

# Using graphical statistics to better understand market segmentation solutions

**Sara Dolnicar**

School of Tourism, University of Queensland, Brisbane, Australia

**Friedrich Leisch**

Institute of Applied Statistics and Computing

University of Natural Resources and Life Sciences, Vienna, Austria

*This is a preprint of an article that has been accepted for publication in the International Journal of Market Research (2013).*

Market segmentation lies “at the heart of successful marketing” (McDonald 2010), yet, market segmentation solutions are not trivial to interpret, especially if consumers are segmented using post-hoc or a posteriori or data-driven segmentation where several consumer characteristics are analyzed simultaneously to identify or construct market segments. In fact, 65 percent of marketing managers admit to having difficulties with the interpretation of data-driven market segmentation solutions. In this study we develop novel ways of visualizing segmentation solutions using graphical statistics methodology. The proposed plots help academics and practitioners to interpret complex market segmentation solutions, thus improving the practical usability of market segmentation, reducing the risk of misinterpretation and contributing to closing the much-lamented “theory-practice divide” in market segmentation.

## Introduction

Market segmentation means “grouping potential customers into sets that are homogeneous in response to some elements of the marketing mix” (Choffrey & Lilien 1980). Smith introduced the concept in 1956, and Haley (1968) later extended segmentation bases to include psychographic elements such as purchase motives and benefits sought. Market segmentation is an extremely popular and commonly used tool in strategic marketing. Not only do organizations focus their marketing activities on target segments which have emerged from segmentation studies, academics also use market segmentation frequently to develop knowledge: according to Zins (2008), for example, about five percent of all tourism publications are segmentation studies.

Market segmentation approaches fall into one of two categories: the *a priori* (Mazanec 2000) or commonsense (Dolnicar 2004) segmentation approach, where managers use a single segmentation criterion, for example, age, to split consumers into homogeneous groups, or the post hoc (Myers & Tauber 1977), *a posteriori* (Mazanec 2000) or data-driven (Dolnicar 2004) segmentation approach, where exploration of more complex sets of variables are undertaken.

*A priori* market segmentation is methodologically simple and intuitively easy to understand, however, a business or destination basing its strategy on *a priori* segmentation approaches may find that it is outperformed by companies using more relevant, for example, psychographic rather than demographic, criteria for segmentation. Such criteria are typically more complex because they consist of sets of variables which together form the so-called *segmentation base* (Wedel & Kamakura 1998). In such cases the data-driven segmentation approach is required.

Results from data-driven segmentation are not trivial to interpret. Substantial evidence points to the fact that managers have serious difficulties understanding and interpreting results correctly (Nairn & Bottomley 2003; Bottomley & Nairn 2004). Typical reasons for these difficulties include poor understanding of segmentation principles and processes (Dibb & Simkin 1994, 1997; Doyle, Saunders & Wong 1985; Palmer & Millier 2004) as well as the existence of misunderstandings surrounding the segmentation process (Greenberg & McDonald 1989). This is at least in part due to segmentation literature heavily emphasizing methodology aspects (Wind 1978) whilst failing to provide guidance to managers regarding interpretation and validity (Dibb & Simkin 1997). The mismatch between academic work in market segmentation and the application of market segmentation in industry is commonly referred to as the “theory-practice divide”.

An empirical study (Dolnicar & Lazarevski 2009) of managerial understanding of data-driven market segmentation solutions confirms these results. Of the 176 marketing managers surveyed, 79 percent acknowledge that they need to understand the technical details of market segmentation studies in order to translate segmentation solutions into marketing action. Forty-three percent state that the segmentation study results they are typically provided with are not self-explanatory and require additional explanation. The following comments were made in response to an open ended question about how results are typically reported to them:

*“A voluminous report from a company such as XXX”*

*“...as a long report that usually contradicts the results”*

*“Groupings of respondents are provided on factors that are too extensive in the range they cover. eg age groups or type of employment are too wide - greater precision is required.”*

*“Has varied on many occasions over the years, rarely with a clear Executive Summary”*

*“In a powerpoint presentation with a slick handout”*

*“...in a rushed slap hazard fashion with the attitude that (sic)“leave the details to us”...”*

*“...mostly incomplete, unless I define the survey myself, there is too much understood (sic) or unclear. The result is usually ranged in numbers and percentages across a few (up to say 10) variables, but mostly insufficiently conclusive.”*

*“...report or spreadsheet...report with percentages”*

*“...often meaningless information”*

When asked whether they feel that the segmentation process is a “black box” to them (where “black box” is defined as a place where data goes in one end and a segmentation solution comes out of the other end without any clear process in between), 71 percent respond saying “yes”. It can be concluded from these responses that, while managers believe that fully understanding the details of a segmentation solution is necessary to translate the solution into successful marketing action, only a minority feel they actually do fully understand the results they receive. Clearly, the situation is not satisfactory if managers are not able to use these segmentation results and translate them, correctly, into marketing action.

The lack of understanding of market segmentation solutions can have two serious negative consequences: (1) managers may misinterpret results, leading to suboptimal strategic decisions; or (2) managers may be aware of their lack of understanding, and consequently avoid complex market segmentation methods entirely, thus missing out on an opportunity to improve their marketing. Improving managerial understanding of complex market segmentation solutions is a necessary prerequisite for the successful practical implementation of market segmentation.

The aim of this study is to contribute to reducing the theory-practice divide by developing novel ways of making data-driven market segmentation solutions easier to understand for users, both academics and practitioners, thus reducing the risk of misinterpretation and increasing the likelihood that results from segmentation studies are translated into marketing action in an optimal way. Specifically, graphical statistics methodology will be applied to develop three visualizations representing key insights that need to be understood by a user of a market segmentation solution, namely (1) how distinct segments resulting from a specific solution are, (2) what the key characteristics of each segment are and (3) which characteristics best discriminate between segments.

Note that we do not suggest that segmentation solutions are never summarized well – visually or otherwise – to make them understandable and actionable by users. It is impossible, however, to access the many market segmentation analyses that are conducted regularly by market research companies for their clients. Based on our review of publicly accessible reports of applied segmentation studies, segmentation solutions are typically not presented in ways that are easy for users to interpret. In addition, focus is mainly if not exclusively on the profiles of market segments with respect to the segmentation base and selected background variables or descriptive. Practically never is information presented about the extent of separation of segmentation solutions. Such information is, however, critical to users because it helps them understand whether the segmentation presented to them has revealed naturally occurring distinct market segments (which is the optimal case) or whether the segments, while being of managerial value, are artificially created. Overall, we therefore argue, that the visualisations we propose have the potential of substantially improving the understanding of segmentation solutions by users.

