

Glocal Segmentation¹

Glocal Segmentation

Luiz Sá Lucas*

MC15 Consultoria, Rio de Janeiro, RJ, Brasil

Leonardo Soares

Instituto COPPEAD da Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil

Wagner Esteves

IBOPE Inteligência, Rio de Janeiro, RJ, Brasil

RESUMO

O artigo apresenta um algoritmo para criar uma Glocal Segmentation, que segmenta não só um 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.

ABSTRACT

The article presents an algorithm to create a Glocal Segmentation, that targets not only a number, but also parts, all this in a coordinated way and not with independent segmentations. The technique is applied to dummy data for five Latin American countries. Another unique aspect of the work is the fact that all the data analysis be done through data visualization.

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

Submissão: 07 junho 2016

Aprovação: 09 dezembro 2016

***Luiz Sá Lucas**

Mestrado em Otimização de Sistemas pelo Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE) da Universidade Federal do Rio de Janeiro. Consultor da MC15 Consultoria.

(CEP 22451-070, Rio de Janeiro, RJ, Brasil).

E-mail: luizsa.lucas@mc15.com.br

Endereço: Rua Vice-Governador Rúbens Berardo, 65, bloco A 406, 22451-070, Rio de Janeiro, RJ, Brasil.

Leonardo Soares

Doutorando em Administração de Empresas pelo Instituto COPPEAD da Universidade Federal do Rio de Janeiro. Consultor da K2 Achievements.

E-mail: leonardosoares@k2a.com.br

Wagner Esteves

Mestrando em Engenharia de Produção pela Universidade Federal Fluminense (UFF). Gerente de Atendimento Técnico e Planejamento do IBOPE Inteligência.

E-mail: wagner.esteves@iboointeligencia.com.br

1 INTRODUÇÃO

1.1 UMA COMUNIDADE GLOBAL?

Levitt (1983) tem sido citado como o criador do termo “globalização”, apresentado no artigo em referência. No entanto, o termo globalização já estava em uso geral desde a década de 1940 e, por economistas, desde a década de 1980. Isso não muda o fato de que Levitt é, provavelmente, o introdutor do termo em marketing. No artigo citado ele descrevia também o que entendia como uma Companhia atuando em uma Comunidade Global. Segundo ele, em uma economia internacional, à medida que os mercados tradicionais se tornavam saturados, era necessário atender à demanda dos novos.

Mas como fazer isso? Se o objetivo fosse atender a cada mercado separadamente, então “o jogo não valia o esforço” (tradução nossa). No entanto, segundo Levitt (1983), isso não era necessário: companhias bem administradas deviam enfatizar a oferta de produtos globalmente padronizados, que fossem avançados, funcionais, confiáveis e de baixo preço.

Ainda segundo Levitt (1983), uma força poderosa provocava uma convergência: a tecnologia. “Quase todo o mundo quer as coisas sobre as quais ouviram falar, viram ou experimentaram por meio das novas tecnologias” (tradução nossa). Cabe, aqui, notar que Levitt (1983) mostrava que cada mercado tinha suas características:

No Brasil, milhares de pessoas vêm, em enxame, da escuridão da Bahia pré-industrial, (sic) numa explosão em direção às cidades costeiras, para rapidamente instalar aparelhos de televisão em lotados barracões de papelão ondulado (sic) e, ao lado de Volkswagen danificados, fazer sacrifícios rituais de frutas e galinhas recém-sacrificadas a espíritos de macumba à luz das velas.” (tradução nossa)

Durante a luta fratricida em Biafra, ainda segundo nosso mestre Levitt (1983), “reportagens diárias na TV mostravam soldados carregando colegas ensanguentados enquanto tomavam Coca-Cola.” (tradução nossa).

Além de sua profunda ignorância sobre o além-mar, Theodore Levitt queria enfatizar que, qualquer que fosse a cultura, uma série de produtos básicos, padronizados e internacionalizados seriam consumidos sem qualquer necessidade de adaptação.

1.2 A COMUNIDADE GLOCAL

Embora a expressão, tal como conhecida hoje, não esteja explícita em seu livro *Cities in Evolution* (Gueddes, 1915), a ideia, aplicada a planejamento urbano, está no texto claramente expressa. Marcas globais devem ser únicas, ter um significado e longa duração em todos os mercados em que atuam. No entanto, uma atuação que leve em conta, como quase sempre é o caso, uma segmentação, tem de estar baseada nas necessidades e benefícios percebidos em cada *target/alvo* e na geografia, não só dentro de um país, mas também num conjunto deles. O *site* BBA-MBA.net apresenta uma expressão relativa ao caso bastante interessante: “Think local. Learn global. Act glocal” (BBA-MBA.net, 2016). A ideia é que companhias de todos os tamanhos, presentes em grandes e pequenos mercados, e que sejam percebidas como universais por seus consumidores, competindo num mercado global, devem focar seus recursos numa estratégia de sucesso em toda a sua geografia. Em outras palavras, seu futuro global só será atingido se a companhia for, em primeiro lugar, bem-sucedida a nível local.

Ora, não faz sentido e nem o texto citado afirma que se tenha que efetuar uma segmentação diferente para cada área ou *alvo/target* específico. Aí então iríamos recair naquilo que Levitt (1983), citou e comentou como “o jogo não vale o esforço”. Por outro lado, uma segmentação única que aborde apenas o total não irá contemplar aspectos específicos de áreas geográficas e/ou alvos que vierem a ser selecionados.

