the desired targets.

In Figure 4 also the groups are indicated, whose analysis, of course, only needs to be done for Total. Thus we have the mothers:

- C - Carefree;
- D - Dedicated;
- MC - Modern Collaborative;
- T - Traditional;
- ME - Modern Education.

These labels were created for groups from the basic axes detected in our synthetic data. We repeat, the data are artificial and do not allow an extension to the real case. In any case, the groups are:

- Carefree: mothers who think they spend adequate time for their children, for themselves and for their husbands, and do not find the work of mothers stressful;
- Dedicated: mothers who are very worried about their children, find the work of a mother stressful, access the Web to know the news etc.;
- Modern Collaborative: fathers take care of children as much as they, mothers, and they like that children discover the world for themselves;
- Traditional: Unlike modern ones, they tend to use more traditional methods for creation without further experimentation;
- Modern Education: take care of children differently from their parents.

Later, in the analysis, we will see how the characterization of the mothers was identified. We will now explore the relative size of groups by country.

Figure 5 corresponds to a *treemap*, which is interesting to identify the relative sizes of each part in the total. The number of observations in each country is equal: 800 observations, so for each country the area is the same. In each country, the largest group(s) has(ve) a larger area and a more intense color:

- In Argentina, we highlight the Modern ones: Collaborative and Education;
- In Brazil, the Dedicated and the Traditional ones;
- In Chile, the Dedicated and the Modern (both, as in Argentina);
- In Colombia, the Dedicated.
- In Mexico, also the Dedicated.

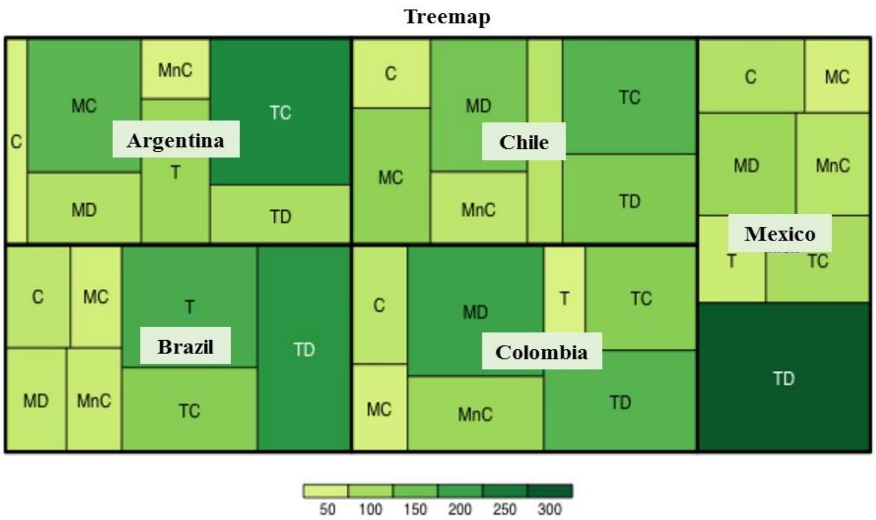

Figure 5 – Group size by country

It should be noted that, in the Total (Figure 4 - Chart 1), the dominant groups are the Dedicated. To a lesser extent, the Modern: Collaborative and Education. In the treemap, it is possible to notice a difficulty that several authors associate to measures that are expressed in area: in Argentina, in the treemap, it is somewhat more difficult to distinguish that Collaborative Moderns have a higher incidence than Modern Education, while this reading is clear in the bar graphs of Figure 4. It is also possible to do this type of analysis by means of a *mosaic chart* which, in this case, has a built-in hypothesis test (Figure 6). The stronger the square in blue, the more the group is above the general average. The more red, the lower.