## Graphical statistics

Data visualization is an integral part of almost every statistical data analysis. Exploratory analysis is impossible without graphics. In the field of statistics, graphical visualization has a long history, and its importance for understanding complex relationships in a data set is widely acknowledged. For example, the American Statistical Association has a separate special interest section for Statistical Graphics which jointly publishes one of the leading journals in the field, the *Journal of Computational and Graphical Statistics*. In addition, Springer Verlag recently published the next volume in its statistical handbook series, the 1,000-page *Handbook of Data Visualization* (Chen, Härdle & Unwin 2008). The books by Tufte (1983, 1997) and Cleveland (1993) are considered classics, and laid the foundations for the now-flourishing field. Wilkinson (2005) goes further and develops a complete “grammar of graphics,” which can help develop new graphics, based on pre-existing basic building plots. Leisch (2008) offers a comprehensive survey on visualization techniques specifically suited for use in conjunction with cluster analysis and mixture models.

Although many scientific journals are still printed in black and white, their online versions of articles are mostly displayed in color. Consequently, statistical data visualization is a highly active research field, and applied articles in statistical journals typically make heavy use of graphics to communicate results. In many cases, good graphics are much better suited for finding patterns in high-dimensional data than are tables of summary statistics or model coefficients. While aesthetics are important (a good-looking figure will draw more attention than a bad-looking one), laws of visual perception are considered to be of primary importance. Programmable computing environments such as R (R Development Core Team 2012) are particularly useful as they allow researchers to develop new graphical methods (Murrell 2006) and share these techniques easily with their peers. All figures in this paper were created using R with the extension package flexclust (Leisch 2006), which is freely available from the Comprehensive R Archive Network (CRAN) at <http://cran.R-project.org>.

Despite the significant advances in the field of graphical statistics, however, the importance of visualization is not yet widely acknowledged by some applied disciplines. In a literature review within tourism (the area identified previously as using segmentation studies heavily) which included all segmentation studies published in the three main outlets for tourism segmentation research (*Journal of Travel Research*, *Tourism Management*, and *Journal of Travel and Tourism Marketing*) between January 2000 and August 2007, it was found that only 44 percent of studies included graphics, of which only 26 percent were actually related to the segmentation solution. The review also considered how the authors demonstrated the quality or managerial usefulness of the proposed segmentation solutions. The vast majority of authors (80 per cent) discussed the managerial usefulness of their results in text form only, and six percent of studies did not discuss managerial usefulness at all. Not a single study in this review of tourism literature utilized graphics as a means of demonstrating the managerial usefulness of the study’s results.

A few articles illustrate this point particularly well. Chung, Oh, Kim and Han (2004) conduct a valuable study aimed at gaining more insight into hotel guests by segmenting them. To report results from their commonsense segmentation analysis using the criteria of business versus pleasure, and package travel versus group and independent travel, they present to the

reader three tables containing a total of 570 numbers. When reporting results from their data-driven benefit segmentation they present eight tables including a total of 272 numbers which the user would be expected to inspect in order to fully understand the meaning of the solution. Less problematic is the presentation of results by Lo, Law and Cheung (2011) who also choose a table and choose to report average values with three after comma digits, but only present 30 numbers (10 variables for three segments) for the segmentation base and another 69 numbers for the additional background variables to describe the segments. This kind of reporting of segmentation results is typical in academic studies. Fuller and Matzler (2008) are among the few who offer the reader a simple chart providing a visual profile of the results segments. Unfortunately, the fact that the article was printed in black and white makes the chart more difficult to read than it would be in color, and the line chart may not necessarily be the optimal choice for visualizing differences between segments, however, this one chart makes is significantly easier for the user to quickly understand the differences between segments. The contrast between the results reporting of Chung et al. (2004) and Fuller and Matzler (2008) illustrates exactly the key problem we are aiming to solve. Note, however, that the visualization in Fuller and Matzler only addresses one of the three key pieces of information which are required for a user to be able to fully assess value and meaning of a market segmentation solution.

To the best of the authors' knowledge, the *International Journal of Market Research* is the only marketing journal which has acknowledged the importance of graphical visualizations of statistical analyses by publishing an article (Bock 2010) which proposed a graphical visualization of results from correspondence analysis for the very purpose of reducing the potential for misinterpretations.

It can be concluded that the importance of visualization for understanding complex statistical findings is not yet generally acknowledged in applied areas despite the fact that they rely heavily on statistics to solve business problems. This is particularly surprising in the context of market segmentation because segmentation methods (cluster analysis and mixture models, for example) are highly exploratory in nature, thus making them particularly well suited to analysis using graphical statistics. As Cleveland (1993) puts it: "Visualization is a necessary part of data analysis. Tools matter".

## **Visualizing Data-Driven Market Segmentation Solutions**

Data-driven market segmentation studies are typically conducted in two steps. First, the segmentation base is partitioned to construct market segments. The resulting segments are then interpreted using the segmentation base profile to understand what each segment stands for. For example, if consumers are segmented using their preferences for specific car characteristics, the actual grouping or segmentation task will result in a number of segments each of which will have a distinct profile with respect to desirable car characteristics. One segment may want a safe, large car, but not place much importance on low fuel consumption or engine performance, another segments may be interested in an attractive, fast car which offers a range of extras internally.

At this stage in the segmentation process, managers need support in understanding:

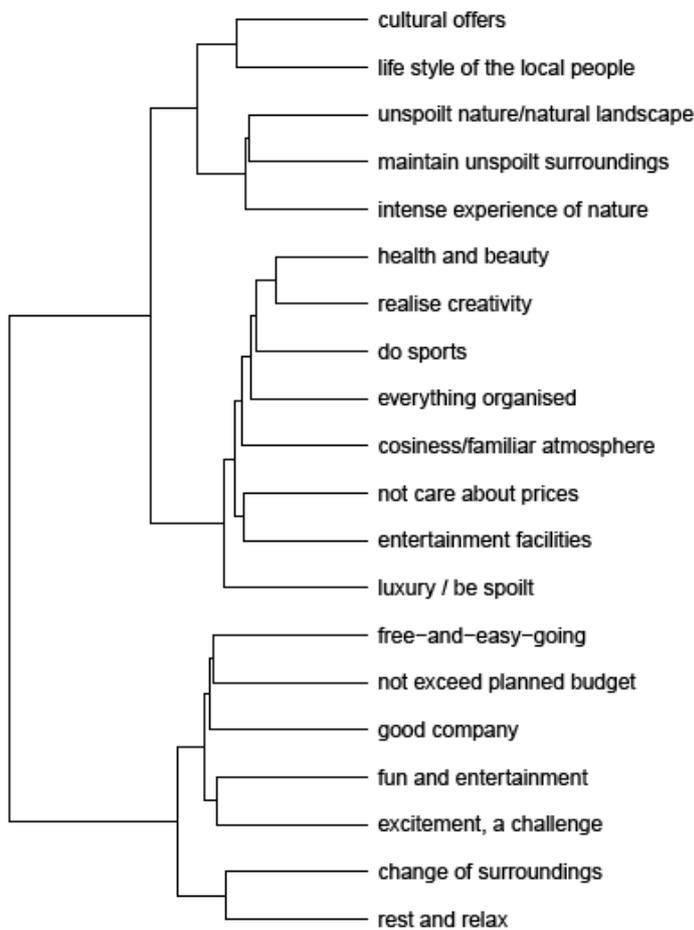
- (1) whether any particular segmentation solution leads to distinctly different segments with respect to the segmentation base;
- (2) how each of the segments is characterized with respect to the segmentation base; and
- (3) which variables from the segmentation base are the most effective for discriminating between segments.