O conceito de segmentação em marketing foi criado em meados dos anos de 1950 (Smith, 1956). Excelentes textos sobre o assunto existem já há um bom tempo. Sobre segmentação e estratégia em marketing vale citar Myers (1996). Sobre técnicas de segmentação, um dos melhores, se não o melhor, ao nosso conhecimento, é o trabalho de Wedel e Kamakura (2000). Existem trabalhos mais novos, mas esses são trabalhos que resistem ao tempo.

A literatura existente trata do problema geral: segmentar o todo, mas como segmentar um todo e, ao mesmo tempo, segmentar suas partes de forma que todos os resultados sejam coerentes? É disso que trata nosso artigo.

2 METODOLOGIA

Imaginemos que uma companhia atue, primordialmente, em cinco países dentro da América Latina: Argentina, Brasil, Colômbia, México e Chile, e que, em sua atuação na área queira levar em consideração as atitudes das mães sobre a educação de seus filhos de até 4 anos.

Vamos supor também que um estudo aplique a uma amostra de 4.000 casos (800 em cada país), uma bateria de 30 atributos com relação a essa educação: métodos tradicionais ou mais modernos, divisão de trabalho com os maridos, acesso à *Web*, tempo dedicado às crianças, trabalhar *versus* dedicar-se apenas aos filhos, colocar bem cedo numa creche ou equivalente, permitir o consumo de refrigerantes etc. Os dados do nosso exemplo são artificiais e não conhecemos estudo algum nesses cinco países que aborde o assunto. Assim, os resultados deste trabalho são **sintéticos e não podem** ser considerados como **válidos** no caso real.

Este método é um meta-método, já que pode ser implementado com vários algoritmos diferentes, e está esboçado na Figura 1. A ideia de se realizar esse tipo de segmentação não é completamente nova: há alguns poucos anos, foi apresentado no Advanced Techniques Forum – ART Forum, da Associação Americana de Marketing, um artigo sobre o assunto (Lee & Kelley, 2010). A metodologia ali descrita é bastante diferente da nossa e fica próxima de um *gibbs sampler*. Pode-se utilizar, nesse caso, o *package clue* do *software R* (que trata de *cluster ensembles*), o método consome muita memória RAM e o tempo exigido de computação pode ser bastante grande, chegando até a quatro dias de processamento para efetuar 600 iterações.

Aperfeiçoando o algoritmo e utilizando apenas parte das informações, Lee e Kelley (2010) conseguiram um tempo de execução de até 23 horas, para 2000 iterações, dependendo do número de observações (os autores não informam esse número). Como dissemos, a ideia é segmentar o todo sem deixar de levar em conta as partes (os países). Aqui queremos utilizar todas as informações ao longo de todo o processo. A ideia, no nosso caso, é efetuar seis segmentações (*clusterings*), mas de alguma forma coordenando esses processamentos. Teríamos então seis problemas de *clustering*, com um nível coordenador, como no esquema mostrado na Figura 1.

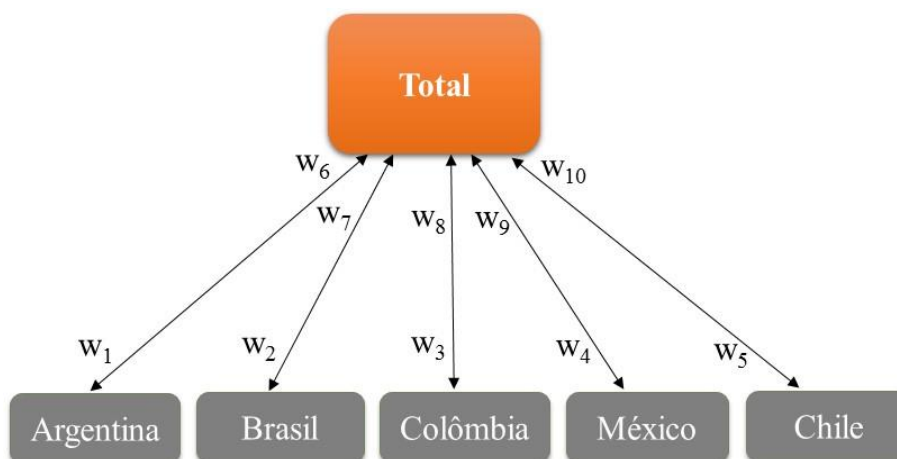


Figura 1 – Meta-heurística

No nível coordenador temos um *clustering* do Total. No nível inferior temos os *clusterings* de cada um dos cinco países. A coordenação é feita pelos valores de w .

Para melhor entendimento do algoritmo vamos introduzir aqui o conceito de mediana de um grupo de observações multivariadas. O conceito é ilustrado pela Figura 2. Na Figura 2 temos oito elementos. O elemento k é aquele que minimiza a soma das distâncias (euclidiana, manhattan etc.) entre os diversos elementos i, j etc. do grupo.

Sejam então m grupos, cada um deles semelhante ao da Figura 2. Sejam também n observações, em que, em cada uma delas, são medidas p características. Essas características podem ser quantitativas (como uma escala de concordância), ou categórica (como sexo, estado civil, p. ex.). Podemos então definir o problema das k -medianas. O objetivo desse problema é obter o conjunto de m medianas tais que a soma das distâncias de cada observação à sua mediana (a mediana do seu grupo) seja mínima. O problema é detalhadamente descrito, por exemplo, em Sá Lucas (1983).

