

Chapter 8

Step 6: Profiling Segments



8.1 Identifying Key Characteristics of Market Segments

The aim of the profiling step is to get to know the market segments resulting from the extraction step. Profiling is only required when data-driven market segmentation is used. For commonsense segmentation, the profiles of the segments are predefined. If, for example, age is used as the segmentation variable for the commonsense segmentation, it is obvious that the resulting segments will be age groups. Therefore, Step 6 is not necessary when commonsense segmentation is conducted.

The situation is quite different in the case of data-driven segmentation: users of the segmentation solution may have decided to extract segments on the basis of benefits sought by consumers. Yet – until after the data has been analysed – the defining characteristics of the resulting market segments are unknown. Identifying these defining characteristics of market segments with respect to the segmentation variables is the aim of profiling. Profiling consists of characterising the market segments individually, but also in comparison to the other market segments. If winter tourists in Austria are asked about their vacation activities, most state they are going alpine skiing. Alpine skiing may characterise a segment, but alpine skiing may not differentiate a segment from other market segments.

At the profiling stage, we inspect a number of alternative market segmentation solutions. This is particularly important if no natural segments exist in the data, and either a reproducible or a constructive market segmentation approach has to be taken. Good profiling is the basis for correct interpretation of the resulting segments. Correct interpretation, in turn, is critical to making good strategic marketing decisions.

Data-driven market segmentation solutions are not easy to interpret. Managers have difficulties interpreting segmentation results correctly (Nairn and Bottomley 2003; Bottomley and Nairn 2004); 65% of 176 marketing managers surveyed in a study by Dolnicar and Lazarevski (2009) on the topic of market segmentation state that they have difficulties understanding data-driven market segmentation solutions,

and 71% feel that segmentation analysis is like a black box. A few of the quotes provided by these marketing managers when asked how market segmentation results are usually presented to them are insightful:

- ... as a long report that usually contradicts the results
- ... rarely with a clear Executive Summary
- ... in a rushed slap hazard fashion with the attitude that ‘leave the details to us’
- ...
- The result is usually arranged 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
- In a PowerPoint presentation with a slick handout

(quotes from the study reported in Dolnicar and Lazarevski 2009).

In the following sections we discuss traditional and graphical statistics approaches to segment profiling. Graphical statistics approaches make profiling less tedious, and thus less prone to misinterpretation.

8.2 Traditional Approaches to Profiling Market Segments

We use the Australian vacation motives data set. Segments were extracted from this data set in Sect. 7.5.4 using the neural gas clustering algorithm with number of segments varied from 3 to 8 and with 20 random restarts. We reload the segmentation solution derived and saved on page 171:

```
R> library("flexclust")
R> data("vacmot", package = "flexclust")
R> load("vacmot-clusters.RData")
```

Data-driven segmentation solutions are usually presented to users (clients, managers) in one of two ways: (1) as high level summaries simplifying segment characteristics to a point where they are misleadingly trivial, or (2) as large tables that provide, for each segment, exact percentages for each segmentation variable. Such tables are hard to interpret, and it is virtually impossible to get a quick overview of the key insights. This is illustrated by Table 8.1. Table 8.1 shows the mean values of the segmentation variables by segment (extracted from the return object using `parameters(vacmot.k6)`), together with the overall mean values. Because the travel motives are binary, the segment means are equal to the percentage of segment members engaging in each activity.

Table 8.1 provides the exact percentage of members of each segment that indicate that each of the travel motives matters to them. To identify the defining characteristics of the market segments, the percentage value of each segment for each segmentation variable needs to be compared with the values of other segments or the total value provided in the far right column.