In the second step, the market segments obtained are compared using further variables, such as general socio-demographic information, additional product preferences, or information sources most frequently used. This analysis may suggest, for example, that members of the “fast car” segment are more frequently young women with high incomes. The most managerially attractive segments are those that are distinctly different from others, as this enables management to develop a customized marketing mix to target chosen segments effectively and efficiently. Managers should have information that enables them to easily identify the differences between segments and to understand the nature of these differences.

In this study new visualization techniques are presented for communicating to managers the key meanings of segmentation solutions (the first step outlined above). Although few authors use graphics in the second step of segmentation analysis, visualizations for descriptive statistics are not novel and therefore do not need to be discussed in the present study.

Newly developed visualizations are illustrated using (1) artificial data of known structure to demonstrate how those visualisations look in prototypical cases and (2) replicating an applied segmentation analysis presented by Dolnicar and Leisch (2008) for which data was collected through an online panel using a 30 minute questionnaire completed by 1000 respondents. The items used as the segmentation base are a set of statements indicating the vacation preferences of respondents (e.g. “I want to rest and relax”, or “I want to do sports”). Figure 1 lists the complete set of items in an order which best represents the similarities between items. This sorting is required for the construction of Figure 4 which will be discussed at a later point.

**Figure 1:** Hierarchical segmentation of the 21 variables in the segmentation base



### Illustration Using Artificial Data

We create two artificial data sets which represent the two extremes that can theoretically occur in customer survey data. One case is that of natural clustering, the other one that of constructive clustering (Dolnicar and Leisch 2010). The case of natural clustering which is characterized by real, well separated density clusters existing in the data, is illustrated for a simple two-dimensional case in the first column and first row of Figure 2. The case of constructive clustering is illustrated in the first column and second row of Figure 2. As can be seen, cases are spread across the entire data space. No density structures exist; there is not even any kind of pseudostructure that could guide the data analyst to select a specific segmentation solution. This is not an uncommon situation with consumer data given that consumers typically vary across all variables of interest, rather than neatly falling into categories (Mazanec 1997, Wedel and Kamakura 1998, Dolnicar and Leisch 2010).

The data were clustered using the neural gas algorithm (Martinetz, Berkovich & Schulten 1993). Note that the particular choice of clustering algorithm is not critical to the remainder of the paper. We could have chosen a standard k-means segmentation, but neural gas resulted in more distinct segments for our data.

Three visualizations are proposed to help managers understand: (1) whether any particular segmentation solution leads to distinctly different segments, (2) how each segment is characterized, and (3) what variables effectively discriminate between segments.

First, the **Segment Separation Plot** (see Figure 2, columns 2 and 3 for the artificial data sets) gives the manager a quick overview of the distinctiveness of the resulting market segments. The segment separation plot does this for all relevant dimensions of the data space, thus providing a comprehensive picture for the manager to interpret. Therefore, the plot looks very simple in the case of two-dimensional data as the artificial data sets used for illustrative purposes in Figure 2. When the data is higher-dimensional, which it typically is if consumer survey data is used as a segmentation base, all relevant two-dimensional plots have to be inspected (as illustrated later for the empirical data set in Figure 3). For the artificial data sets we provide two versions of the segment separation plot, once we include the actual data points, once we do not. The difference between the plots for the natural and the constructive segmentation are clearly visible in those illustrations: the segment separation plot indicated clearly that natural clusters exist by showing compact clusters which are far away from each other. In case of constructive clustering the segment separation plot including the data illustrates clearly that the shown split into five clusters is random because there is data in the entire space and any chosen number of clusters will lead to an even distribution of respondents to clusters. The segment separation plot without data indicates that the solution is not as clear because the clusters are not as compact, they are larger in size indicating larger spread of respondents and they are closer together indicating worse separation.

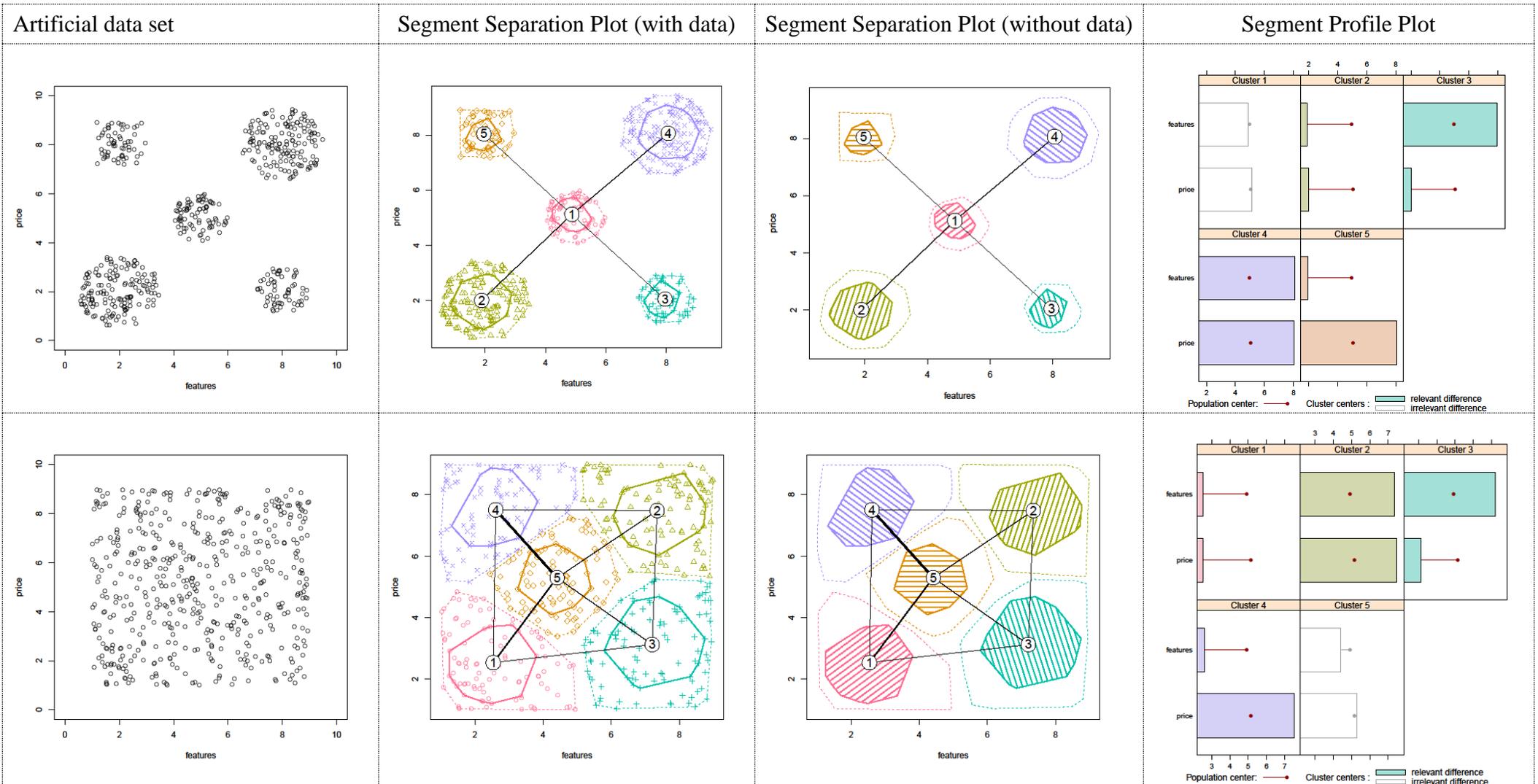
From a graphical statistics point of view, the segment separation plot depicts the convex hulls of the inner points of each cluster, projected pair-wise onto the first  $p$  principle components (here we chose  $p = 3$ ). The *inner points* of each cluster are defined as the 50 percent of all points in each cluster with minimal distance to the cluster centroid (for details see Leisch 2010). The shaded area of each cluster is comparable to the box in a box-whisker plot (Tukey 1977) which also contains the inner half of a sample. The dashed line around each cluster depicts the convex hull of 95 percent of all points in the cluster — that is, everything but a few outliers.