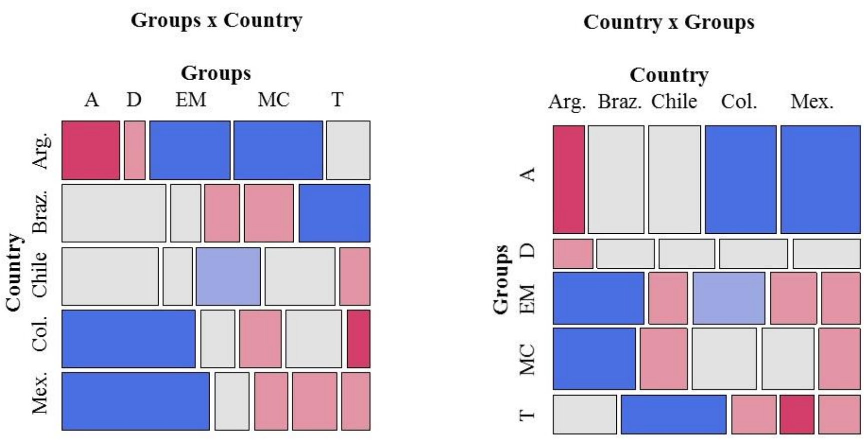

Figure 6 – Group size by country

It is also possible to analyze the total socioeconomic composition of the groups. This is done here through a multiple of mosaic charts (without hypothesis testing). There is no excessive variation in age groups, marital status, employment, and SEC (Figure 7).

Figure 7 – Socioeconomic composition of the groups

Finally, the same analysis can be done by a heatmap (Figure 8), which conglomerates not only the segments but also the socioeconomic levels, but again it is verified that there is not such a great difference in the demographic. A description of heatmaps can be found in Evergreen (2017).

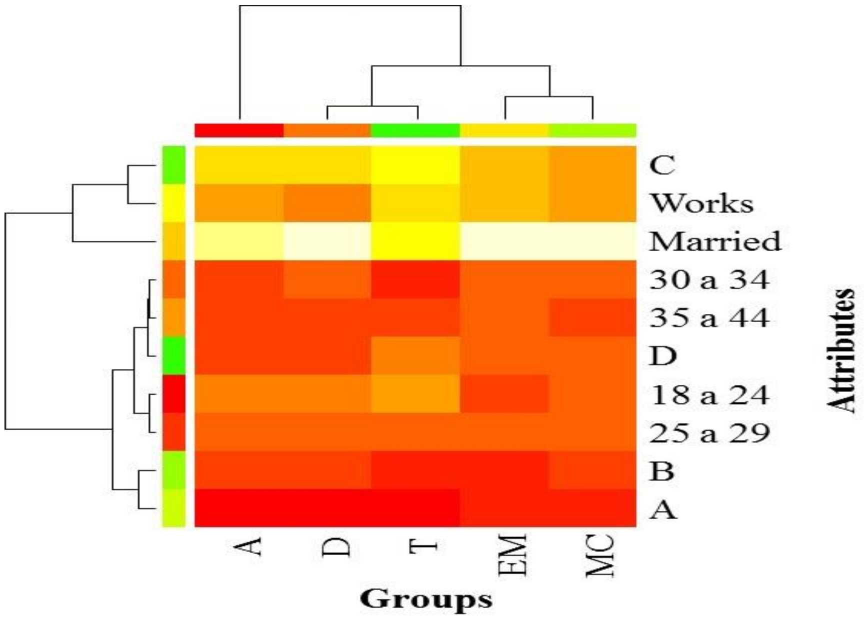

Figura 8 – Socio-economic traits by segment

3.3 RESULTS - GROUP ANALYSIS

The groups were characterized by *correspondence analysis* and *parallel sets* [1] (Venables & Ripley, 2002) and heatmap (Evergreen, 2017). Correspondence analysis is well known and is presented in Figure 9 in 3D: x, y and z axes. The figures in percentage (%) present the percentage of inertia of each axis (33.0 + 27.6 + 21.6 = 82.2 per cent of total inertia).