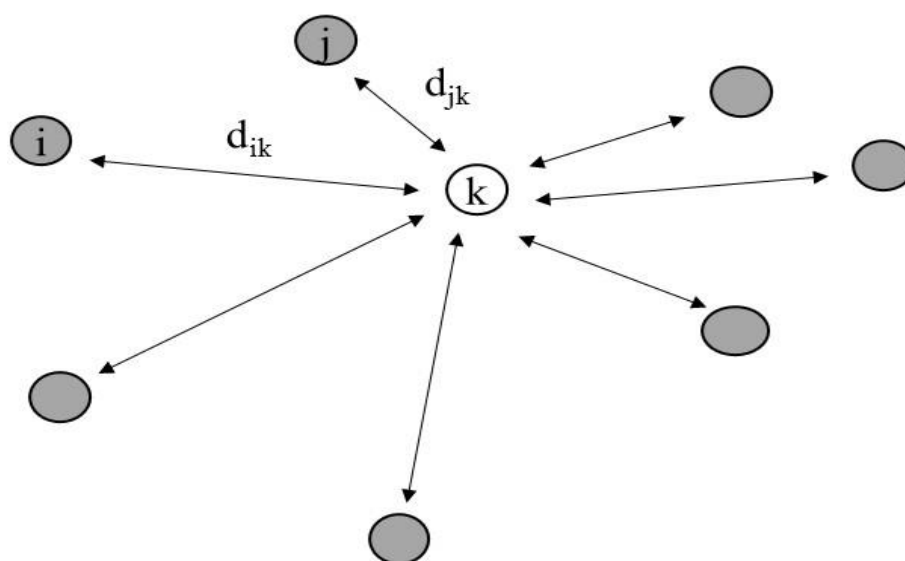


Figura 2 – Mediana de um grupo

Assim, poderá ficar mais claro o que significa cada w . Se temos m grupos, temos m medianas. Cada uma delas tem p observações (as características com base nas quais estamos efetuando a segmentação/clusterização). Desta forma, as m medianas de cada problema (na verdade de cada um dos seis subproblemas no nosso caso) podem ser consideradas como um conjunto de m observações que podem ser agregadas w vezes ao outro subproblema correspondente. A correspondência é dada pelas setas da Figura 1. A intensidade da coordenação é dada pelos valores de w que, é claro, são números inteiros. Imaginemos que m seja igual a cinco (cinco) grupos para o programa mestre (o problema *Total*). Se, por exemplo, w_1 é igual a 3 (três), estaremos agregando as cinco medianas do problema *Total* ao subproblema *Argentina* três vezes ($w_1=3$). Por outro lado, se w_6 for igual a 2 (dois), e se o subproblema *Argentina* tiver 6 (seis) grupos, estaremos agregando as seis medianas da *Argentina* ao *Total* duas vezes.

Note-se que as segmentações do *Total* e dos países não precisam ter o mesmo número de grupos. Vale a pena, no entanto, fazer com que o número de grupos no *Total* seja maior ou igual ao número de grupos dos subproblemas que, por sua vez, podem ter número de grupos diferentes entre si.

Os valores de w indicam o grau de influência de um problema no outro, isto é, a ênfase da coordenação entre eles. Cabe lembrar que os dados do Brasil (800 observações) são os mesmos no problema *Total* e no problema Brasil. A exceção são os dados derivados da atividade dos w .

O objetivo final da meta-heurística é que, por exemplo, a afinidade entre a segmentação do Brasil derivada do problema Brasil e a segmentação do Brasil obtida no problema *Total* seja a maior possível, embora essas segmentações não sejam necessariamente as mesmas.

A medida de afinidade entre as segmentações de cada país no subproblema e no total foi a medida pelo coeficiente *Kappa* (Viera & Garrett, 2005). Esse coeficiente varia de 0 (zero) – concordância devido ao acaso – a 1(um) – os segmentos são os mesmos. A média geométrica dos cinco *Kappas*, um para cada país, também irá se situar entre zero e um, e quão mais próximo esse valor for de 1(um) melhor a solução global obtida.

Resumindo: temos um problema de otimização bastante complexo que envolve definir os parâmetros de coordenação w , inteiros que, resolvendo em cada iteração seis problemas de *clustering*, maximizam a média geométrica dos *Kappas* de cada país. A solução final do problema é dada pela segmentação do problema Total. No nosso caso, a solução do problema, ou seja, o conjunto dos valores de w_i , foi obtida por meio de um algoritmo genético.

Como é o caso dos algoritmos genéticos, a solução w obtida não é necessariamente ótima, mas a concordância global, por ser uma média geométrica de *Kappas*, pode ser avaliada por seu valor final: quanto mais próxima de 1(um) melhor.

3 RESULTADOS

Os resultados serão aqui apresentados também de uma maneira original. Um conjunto de técnicas, algumas seculares outras bastante recentes, denominado de Visualização de Dados (*Data Visualization*) - VD, vem sendo cada vez mais utilizado na literatura. Uma apresentação interessante dessa técnica é apresentada em Knaflitz (2015). Nossa apresentação abordará os dados dessa forma, o que dará ao leitor a oportunidade de examinar sua utilidade prática.

No entanto, embora a VD seja normalmente usada para apresentação dos dados a uma audiência (gerência da empresa, clientes etc.), vamos adotar aqui uma estratégia diferente. A visualização será utilizada para a análise dos dados, não envolvendo necessariamente a apresentação para qualquer audiência. Isso vai permitir que, aqui, o analista possa usar os mesmos gráficos mais complexos para a análise, como se verá a seguir.