The decision to plot convex hulls instead of scatter plots of data points follows one of the main principles of Tufte (1983); namely, to avoid chart junk and maximize the share of data/ink. Or, as Chambers, Cleveland, Kleiner and Tukey (1993, p. 327) state: “The amount of uninformative detail and clutter in a plot should be minimized, since they interfere with our perception of the important information... Just getting all the data into a plot does not necessarily make it informative.”

A principal component projection was used because it is very popular for creating perceptual maps (for example, Hsu & Lee 2002), and the resulting plots are easy to interpret. Other possible choices are linear discriminant analysis with the cluster membership as

dependent variable, or asymmetric plots separating one cluster from the rest of the data (see Hennig 2004).

**Figure 2:** Artificial data-sets and resulting visualisations



We can further use the segment separation plot as a basis for understanding how segments can be roughly interpreted by plotting marker variables directly into the plot (as illustrated in Figure 4 for the tourist data, for the artificial data sets this plot is not required because the axes directly represent the two dimensions and the clusters can thus be directly interpreted from the segment separation plot). Such a **Segment Positioning Plot** helps managers to develop an initial understanding of the nature of segments.

Finally, the **Segment Profile Plot** (see column 4 of Figure 2 for an illustration of this plot for the two extreme cases modeled with artificial data) enables managers to develop a more detailed understanding of the nature of resulting market segments without overloading them with unnecessary information. The horizontal bars illustrate the percentage of consumers who seek certain benefits; the bullets and horizontal lines show the percentage of benefits sought by the entire market. Differences between the segment and sample percentages form the basis of segment profiling: the higher the difference, the more distinct the segment. Consequently, some variables are particularly important for the description of a segment (marker variables); whereas others are not very distinct, and therefore do not contribute much to the understanding of the nature of each segment.

To make it easy for managers to detect market variables quickly, a pre-specified rule is used to highlight only marker variables in the segment profile chart: a variable is called a marker variable if the absolute deviation from the overall mean is 25 percent of the maximum value seen, or if the relative deviation is 50 percent. Let  $M$  be the maximum value of a variable,  $mg$  the global mean of the variable, and  $mk$  the mean in cluster  $k$ , then a variable is designated as a marker variable for cluster  $k$  if one of the following two conditions is satisfied:

$$|mg - mk| \geq 0.25M \text{ or } |mg - mk| \geq 0.5m.$$

The values 0.25 for absolute deviations and 0.5 for relative deviations are configurable in the software implementation and have been chosen based on experiments with a collection of data sets. For the binary data set  $M=1$  hence, absolute differences of 0.25 and relative differences of 0.5 are marked.

The variables are sorted in a way that best represents the similarities between variables, which resulted from a preliminary hierarchical cluster analysis of the segmentation base (as shown in Figure 1 for the tourist survey data example). Thus, variables which have similar values for different tourists are clustered together, further improving the ease of interpretation for managers. The design of the Segment Profile Plot thus enables the user to instantly see the key features of each segment.

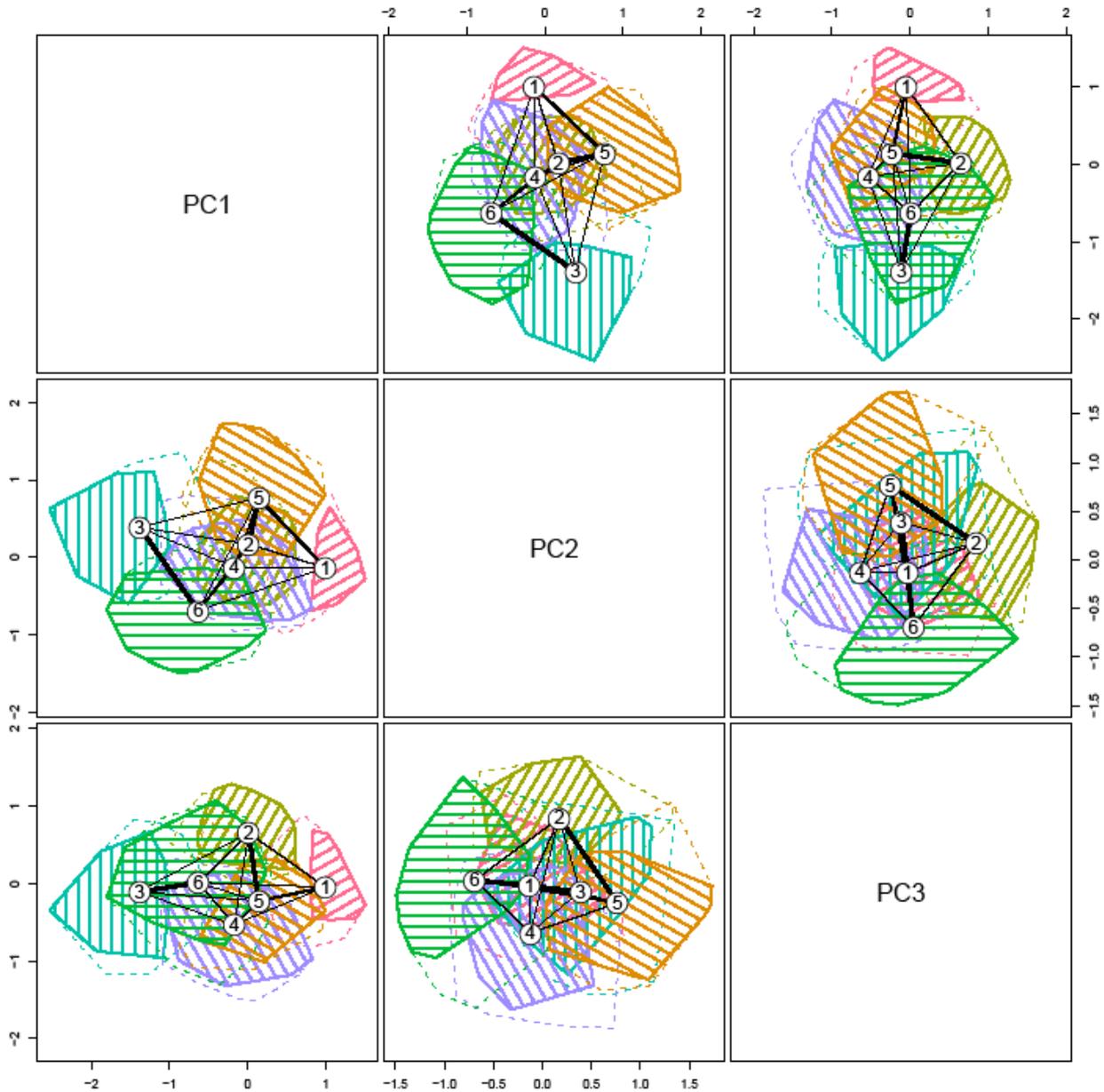
## Illustration Using Customer Survey Data