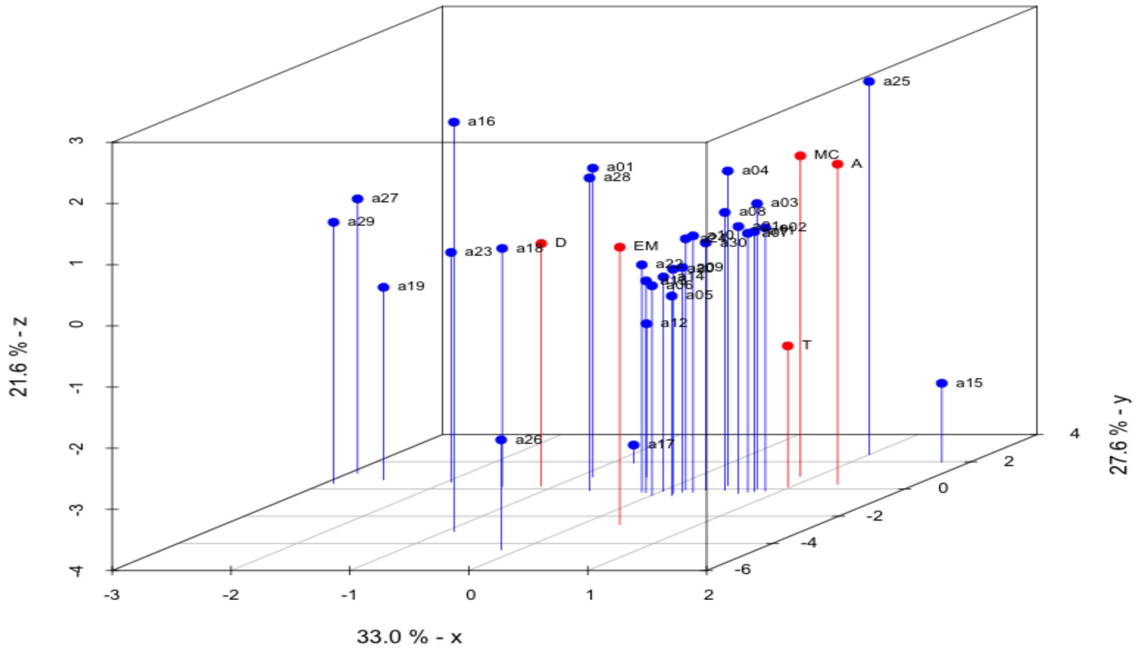

Figura 9 – Correspondence analysis in 3D

Here, in this case, Figure 9 serves to indicate that:

- There is a group on the right {MC-Modern Collaborative, D-Dedicated} and {T-Traditional} where:

[1] Parallel sets are also called parallel coordinate plots. They are described, for example, in Venables e Riplay (2002).

   - {MC, D} are situated above, denoting modernity;
   - {T} is below, denoting traditionalism;
- There is another group on the left {C-Carefree and ME-Modern Education}:
   - {C-Carefree} more intense on non-concern.

The heatmap graph shown in Figure 10 addresses the problem in another way: from the point of view of the original table. We will not perform a detailed analysis here, so as not to make the article more extensive.