Table 8.1 Six segments computed with the neural gas algorithm for the Australian travel motives data set. All numbers are percentages of people in the segment or in the total sample agreeing to the motives

	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6	Total
Rest and relax	83	96	89	82	98	96	90
Change of surroundings	27	82	73	82	87	77	67
Fun and entertainment	7	71	81	60	95	37	53
Free-and-easy-going	12	65	58	45	87	75	52
Not exceed planned budget	23	100	2	49	84	73	51
Life style of the local people	9	29	30	90	75	80	46
Good company	14	59	40	58	77	55	46
Excitement, a challenge	9	17	39	57	76	36	33
Maintain unspoilt surroundings	9	10	16	7	67	95	30
Cultural offers	4	2	5	96	62	38	28
Luxury / be spoilt	19	24	39	13	89	6	28
Unspoilt nature/natural landscape	10	10	13	15	69	64	26
Intense experience of nature	6	8	9	21	50	58	22
Cosiness/familiar atmosphere	11	24	12	7	49	25	19
Entertainment facilities	5	25	30	14	53	6	19
Not care about prices	8	7	43	19	29	10	18
Everything organised	7	21	15	12	46	9	16
Do sports	8	12	13	10	46	7	14
Health and beauty	5	8	10	8	49	16	12
Realise creativity	2	2	3	8	29	14	8

Using Table 8.1 as the basis of interpreting segments shows that the defining characteristics of segment 2, for example, are: being motivated by rest and relaxation, and not wanting to exceed the planned travel budget. Also, many members of segment 2 care about a change of surroundings, but not about cultural offers, an intense experience of nature, about not caring about prices, health and beauty and realising creativity. Segment 1 is likely to be a response style segment because – for each travel motive – the percentage of segment members indicating that a travel motive is relevant to them is low (compared to the overall percentage of agreement).

Profiling all six market segments based on Table 8.1 requires comparing 120 numbers if each segment's value is only compared to the total (for each one of 20 travel motives, the percentages for six segments have to be compared to the percentage in the total column). If, in addition, each segment's value is compared to the values of other segments, $(6 \times 5)/2 = 15$ pairs of numbers have to be compared for each row of the table. For the complete table with 20 rows, a staggering $15 \times 20 = 300$ pairs of numbers would have to be compared between segments. In total this means 420 comparisons including those between segments only and between segments and the total.

Imagine that the segmentation solution in Table 8.1 is not the only one. Rather, the data analyst presents five alternative segmentation solutions containing six

segments each. A user in that situation would have to compare $5 \times 420 = 2100$ pairs of numbers to be able to understand the defining characteristics of the segments. This is an outrageously tedious task to perform, even for the most astute user.

Sometimes – to deal with the size of this task – information is provided about the statistical significance of the difference between segments for each of the segmentation variables. This approach, however, is not statistically correct. Segment membership is directly derived from the segmentation variables, and segments are created in a way that makes them maximally different, thus not allowing to use standard statistical tests to assess the significance of differences.

8.3 Segment Profiling with Visualisations

Neither the highly simplified, nor the very complex tabular representation typically used to present market segmentation solutions make much use of graphics, although data visualisation using graphics is an integral part of statistical data analysis (Tufté 1983, 1997; Cleveland 1993; Chen et al. 2008; Wilkinson 2005; Kastellec and Leoni 2007). Graphics are particularly important in exploratory statistical analysis (like cluster analysis) because they provide insights into the complex relationships between variables. In addition, in times of big and increasingly bigger data, visualisation offers a simple way of monitoring developments over time. Both McDonald and Dunbar (2012) and Lilien and Rangaswamy (2003) recommend the use of visualisation techniques to make the results of a market segmentation analysis easier to interpret. Haley (1985, p. 227), long before the wide adoption of graphical statistics, pointed out that the same information presented in tabular form is not nearly so insightful. More recently, Cornelius et al. (2010, p. 170) noted, in a review of graphical approaches suitable for interpreting results of market structure analysis, that a single two-dimensional graphical format is preferable to more complex representations that lack intuitive interpretations.

A review of visualisation techniques available for cluster analysis and mixture models is provided by Leisch (2008). Examples of prior use of visualisations of segmentation solutions are given in Reinartz and Kumar (2000), Horneman et al. (2002), Andriotis and Vaughan (2003), Becken et al. (2003), Dolnicar and Leisch (2003, 2014), Bodapati and Gupta (2004), Dolnicar (2004), Beh and Bruyere (2007), and Castro et al. (2007).