As was the case for the artificial data, the data were again clustered using the neural gas algorithm (Martinetz, Berkovich & Schulten 1993) and ten computations were run for each number of clusters between two and ten for the purpose of selecting the best solution in terms of the within-sum-of-squares measure. No solution containing a certain number of clusters was superior to the other computation results. It was therefore assumed that the data set does not contain naturally occurring segments. Consequently, segment differentiation was used as the criterion for choosing the number of clusters: the solution containing six clusters was selected.

The **Segment Separation Plot** for the tourist data is provided in Figure 3 and indicates that only Segments 1 and 3 are distinct from most perspectives. Plotting the first against the second principal components separates these two segments best. As will be seen later, this dimension effectively represents the frequency of agreement with all questions, therefore capturing response styles, not content segments. Specifically, members of Segment 1 show lower levels of agreement with respect to all questions and members of Segment 3 show above average levels of agreement with all questions.

Plotting principal components two against three illustrates the existence of four segments which are not merely response styles, yet are quite well separated. From this plot it can therefore be derived that the segmentation solution identified is a useful one: the segments are far from being located “on top of each other” which would indicate few differences between them.

**Figure 3:** Segment separation plot (tourist survey data)

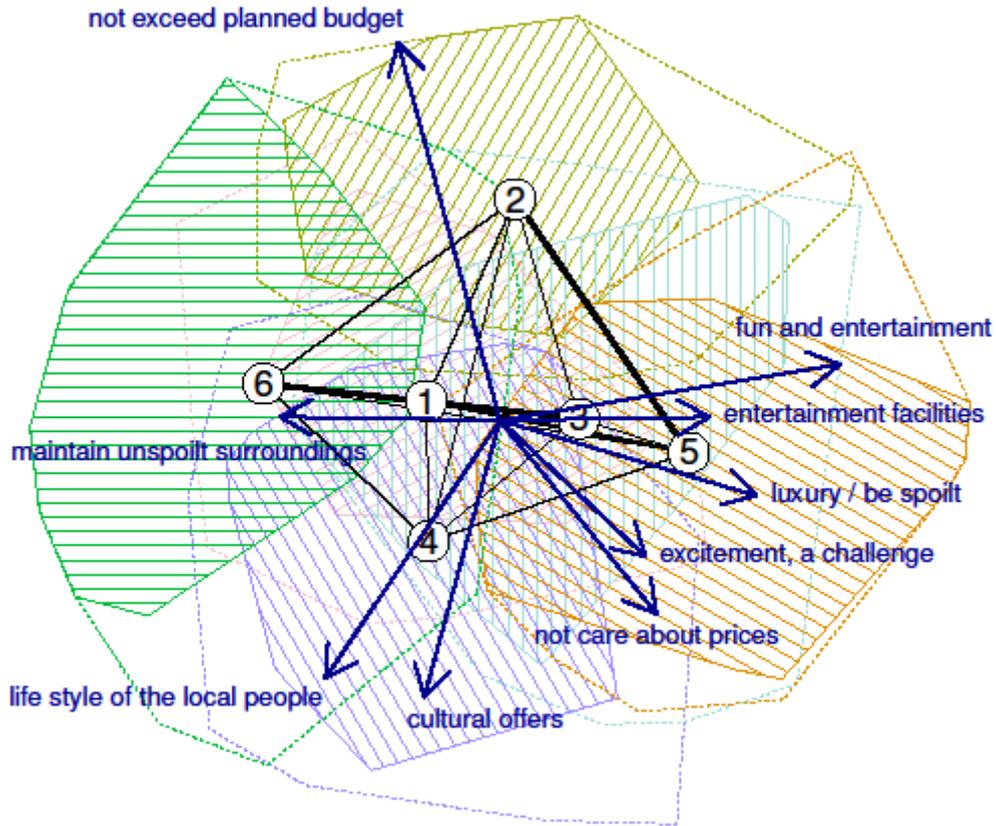


Although the level of segment separation is critical information for a user of a segmentation study, indicating how distinctly different segments of consumers are from each other, previous studies rarely use segment separation plots to illustrate outcomes of data-driven market segmentation solutions, making it impossible for a manager to assess the extent to which resulting segments represent distinct entities. Bennion (1986) represents one of the rare exceptions. However, Bennion's illustration seems to be hand-drawn; the present study proposes a fully transparent and automated algorithm to compute cluster areas. In addition,

by defining inner and outer areas of clusters, the core area of the cluster (key customers in the market segment) can be distinguished from outlying observations (customers who are not strongly associated with the segment). Becken, Simmons and Frampton (2003) use ellipses to represent segments; however, clusters in real data are not necessarily of ellipsoidal shape and so the ellipse is not considered the best format for visualizing segments.