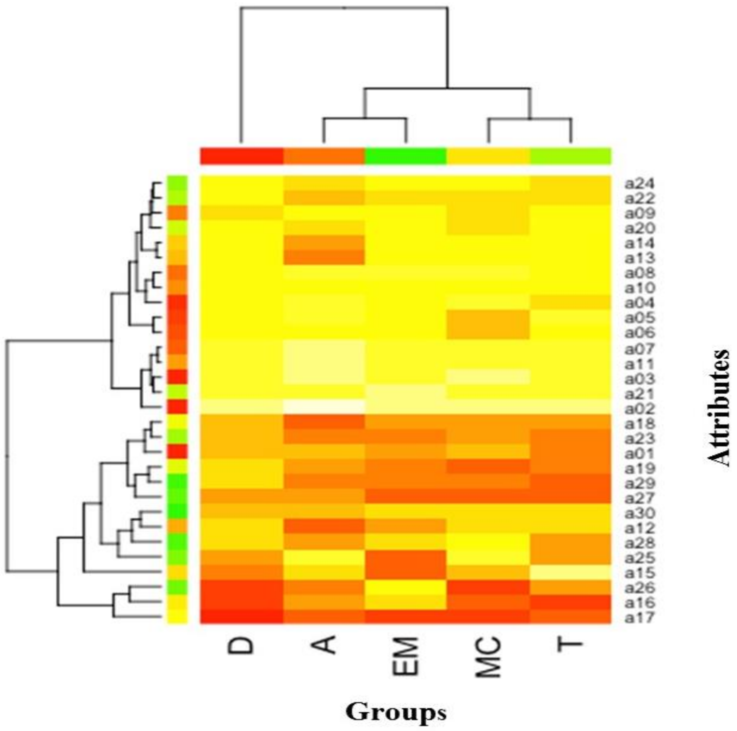

Figure 10 – Heatmap – Groups and attributes

Finally, we should focus on parallel sets. Initially, we will illustrate this graph (Figure 11) with its perhaps most well-known form: the database of the famous and tragic case of the Titanic transatlantic.

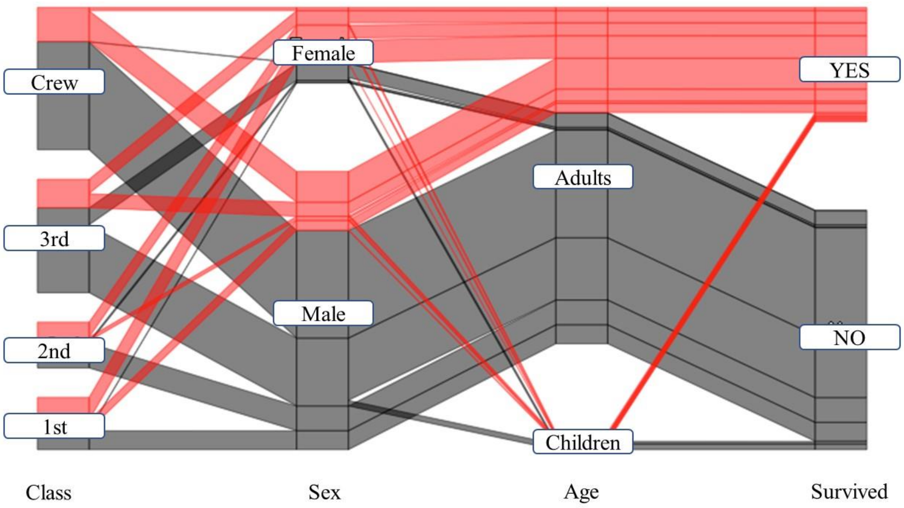

Figure 11 – Parallel Sets – Titanic

To the right of the graph (Figure 11) we have the variable Survived. From left to right we have the variables Class, Sex and Age. It is noteworthy, by examining the width of each section corresponding to the YES level (above right), that, generally speaking, women of first, second, and third adult classes (almost no children on board) survived. This type of chart allows fairly complex analyzes to be performed in a relatively simple way: a little training is enough.

Let's now illustrate how we proceeded with the parallel sets for the Dedicated group. Figure 12 shows the results for the Dedicated. The variable grp2 (Group 2 - Dedicated) indicates YES, for whom is in Group 2 and No for whom belongs to the other groups. It's an analysis of the kind that is sometimes called one against all.