Visualisations are useful in the data-driven market segmentation process to inspect, for each segmentation solution, one or more segments in detail. Statistical graphs facilitate the interpretation of segment profiles. They also make it easier to assess the usefulness of a market segmentation solution. The process of segmenting data always leads to a large number of alternative solutions. Selecting one of the possible solutions is a critical decision. Visualisations of solutions assist the data analyst and user with this task.

8.3.1 Identifying Defining Characteristics of Market Segments

A good way to understand the defining characteristics of each segment is to produce a *segment profile plot*. The segment profile plot shows – for all segmentation variables – how each market segment differs from the overall sample. The segment profile plot is the direct visual translation of tables such as Table 8.1.

In figures and tables, segmentation variables do not have to be displayed in the order of appearance in the data set. If variables have a meaningful order in the data set, the order should be retained. If, however, the order of variables is independent of content, it is useful to rearrange variables to improve visualisations.

Table 8.1 sorts the 20 travel motives by the total mean (last column). Another option is to order segmentation variables by similarity of answer patterns. We can achieve this by clustering the columns of the data matrix:

```
R> vacmot.vdist <- dist(t(vacmot))
R> vacmot.vclust <- hclust(vacmot.vdist, "ward.D2")
```

The `t()` around the data matrix `vacmot` transposes the matrix such that distances between columns rather than rows are computed. Next, hierarchical clustering of the variables is conducted using Ward's method. Figure 8.1 shows the result.

Tourists who are motivated by cultural offers are also interested in the lifestyle of local people. Tourists who care about an unspoilt natural landscape also show interest in maintaining unspoilt surroundings, and seek an intense experience of nature. A segment profile plot like the one in Fig. 8.2 results from:

```
R> barchart(vacmot.k6, shade = TRUE,
+   which = rev(vacmot.vclust$order))
```

Argument `which` specifies the variables to be included, and their order of presentation. Here, all variables are shown in the order suggested by hierarchical clustering of variables. `shade = TRUE` identifies so-called *marker variables* and depicts them in colour. These variables are particularly characteristic for a segment. All other variables are greyed out.

The segment profile plot is a so-called *panel plot*. Each of the six panels represents one segment. For each segment, the segment profile plot shows the cluster centres (centroids, representatives of the segments). These are the numbers contained in Table 8.1. The dots in Fig. 8.2 are identical in each of the six panels, and represent the total mean values for the segmentation variables across all observations in the data set. The dots are the numbers in the last column in Table 8.1. These dots serve as reference points for the comparison of values for each segment with values averaged across all people in the data set.

To make the chart even easier to interpret, marker variables appear in colour (solid bars). The remaining segmentation variables are greyed out. The definition of marker variables in the segment profile plot used by default in `barchart()` is suitable for binary variables, and takes into account the absolute and relative difference of the segment mean to the total mean. Marker variables are defined as variables which deviate by more than 0.25 from the overall mean. For example, a