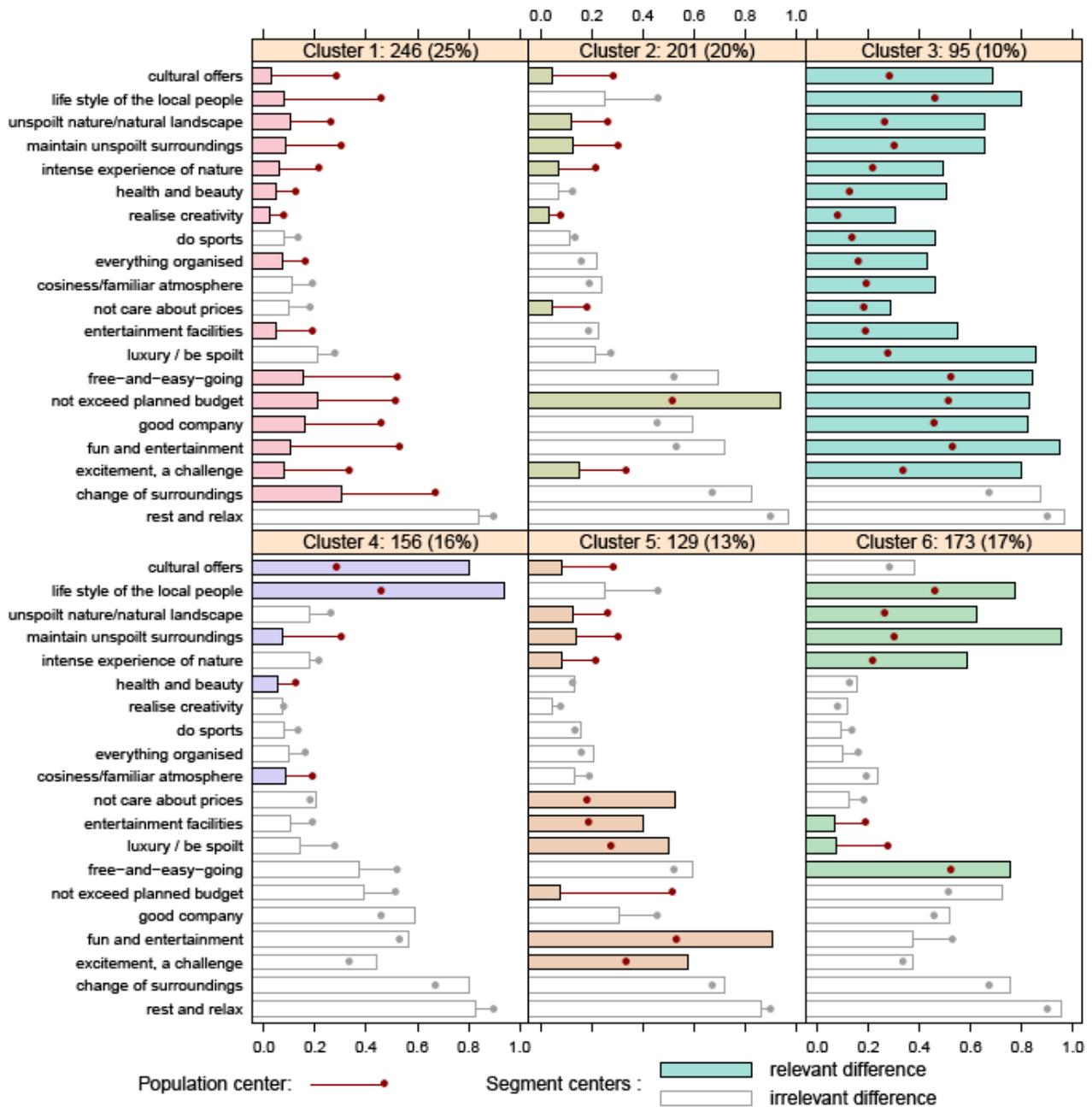
The **Segment Positioning Plot** for the tourist survey data is provided in Figure 4. As can be seen, Segment 2 wants to ensure the travel budget is not exceeded, whereas segments on the other end of the plane (4 and 5) care little about prices. Segment 6 wants to maintain unspoiled surroundings, Segment 4 is interested in cultural offerings and the lifestyle of the local people, and Segment 5 is located on the “fun and entertainment” part of the map.

**Figure 4:** Segment separation and positioning plots



Finally, the **Segment Profile Plot** is for the tourist survey data is provided in Figure 5. It can easily be seen from this plot that Segment 3 tends to agree with all statements, Segment 1 does not agree with any of the statements, Segment 2's most distinct feature is that the travel budget must not be exceeded, Segment 4 differs mainly because of the interest in culture and local people, Segment 5 seeks fun, entertainment, excitement and challenges, does not care about prices, wants to indulge in luxury, and cares about entertainment facilities at the destination. Finally, Segment 6 cares about nature playing a major part in vacations.

**Figure 5: Segment profile plot**



Previous studies typically present segment profiles in the form of tables (for example, Mazumdar & Papatla 2000; Varki & Chintagunta 2004; Ailawadi, Neslin & Gedenk 2001). Table 1 shows the segment profiles in tabular form for the vacation preference example. This tabular format is representative of how results from publicly accessible applied segmentation studies are presented and in line with Ehrenberg’s recommendation on well laid out tables (Ehrenberg 1977, 1978), listing variables in order of agreement percentage of the total sample, providing averages, thus enabling comparison across segments as well as comparison

with the sample total in columns, providing percentages rounded to full numbers and keeping columns close together to enable easier comparison. A further simplification of the table is impossible given the number of segments and the number of variables. The Table therefore does not and cannot satisfy Ehrenberg's (1977, p. 278) "Strong Criterion for a Good Table: The patterns and exceptions should be obvious at a glance", but it does pass what Ehrenberg describes as the more important "Weak Criterion: The patterns and exceptions in a table should be obvious at a glance once one has been told what they are". We argue that tables representing profiles of segments, no matter how well laid out, are substantially more difficult to interpret than the Segment Profile Plot in Figure 5 unless the number of segments and variables is untypically low for applied segmentation studies.

**Table 1: Segment profiles (in percent of respondents)**

	<b>Seg. 1</b>	<b>Seg. 2</b>	<b>Seg. 3</b>	<b>Seg. 4</b>	<b>Seg. 5</b>	<b>Seg. 6</b>	<b>Total</b>
rest and relax	84	97	97	83	86	95	90
change of surroundings	30	83	87	80	72	76	67
fun and entertainment	11	72	95	56	91	38	53
free-and-easy-going	15	70	84	37	60	75	52
not exceed planned budget	21	94	83	39	8	72	51
life style of the local people	8	25	80	94	25	77	46
good company	16	59	82	59	31	51	46
luxury / be spoilt	21	21	85	14	50	8	28
do sports	8	11	46	8	16	9	14
excitement, a challenge	8	15	80	44	57	38	33
maintain unspoiled surroundings	9	12	65	7	14	95	30
cultural offerings	3	4	68	80	9	38	28
unspoiled nature/natural landscape	10	12	65	18	12	62	26
intense experience of nature	6	7	49	18	9	58	22
entertainment facilities	4	23	55	10	40	7	19
coziness / familiar atmosphere	11	24	46	8	13	24	19
not care about prices	10	4	28	21	53	12	18
everything organized	7	22	43	10	21	10	16
health and beauty	4	7	51	5	13	16	12
realise creativity	2	3	31	7	5	12	8

Profiling segments according to their segmentation base is how most graphical visualizations were utilized in the past (for example, Reinartz & Kumar 2000; Bodapati & Gupta 2004; Andriotis & Vaughan 2003; Dolnicar & Leisch 2003; Becken et al. 2003; Dolnicar 2004; Beh & Bruyere 2007; Castro, Armario & Ruiz 2007; Horneman et al. 2002). However, while such plots give users of segmentation solutions a visual explanation of the key segment characteristics, they typically fail to guide users to the most relevant pieces of information, thus often leading to visual information overflow. The segment profile plot represents a significant improvement over previous visualizations, because it highlights marker variables while pushing into the background variables that do not contribute much to the understanding of the specifics of market segments. Interested readers can undertake a direct comparison of Figure 4 with Figure 6 in Dolnicar and Leisch (2008) to be convinced of the improved usability of the proposed segment profile plot.