3.1 RESULTADOS - ALGORITMO GENÉTICO

A Figura 3 mostra a evolução do algoritmo genético. Esse algoritmo trabalha, em cada iteração, com várias soluções (aqui denominadas de população). Em cada iteração temos, para cada população:

- O melhor valor obtido até a iteração (o valor máximo);
- O valor médio;
- O terceiro quartil;
- O valor mediano;
- O primeiro quartil;
- O valor mínimo.

No gráfico, as soluções em cada iteração estão apresentadas na ordem descrita acima. No algoritmo, utilizamos uma população de 50 soluções em cada iteração (o máximo de iterações foi igual a 30), conforme mostra a Figura 2.

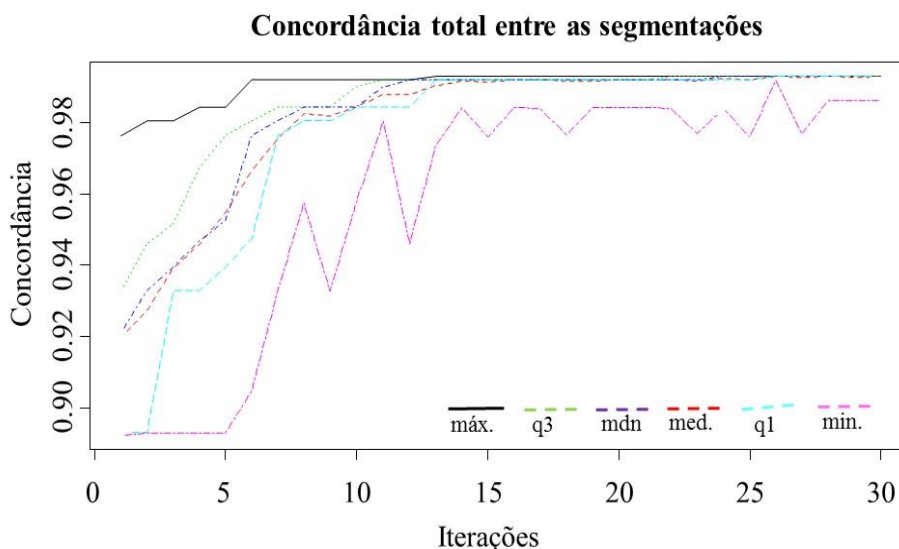


Figura 3 – Evolução da média geométrica dos *Kappas*

É possível notar que, desde a sétima iteração, um valor muito próximo do melhor já havia sido obtido, e esse melhor valor fica bem próximo de 0,99. Esse número corresponde à média geométrica dos *Kappas* de cada país, e é como se vê, praticamente igual a 1 (valor ótimo).

3.2 RESULTADOS – GRUPOS OBTIDOS

A solução obtida para o Total e cada subsolução para cada país, em cinco grupos, estão apresentadas na Figura 4.

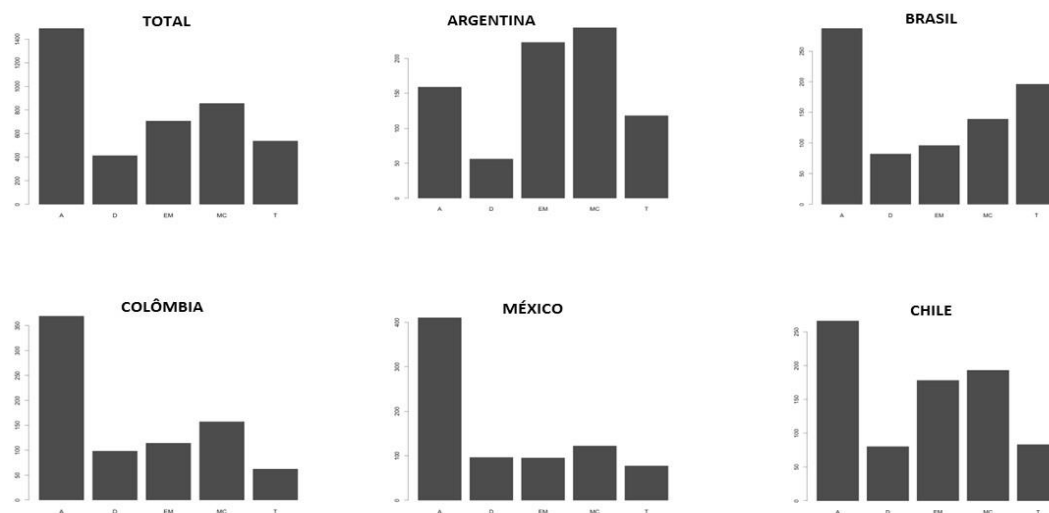


Figura 4 – Segmentação em cinco grupos

Nota-se que a solução global total, acima à esquerda, tem números de observações em cada grupo diferentes de cada país individual. Assim, a incidência dos mesmos grupos em cada país é diferente: uma estratégia global, única, pode ser adaptada para cada país, enfatizando os alvos desejados.

Na Figura 4 também são indicados os grupos, cuja análise, é claro, só precisa ser efetuada para o Total. Assim temos as mães:

- D – Despreocupada;
- A – Antenada;

- MC - Moderna Colaborativa;
- T – Tradicional;
- EM – Educação Moderna.

Esses rótulos foram criados para os grupos a partir de eixos básicos detectados em nossos dados sintéticos. Repetimos, **os dados são artificiais** e não permitem uma extensão para o caso real. De qualquer forma, os grupos são:

- Despreocupadas: são mães que acham que dedicam tempo adequado para os filhos, para si e para o marido, e não acham o trabalho de mãe estressante;
- Antenas: mães muito preocupadas com os filhos, acham o trabalho de mãe estressante, acessam a *Web* para saber as novidades etc.;
- Modernas Colaborativas: os pais cuidam tanto das crianças quanto elas, as mães, e gostam que os filhos descubram o mundo por si mesmos;
- Tradicionais: ao contrário das Modernas, tendem a utilizar métodos de criação mais tradicionais, sem maior experimentação;
- Educação Moderna: cuidam das crianças de forma diferente da que foram criadas pelos pais.