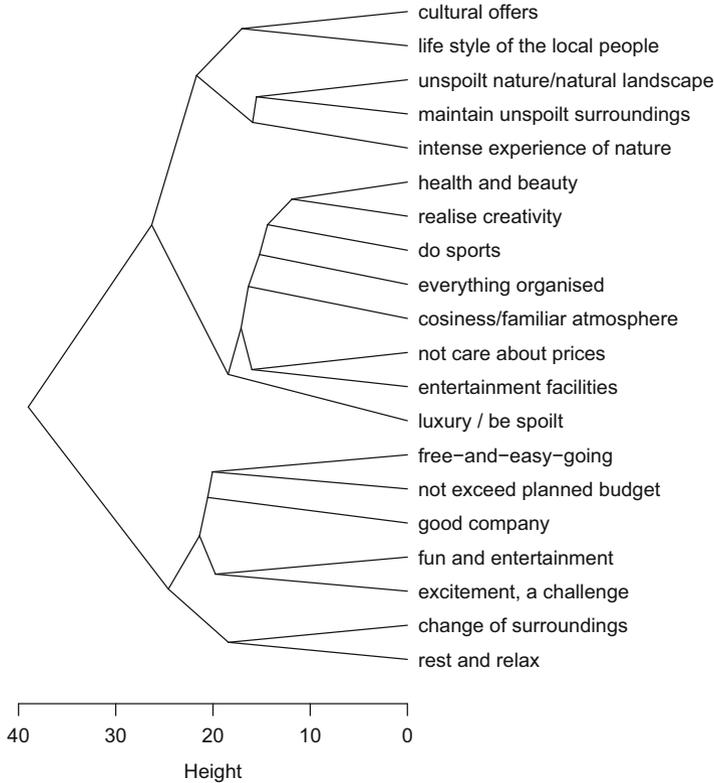


Fig. 8.1 Hierarchical clustering of the segmentation variables of the Australian travel motives data set using Ward's method

variable with a total sample mean of 0.20, and a segment mean of 0.60 qualifies as marker variable ($0.20 + 0.25 = 0.45 < 0.60$). Such a large absolute difference is hard to obtain for segmentation variables with very low sample means. A relative difference of 50% from the total mean, therefore, also makes the variable a marker variable.

The deviation figures of 0.25 and 50% have been empirically determined to indicate substantial differences on the basis of inspecting many empirical data sets, but are ultimately arbitrary and, as such, can be chosen by the data analyst and user as they see fit. In particular if the segmentation variables are not binary, different thresholds for defining a marker variable need to be specified.

Looking at the travel motive of HEALTH AND BEAUTY in Fig. 8.2 makes it obvious that this is not a mainstream travel motive for tourists. This segmentation variable has a sample mean of 0.12; this means that only 12% of all the people who participated in the survey indicated that HEALTH AND BEAUTY was a travel motive for them. For segments with HEALTH AND BEAUTY outside of the interval 0.12 ± 0.06 this vacation activity will be considered a marker variable, because 0.06 is 50% of 0.12.