## Conclusions

Data-driven market segmentation is a powerful tool for gaining insight into the market and understanding the heterogeneity of customers. Because data-driven market segmentation requires the application of statistical procedures that are not intuitively understandable by

marketing managers, it is crucial to translate the solutions of market segmentation studies into language easy to understand by users. The recommended language is, in this case, graphical. Using graphical means for understanding segmentation data, managers should have more confidence in segmentation solutions and make increased use of segmentation for marketing purposes. Naturally, there is a learning curve involved, because managers need to familiarize themselves with new graphical techniques. However, once the concepts of the graphs are understood, a quicker overview is gained of patterns in the data. It is also much less tiresome to look at graphs than to scan numeric tables for patterns in data.

This paper proposes one possible avenue for making data-driven market segmentation results more managerially useful. Using graphical statistics, we introduce three plots which illustrate the key findings in a statistically correct and managerially meaningful manner and provide reference plots which show the behavior of these plots under optimal segmentation conditions (when actual density clusters exist in the data) and under worst-case conditions (when the data set is entirely unstructured) . The segment separation plot offers managers an insight into the extent to which segments are separated; the segment positioning plot enables managers to develop an initial understanding about the nature of the segments; and the segment profile plot gives a detailed insight into the key characteristics of segments. The information contained in the segment profile plot is the most commonly provided information by data analysts to users. However, based on our review of publicly accessible reports of segmentation studies, segment profiles are typically presented in large tables which are very difficult to interpret, both in tabular formats typically used in segmentation studies as well as in tabular formats as recommended by Andrew Ehrenberg, an expert in data presentation. Even if profile plots are made available in graphical format, they presented variables in an unsorted manner and leave it up to the viewer to assess whether differences between the overall population and the segment are sufficient to be characteristic for the segment or not. The segment profile plot presented in this study overcomes both these difficulties. Information relating to segment separation is hardly ever made available to users of segmentation solutions, making it impossible for them to judge whether the segmentation solution represents the identification of naturally occurring market segments or whether, in absence of structure in the data, is represents one of many possible artificial ways in which a data set can be grouped. Providing this information to users is important as it is likely to affect the level of confidence they have in the results.

It is the authors' hope that this study will lead to increased use of graphical visualizations of segmentation solutions in order to improve the interpretability, and consequently the practical usability, of market segmentation. The full R code required to generate the proposed graphics is provided in the Appendix.

## References

- Ailawadi, K.L., Neslin, S.A., & Gedenk, K. (2001) Pursuing the value-conscious consumer: store brands versus national brand promotions. *Journal of Marketing*, 65, 71–89.
- Andriotis, K., & Vaughan, R.D. (2003) Urban residents' attitudes toward tourism development: the case of Crete. *Journal of Travel Research*, 42, 172–185.

- Becken, S., Simmons, D., & Frampton, C. (2003) Segmenting tourists by their travel pattern for insights into achieving energy efficiency. *Journal of Travel Research*, 42, 48–53.
- Beh, A., & Bruyere, B.L. (2007) Segmentation by visitor motivation in three Kenyan national reserves. *Tourism Management*, 28(6), 1464–1471.
- Bennion, M.L. Jr. (1986) Segmentation and positioning in a basic industry. *Industrial Marketing Management*, 16(9), 9–18.
- Bock, T. (2010) Improving the display of correspondence analysis using moon plots. *International Journal of Market Research*, 53(3), 307-326.
- Bodapati, A.V., & Gupta, S. (2004) The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research*, 41, 351–364.
- Bottomley, P. & Nairn, A. (2004) Blinded by science: the managerial consequences of inadequately validated cluster analysis solutions. *International Journal of Market Research*, 46(2), 171-187.
- Castro, C.B., Armario, E.M., & Ruiz, D.M. (2007) The influence of market heterogeneity on the relationship between a destination's image and tourists' future behavior. *Tourism Management*, 28, 175–187.
- Chambers, J.M., Cleveland, W.S., Kleiner, B., & Tukey, P.A. (1993) *Graphical Methods for Data Analysis*. Belmont, CA: Wadsworth International.
- Chen, C-H., Härdle, W.K., & Unwin, A. (Eds.) (2008) *Handbook of Data Visualization*. Springer Handbooks of Computational Statistics. Berlin: Springer Verlag.
- Choffrey, J-M., & Lilien, G.L. (1980) Industrial market segmentation. In *Marketing Planning for New Industrial Products*, J-M. Choffrey, G.L. Lilien (eds.). New York: J. Wiley and Sons, pp 74–91.
- Chung, K.Y., Oh, S.Y., Kim, S.S., & Han, S.Y. (2004) Three representative market segmentation methodologies for hotel guest room customers. *Tourism Management*, 25, 429-441.
- Cleveland, W.S. (1993) *Visualizing Data*. Summit, NJ: Hobart Press.
- Dibb, S., & Simkin, L. (1994) Implementation problems in industrial market segmentation. *Industrial Marketing Management*, 23, 55–63.
- Dibb, S., & Simkin, L. (1997) A program for implementing market segmentation. *Journal of Business and Industrial Marketing*, 12, 51–65.
- Dolnicar, S. (2002) Review of Data-Driven Market Segmentation in Tourism. *Journal of Travel & Tourism Marketing*, 12(1), 1-22.