Mais adiante, na análise, veremos como a caracterização das mães foi identificada. A seguir exploraremos o tamanho relativo dos grupos por país.

A Figura 5 corresponde a um *treemap*, que é interessante para identificar os tamanhos relativos de cada parte no total. O número de observações em cada país é igual: 800 observações. Em cada país, o(s) grupo(s) maior(es) tem maior área e cor mais intensa:

- Na Argentina, ressaltam-se as Modernas: Colaborativas e Educação;
- No Brasil, as Antenas e as Tradicionais;
- No Chile, as Antenas e as Modernas (as duas, como na Argentina);
- Na Colômbia, as Antenas.
- No México, também as Antenas.

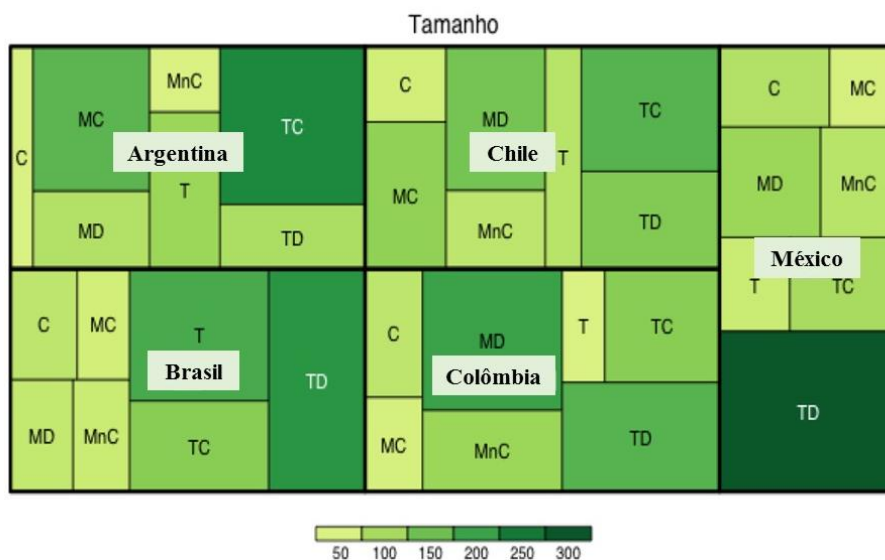


Figura 5 – Tamanho dos grupos por país

Cabe notar que, no Total (Figura 4 - Gráfico1), os grupos dominantes são as Antenas. Em menor grau, as Modernas: Colaborativas e Educação.

No *treemap* é possível notar uma dificuldade que vários autores associam a medidas que são expressas em área: na Argentina, no *treemap*. É um pouco mais difícil distinguir que as Modernas Colaborativas têm maior incidência do que as Educação Moderna, enquanto essa leitura fica clara nos gráficos de barra da Figura 4.

É possível também fazer esse tipo de análise por meio de um gráfico de mosaico que, nesse caso, tem um teste de hipótese embutido (Figura 6). Quanto mais forte o quadrado em cor azul, mais o grupo está mais acima da média geral. Quanto mais vermelho, abaixo.

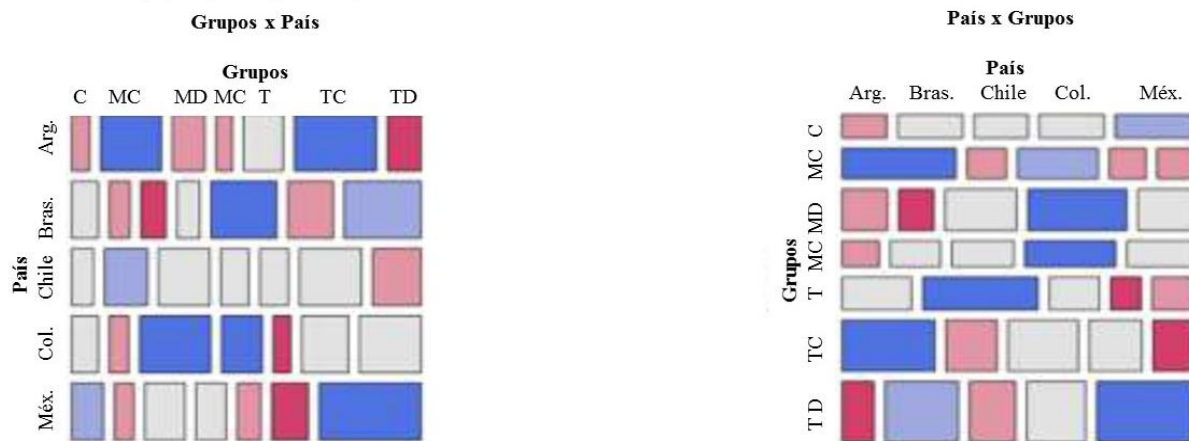


Figura 6 – Tamanho dos grupos por país

É possível também fazer uma análise para o total da composição socioeconômica dos grupos. Isso aqui é feito por meio de um múltiplo de gráficos de mosaico (sem teste de hipótese). Não há variação excessiva em grupos de idade, estado civil, emprego e SEC (Figura 7).