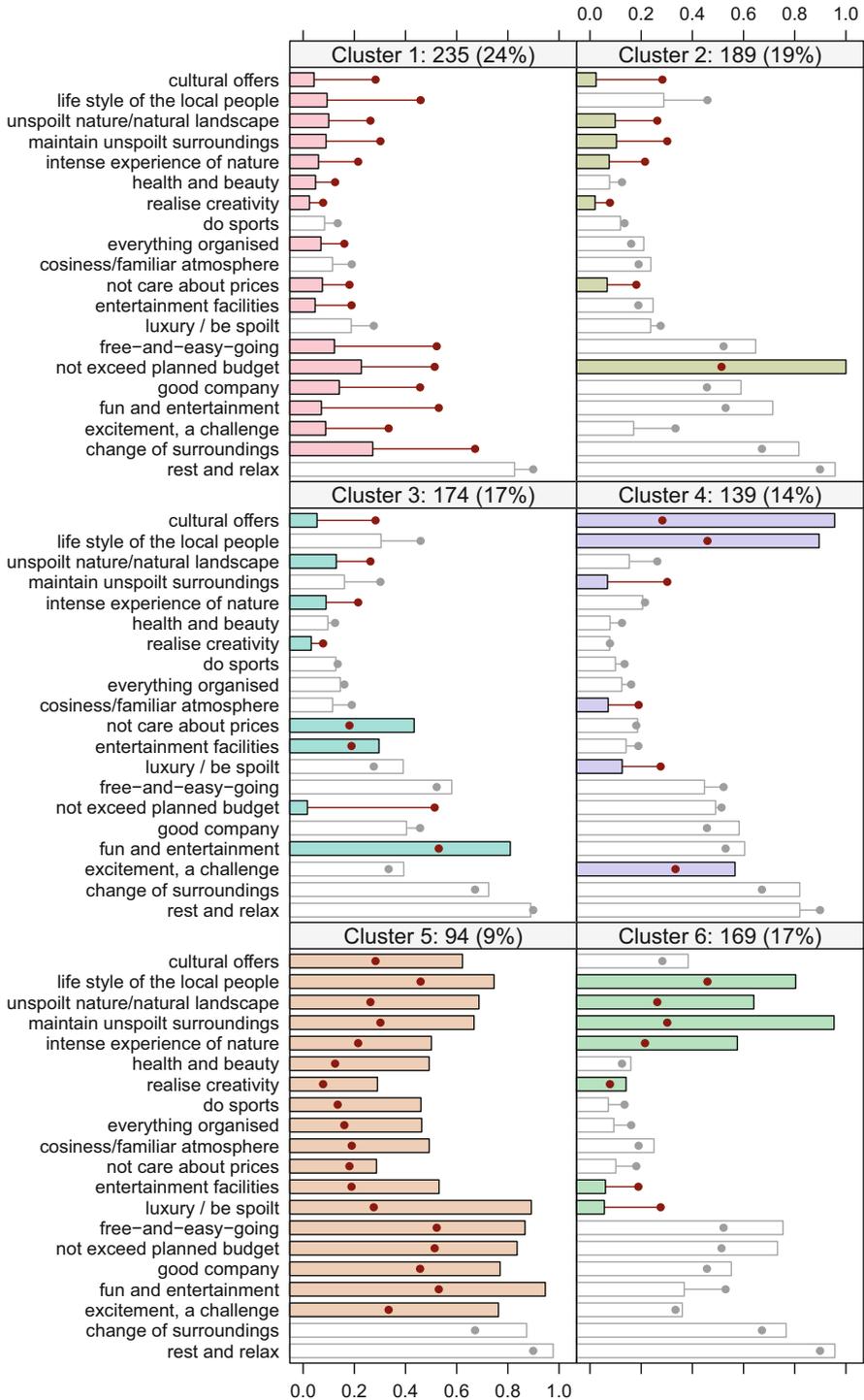


Fig. 8.2 Segment profile plot for the six-segment solution of the Australian travel motives data set

The segment profile plot in Fig. 8.2 contains the same information as Table 8.1: the percentage of segment members indicating that each of the travel motives matters to them. Marker variables are highlighted in colour. As can be seen, a segmentation solution presented using a segment profile plot (such as the one shown in Fig. 8.2) is much easier and faster to interpret than when it is presented as a table, no matter how well the table is structured. We see that members of segment 2 are characterised primarily by not wanting to exceed their travel budget. Members of segment 4 are interested in culture and local people; members of segment 3 want fun and entertainment, entertainment facilities, and do not care about prices. Members of segment 6 see nature as critical to their vacations. Finally, segments 1 and 5 have to be interpreted with care as they are likely to represent response style segments.

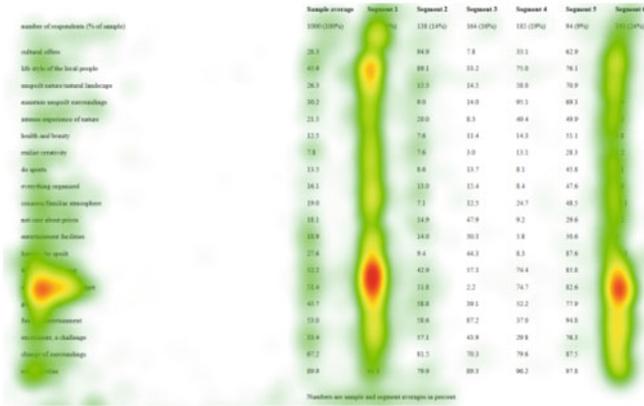
An eye tracking study conducted by Nazila Babakhani as part of her PhD studies investigated differences in people's ability to interpret complex data analysis results from market segmentation studies presented in traditional tabular versus graphical statistics format. Participants saw one of three types of presentations of segmentation results: a table; an improved table with key information bolded; and a segment profile plot. Processing time of information was the key variable of interest. Eye tracking plots indicate how long a person looked at something.

A heat map showing how long one person was looking at each section of the table or figure is shown in Fig. 8.3. We see that this person worked harder to extract information from the tables; the heat maps of the tables contain more yellow and red colouring, representing longer looking times. Longer looking times indicate more cognitive effort being invested in the interpretation of the tables. Also, the person looked at a higher proportion of the table; they were processing a larger area in the attempt to answer the question. In contrast, the heat map of the segment profile plot in Fig. 8.3 shows that the person did not need to look as long to find the answer. They also inspected a smaller surface area. The heat map suggests that it took less effort to find the information required to answer the question. It is therefore well worth spending some extra time on presenting results of a market segmentation analysis as a well designed graph. Good visualisations facilitate interpretation by managers who make long-term strategic decisions based on segmentation results. Such long-term strategic decisions imply substantial financial commitments to the implementation of a segmentation strategy. Good visualisations, therefore, offer an excellent return on investment.