- Dolnicar, S. (2004) Beyond “commonsense segmentation”: a systematics of segmentation approaches in tourism. *Journal of Travel Research*, 42, 244–250.
- Dolnicar, S., & Lazarevski, K. (2009) Methodological reasons for the theory/practice divide in market segmentation. *Journal of Marketing Management*, 25(3-4), 357-373.
- Dolnicar, S., & Leisch, F. (2003) Winter tourist segments in Austria: identifying stable vacation styles using bagged clustering techniques. *Journal of Travel Research*, 41, 281–292.
- Dolnicar, S., & Leisch, F. (2008) An investigation of tourists’ patterns of obligation to protect the environment. *Journal of Travel Research*, 46(4), 381-391.
- Dolnicar, S., & Leisch, F. (2010) Evaluation of structure and reproducibility of cluster solutions using the bootstrap. *Marketing Letters*, 21(1), 83-101.
- Doyle, P., Saunders, J., & Wong, V. (1985) Japanese marketing strategies in the UK: a comparative study. *Journal of International Business Studies*, 17, 27-46.
- Ehrenberg, A.S.C. (1977) Rudiments of numeracy. *Journal of the Royal Statistical Society, A*, 140(3), 277-297.
- Ehrenberg, A.S.C. (1978) Graphs or Tables. *The Statistician*, 27(2), 87-96.
- Fuller, J., & Matzler, K. (2008) Customer delight and market segmentation: An application of the three-factor theory of customer satisfaction on life style groups. *Tourism Management*, 29, 116-126.
- Greenberg, M., & Schwartz McDonald, S. (1989) Successful needs/benefits segmentation: a user’s guide. *Journal of Consumer Marketing*, 6(3), 29–37.
- Haley, R.I. (1968) Benefit segmentation: a decision oriented research tool. *Journal of Marketing*, 32(30), 30-35.
- Hennig, C. (2004) Asymmetric linear dimension reduction for classification. *Journal of Computational and Graphical Statistics*, 13(4), 1–17.
- Horneman, L., Carter, R.W., Wei, S., & Ruys, H. (2002) Profiling the senior traveler: an Australian perspective. *Journal of Travel Research*, 41, 23–37.
- Hsu, C.H.C., & Lee, E-J. (2002) Segmentation of senior motorcoach travelers. *Journal of Travel Research*, 40, 363–373.
- Lee, H., & Law, R. (2012) Diversity in Statistical Research Techniques: An Analysis of Refereed Research Articles in the Journal of Travel & Tourism Marketing Between 1992 and 2010. *Journal of Travel & Tourism Marketing*, 29(1), 1-17
- Leisch, F. (2006) A toolbox for k-centroids cluster analysis. *Computational Statistics and Data Analysis*, 51(2), 526–544.

- Leisch, F. (2008) Visualization of cluster analysis and finite mixture models. In *Handbook of Data Visualization*, Springer Handbooks of Computational Statistics, C-H. Chen, W. Härdle, A. Unwin (eds.). Berlin: Springer, Verlag.
- Leisch, F. (2010) Neighborhood graphs, stripes and shadow plots for cluster visualization. *Statistics and Computing*, 20(4), 457-469.
- Lo, A., Law, R., & Cheung, C. (2011) Segmenting Leisure Travelers by Risk Reduction Strategies. *Journal of Travel & Tourism Marketing*, 28(8), 828-839
- McDonald, M. (2010) Existentialism – a school of thought based on a conception of the absurdity of the universe. *International Journal of Market Research*, 52(4), 427-430.
- Martinetz, T.M., Berkovich, S.G., & Schulten, K.J. (1993) “Neural-gas” network for vector quantization and its application to time-series prediction. *IEEE Transactions on Neural Networks*, 4(4), 558–569.
- Mazanec, J.A. (1997) Segmenting city tourists into vacation styles. In *International City Tourism: Analysis and Strategy. The Cutting Edge of Tourism*. K. Grabler, G. Maier, J.A. Mazanec, K. Wöber (eds.). New York: Continuum International Publishing, pp. 114-128
- Mazanec, J.A. (2000) Market segmentation. In *Encyclopedia of Tourism*. J. Jafari (ed.). London: Routledge.
- Mazumdar, T., & Papatla, P. (2000) An investigation of reference price segments. *Journal of Marketing Research*, 37, 246–258.
- Murrell, P. (2006) *R Graphics*. London: Chapman & Hall/CRC.
- Myers, J.H., & Tauber, E.M. (1977) *Market structure analysis*. Chicago: American Marketing Association.
- Nairn, A. & Bottomley, P. (2003) Something approaching science? Cluster analysis procedures in the CRM era. *International Journal of Market Research*, 45(2), 241-261.
- Palmer, R.A., & Millier, P. (2004) Segmentation: identification, intuition, and implementation. *Industrial Marketing Management*, 33(8), 779–785.
- R Development Core Team. (2012) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Reinartz, W.J., & Kumar, V. (2000) On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. *Journal of Marketing*, 64, 17–35.
- Smith, W.R. (1956) Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21, 3–9.

- Tufte, E.R. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
- Tufte, E.R. (1997) *Visual Explanations*. Cheshire, Connecticut: Graphics Press.
- Tukey, J.W. (1977) *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Varki, S., & Chintagunta, P.K. (2004) The augmented latent class model: incorporating additional heterogeneity in the latent class model for panel data. *Journal of Marketing Research*, 41, 226–233.
- Wedel, M., & Kamakura W.A. (1998) *Market segmentation: conceptual and methodological foundations*. Boston: Kluwer Academic Publishers.
- Wilkinson, L. (2005) *The Grammar of Graphics*. New York: Springer Verlag.
- Wind, Y. (1978) Issues and advances in segmentation research. *Journal of Marketing Research*, 15(3), 317–338.

## APPENDIX

*Reviewers please note that in the final version the readers will not only be given the code, but also provided with the data set we used in this study so they can replicate our plots before using them on their own data. The data will be available from the authors' home page by the time this paper is available online, if accepted.*

R and extension package flexclust are open source and freely available at <http://cran.r-project.org>

### **R code for reproducing the clustering and PCA we used**

```
## install and load package flexclust
install.packages("flexclust")
library("flexclust")

## recreate cluster object for current version of
## flexclust (make sure we have the same
## solution and not a different local optimum of the
## cluster algorithm), and have a look at the result
load("vacpref.RData")
cl6 <- kcca(vacpref, k=vacpref6, control=list(iter=0),
           simple=FALSE, save.data=TRUE)
summary(cl6)

## hierarchical clustering of the variables
varhier <- hclust(dist(t(vacpref)), "ward")
par(mar=c(0,0,0,15))
plot(as.dendrogram(varhier), xlab="", horiz=TRUE,
     yaxt="n")

## principalcomponent projection
vacpca <- prcomp(vacpref)
```

### **R code for generating the Segment Separation Plot**

```
pairs(cl6, project=vacpca, which=1:3, asp=TRUE,
      points=FALSE, hull.args=list(density=10))
```

### **R code for generating the Segment Positioning Plot**

```
## clusters 1 and 3 get lighter colors
col <- flxColors(1:6)
col[c(1,3)] <- flxColors(1:4, "light")[c(1,3)]
par(mar=rep(0,4))
```

```
plot(c16, project=vacpca, which=2:3, col=col,  
     asp=TRUE, points=F, hull.args=list(density=10),  
     axes=FALSE)  
projAxes(vacpca, minradius=.5, which=2:3, lwd=2,  
         col="darkblue")
```

**R code for generating the Segment Profile Plot**

```
barchart(c16, shade=TRUE, which=rev(varhier$order),  
         legend=TRUE)
```