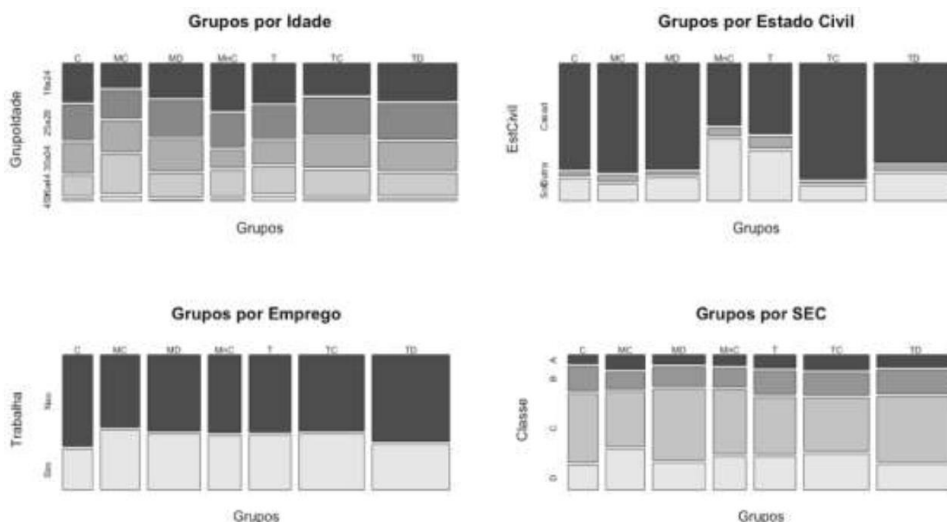


Figura 7 – Traços socioeconômicos por segmento

Finalmente, a mesma análise pode ser feita por um *heatmap* (Figura 8), que conglera não só os segmentos, como também os níveis socioeconômicos, mas novamente se verifica que não há uma diferença tão grande assim nos demográficos. Uma descrição de *heatmaps* pode ser encontrada em Evergreen (2017).

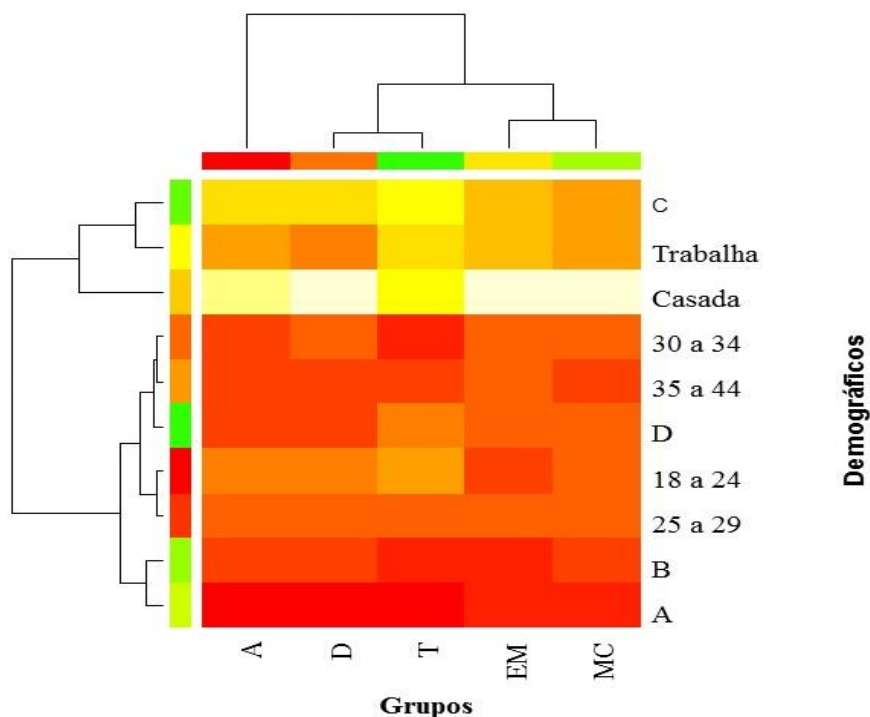


Figura 8 – Traços socioeconômicos por segmento

3.3 RESULTADOS – ANÁLISE DOS GRUPOS

A caracterização dos grupos foi feita por meio de análise de correspondência e *parallel sets*¹ (Venables & Ripley, 2002) e *heatmap* (Evergreen, 2017). A análise de correspondência é bastante conhecida e é apresentada na Figura 9 em 3D: eixos *x*, *y* e *z*.

Aqui, neste caso, a Figura 9 serve para indicar que:

- Existe um grupo à direita {MC-Moderna Colaborativa, Antenada} e {T-Tradicional} em que:
 - {MC, A} se situam acima, denotando modernidade;
 - {T} fica abaixo, denotando tradicionalismo;
- Existe outro grupo à esquerda {D-Despreocupada e EM-Educação Moderna}:
 - {D-Despreocupada} mais intensa no quesito despreocupação.

¹ *Parallel sets* são também denominados de *parallel coordinate plots*. Estão descritos, por exemplo, em Venables e Ripley (2002).