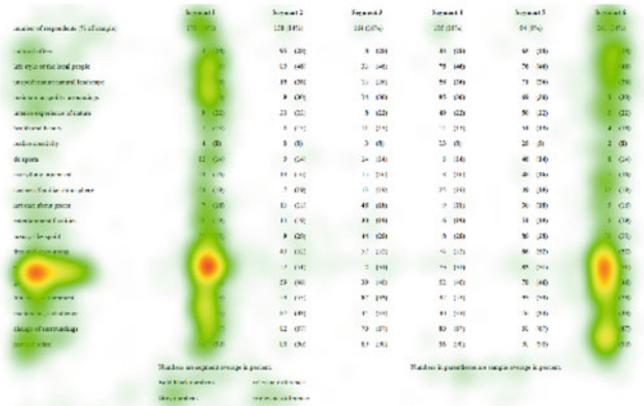
8.3.2 *Assessing Segment Separation*

Segment separation can be visualised in a *segment separation plot*. The segment separation plot depicts – for all relevant dimensions of the data space – the overlap of segments.

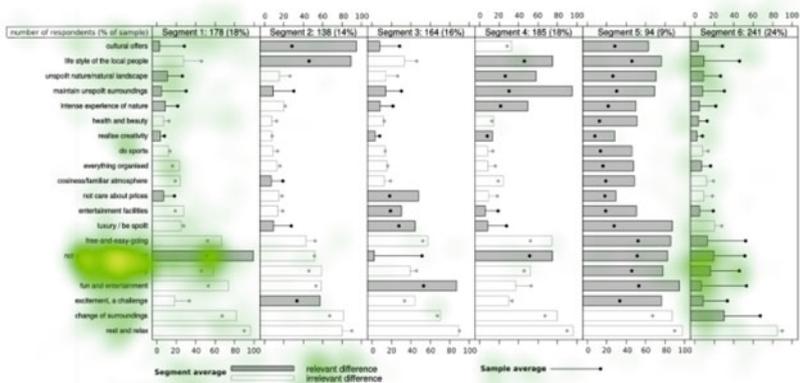
Segment separation plots are very simple if the number of segmentation variables is low, but become complex as the number of segmentation variables increases. But even in such complex situations, segment separation plots offer data analysts and users a quick overview of the data situation, and the segmentation solution.



(a)



(b)



(c)

Fig. 8.3 One person’s eye tracking heat maps for three alternative ways of presenting segmentation results. (a) Traditional table. (b) Improved table. (c) Segment profile plot