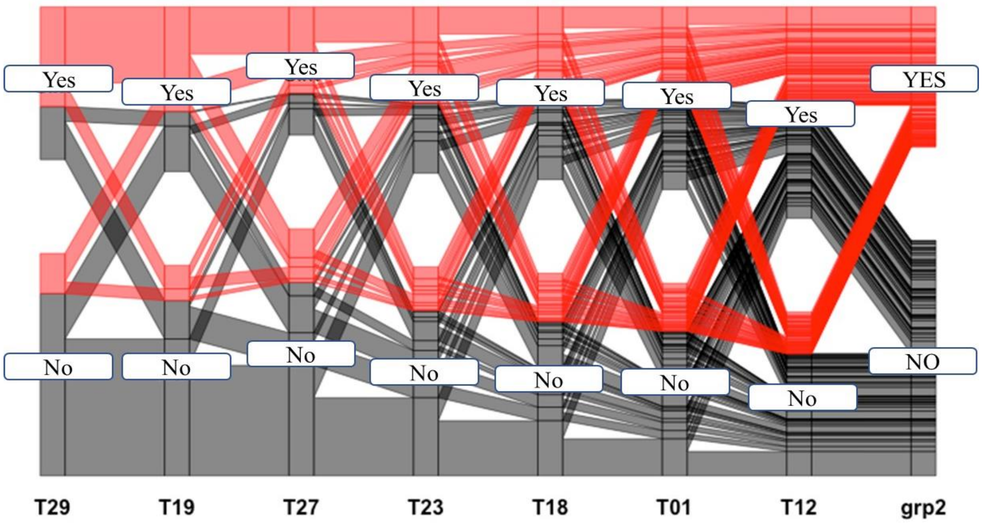

Figure 12 – *Parallel sets* – Dedicated

Attributes T29, T19 and T23 indicate that the interviewee does not think that she devotes little time to the family. T27 and T18 indicate that this interviewee does not think that she is a stressful mother and accesses the Web to know the news. The other groups were analyzed in a similar way.

4 CONCLUSION

The reader should have observed that practically no number was used, even though a purely quantitative analysis was performed. This shows the strength of Data Visualization. As we said earlier in Section 3, an interesting presentation of this technique is given in Knaflic (2015). In addition, our meta-heuristics needs improvement. The genetic algorithm, for 30 iterations, took almost four hours of processing. Even 10 iterations, which would make it possible to verify that near-optimality had been reached (affinity value close to 1, as shown in Figure 3), would take about a third of this (1h30min, for example). It is still an excessive time.

In spite of the lack of memory excess, even with a typical size of Marketing Research (4000 observations in Total and 5 x 800 for the set of the five countries, a total of 4000 + (5 x 800) = 8000 observations), an hour and a half can be considered much. It is worth improving the technique for a better processing time.

Finally, it should be emphasized that our algorithm does not apply only to disjoint sets, like the countries in our example. It could be applied, for example, to consumer groups defined by linking to certain brands, even if a consumer had some kind of attachment to more than one. This consumer would then be included in all sub-problems (brands) to which he was bound.

REFERÊNCIAS

BBA-MBA.net. (2016). Recuperado em 10 fev., 2016, de http://bba-mba.net/think-local-learn-global-act-glocal.htmlace

Evergreen, S. (2017). *Effective data visualization: The right chart for the right data*. Thousand Oaks: Sage Publications.

Gueddes, P. (1915). *Cities in evolution*. Londres: Williams.

Knaflic, C. (2015). *Storytelling with data: A data visualization guide for business professionals*. New Jersey: John Wiley & Sons.

Lee, B., & Kelley, R. (2010). Gibbs-inspired consensus-based. *Hybrid Global Segmentation*, ART Forum – American Marketing Association.

Levitt, T. (1983, may). The globalization of markets. *Harvard Business Review*. Retrieved for https://hbr.org/1983/05/the-globalization-of-markets?cm_sp=Article-_-Links-_-

Myers, J. (1996). *Segmentation and positioning for strategic marketing decisions*. Chicago: American Marketing Association.

Sá Lucas, L. (1983). Análise de Grupamento. *Revista Brasileira de Estatística*, *43*(172), out/dez., IBGE.

Smith, W. (1956, Jul.). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21, 3-8.

Viera, A., & Garrett, J. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5):360-3.

Venables, W., & Ripley, B. (2002). *Modern applied statistics with S*. New York: Springer-Verlag.

Wedel, M., & Kamakura, W. (2000). *Market segmentation: Conceptual and methodological foundations*. Dordrecht: Kluwer Academic Publishers.

[1] This work was presented at the seventh Brazilian Congress of Market Research, Opinion and Media of ABEP (held in April 2016), transformed into article by its authors, submitted to PMKT and approved for publication.