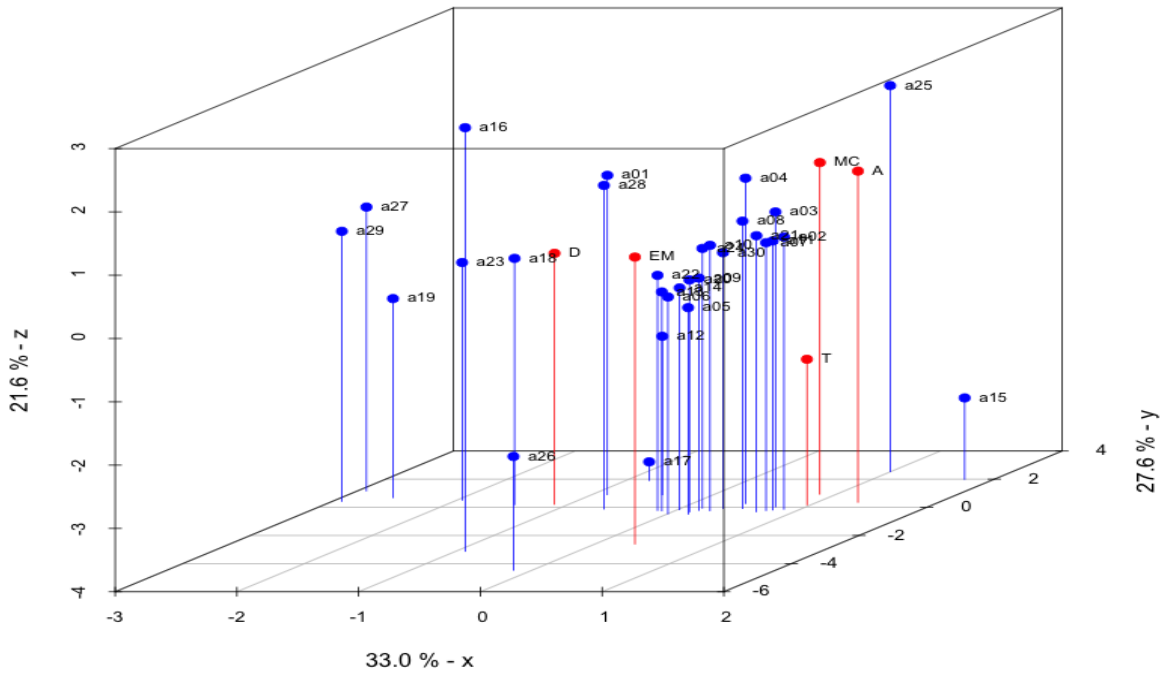


Figura 9 – Análise de correspondência em 3D

O gráfico de *heatmap* mostrado na Figura 10 aborda o problema de outra forma: do ponto de vista da tabela original. Não vamos aqui fazer essa análise detalhada, para não tornar o artigo mais extenso.

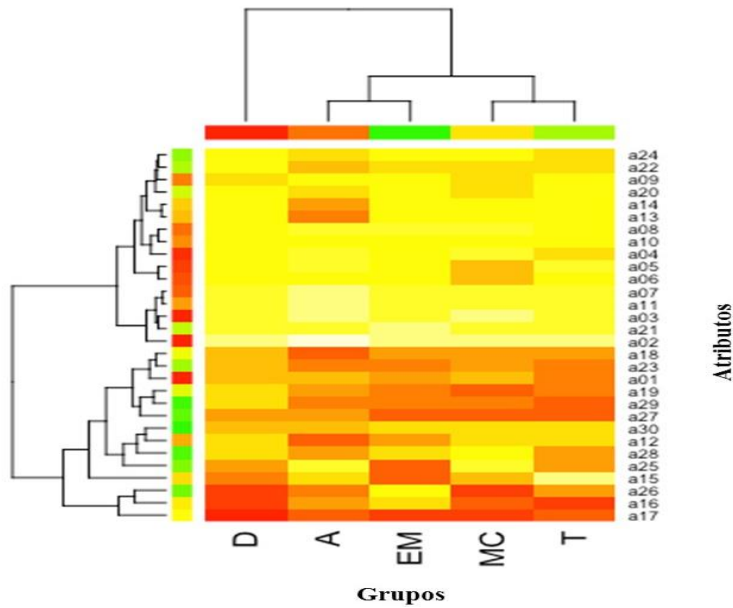


Figura 10 – Heatmap – Grupos e atributos

Finalmente, cabe focar o *parallel sets*. Inicialmente, vamos ilustrar esse gráfico (Figura 11) com sua forma talvez mais conhecida: o banco de dados do famoso e trágico caso do transatlântico Titanic.

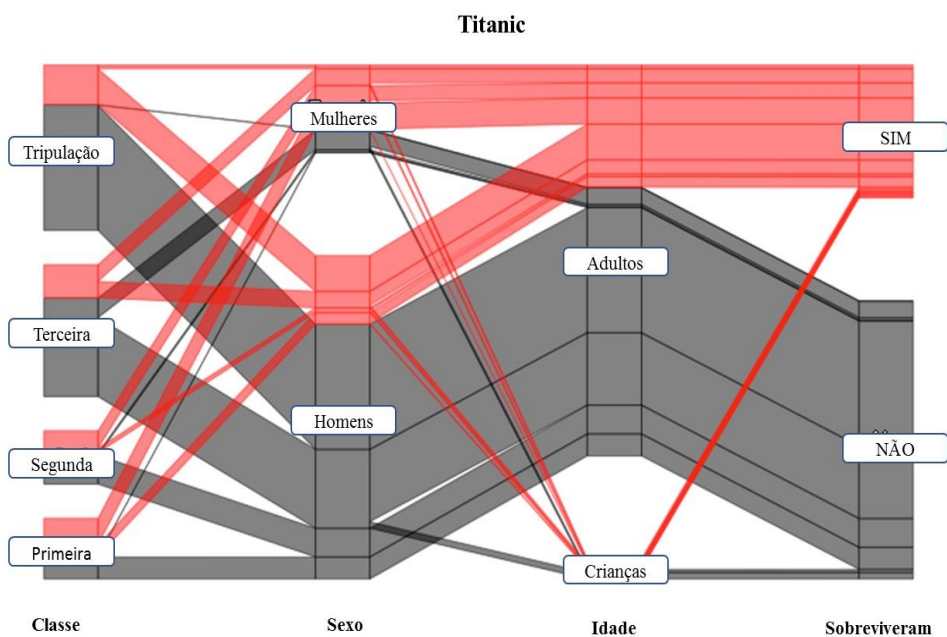


Figura 11 – Parallel Sets – Titanic

À direita do gráfico (Figura 11) temos a variável *Sobreviveu*. Da esquerda para a direita temos as variáveis *Classe*, *Sexo* e *Idade*. Nota-se, examinando a largura de cada trecho correspondente ao nível *Sim* (acima à direita) que, de forma geral, as mulheres de primeira, segunda e terceira classes e adultas (quase não havia crianças no navio) sobreviveram. Esse tipo de gráfico permite análises bastante complexas de forma relativamente simples: basta um pouco de treino.

Vamos agora exemplificar como procedemos com o *parallel sets* para o grupo Antenada. A Figura 12 apresenta os resultados para a Antenada. A variável *grp2* (Grupo 2 – Antenada) indica Sim, para quem é do Grupo 2 e Não para quem pertence aos outros grupos. É uma análise do tipo que, às vezes, se chama de *one against all*.

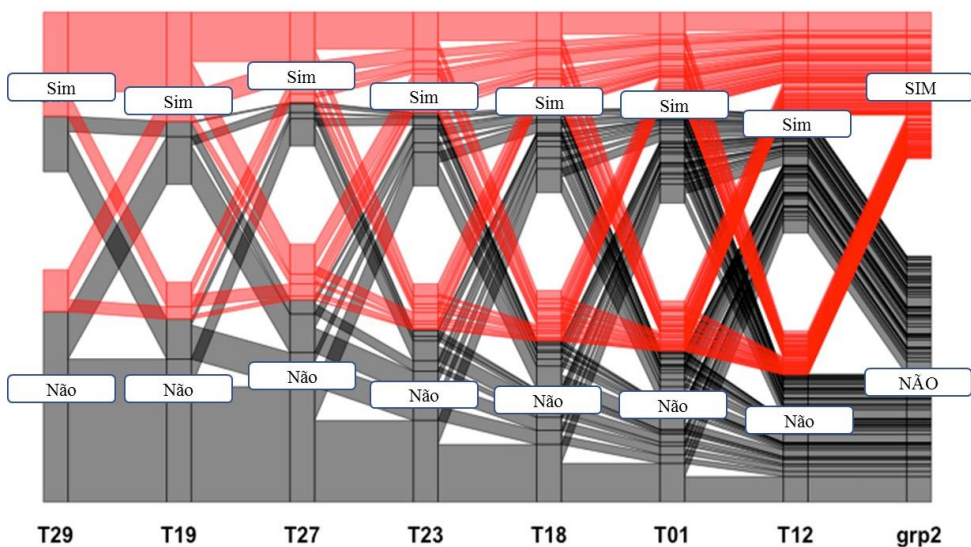


Figura 12 – Parallel sets – Antenada

Os atributos T29, T19 e T23 indicam que a entrevistada não acha que dedica pouco tempo à família; T27 e T18 indicam que essa entrevistada não acha que é uma mãe estressante e acessa a *Web* para saber as novidades. Os demais grupos foram analisados de forma semelhante.

4 CONCLUSÃO

O leitor deve ter observado que praticamente não se utilizou número algum, mesmo sendo realizada uma análise meramente quantitativa. Isso mostra a força da visualização de dados. Como dissemos anteriormente, uma apresentação interessante dessa técnica é apresentada em Knaflic (2015).

Além disso, a nossa meta-heurística precisa de melhorias. O algoritmo genético para 30 iterações, levou quase quatro horas de processamento. Mesmo 10 iterações, que permitiriam verificar que uma solução quase ótima havia sido atingida (valor de afinidade próximo de 1, conforme apresentado na Figura 3), levariam cerca de um terço disso (1h30, por exemplo). Ainda é um tempo excessivo.

Apesar de não apresentar problema de excesso de memória, mesmo com um tamanho típico de Pesquisa de Marketing (4000 observações no Total e 5 x 800 para o conjunto dos cinco países, num total de $4000 + (5 \times 800) = 8000$ observações), uma hora e meia pode ser considerado muito. Vale melhorar a técnica para um tempo de processamento melhor.

Finalmente, cabe enfatizar que nosso algoritmo não se aplica apenas a conjuntos disjuntos, como os países do nosso exemplo. Poderia ser aplicado, por exemplo, no caso de grupos de consumidores definidos pela vinculação com determinadas marcas, mesmo que um consumidor tivesse algum tipo de vinculação com mais de uma delas. Esse consumidor seria incluído então, em todos os subproblemas (marcas) a que ele estivesse vinculado.

REFERÊNCIAS

- BBA-MBA.net. (2016). Recuperado em 10 fev., 2016, de <http://bba-mba.net/think-local-learn-global-act-glocal.html>
- 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 Este trabalho foi apresentado no 7º Congresso Brasileiro de Pesquisa - Mercado, Opinião e Mídia da ABEP (realizado em abril de 2016), transformado em artigo por seus autores, submetido à PMKT e aprovado para publicação.