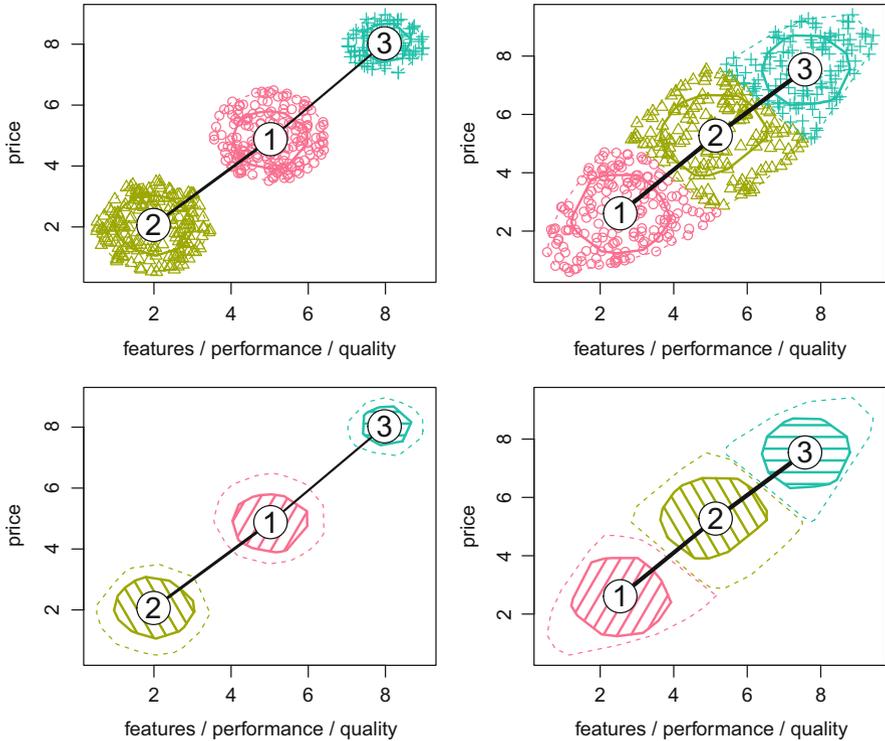


Fig. 8.4 Segment separation plot including observations (first row) and not including observations (second row) for two artificial data sets: three natural, well-separated clusters (left column); one elliptic cluster (right column)

Examples of segment separation plots are provided in Fig. 8.4 for two different data sets (left compared to right column). These plots are based on two of the artificial data sets used in Table 2.3: the data set that contains three distinct, well-separated segments, and the data set with an elliptic data structure. The segment separation plot consists of (1) a scatter plot of the (projected) observations coloured by segment membership and the (projected) cluster hulls, and (2) a neighbourhood graph.

The artificial data visualised in Fig. 8.4 are two-dimensional. So no projection is required. The original data is plotted in a scatter plot in the top row of Fig. 8.4. The colour of the observations indicates true segment membership. The different cluster hulls indicate the shape and spread of the true segments. Dashed cluster hulls contain (approximately) all observations. Solid cluster hulls contain (approximately) half of the observations. The bottom row of Fig. 8.4 omits the data, and displays cluster hulls only.

Neighbourhood graphs (black lines with numbered nodes) indicate similarity between segments (Leisch 2010). The segment solutions in Fig. 8.4 contain three

segments. Each plot, therefore, contains three numbered nodes plotted at the position of the segment centres. The black lines connect segment centres, and indicate similarity between segments. A black line is only drawn between two segment centres if they are the two closest segment centres for at least one observation (consumer). The width of the black line is thicker if more observations have these two segment centres as their two closest segment centres.

As can be seen in Fig. 8.4, the neighbourhood graphs for the two data sets are quite similar. We need to add either the observations or the cluster hulls to assess the separation between segments.

For the two data sets used in Fig. 8.4, the two dimensions representing the segmentation variables can be directly plotted. This is not possible if 20-dimensional travel motives data serve as segmentation variables. In such a situation, the 20-dimensional space needs to be projected onto a small number of dimensions to create a segment separation plot. We can use a number of different projection techniques, including some which maximise separation (Hennig 2004), and principal components analysis (see Sect. 6.5). We calculate principal components analysis for the Australian travel motives data set with the following command:

```
R> vacmot.pca <- prcomp(vacmot)
```

This provides the rotation applied to the original data when creating our segment separation plot. We use the segmentation solution obtained from neural gas on page 171, and create a segment separation plot for this solution:

```
R> plot(vacmot.k6, project = vacmot.pca, which = 2:3,
+       xlab = "principal component 2",
+       ylab = "principal component 3")
R> projAxes(vacmot.pca, which = 2:3)
```

Figure 8.5 contains the resulting plot. Argument `project` uses the principal components analysis projection. Argument `which` selects principal components 2 and 3, and `xlab` and `ylab` assign labels to axes. Function `projAxes()` enhances the segment separation plot by adding directions of the projected segmentation variables. The enhanced version combines the advantages of the segment separation plot with the advantages of perceptual maps.

Due to the overlap of market segments (and the sample size of $n = 1000$), the plot in Fig. 8.5 is messy and hard to read. Modifying colours (argument `col`), omitting observations (`points = FALSE`), and highlighting only the inner area of each segment (`hull.args = list(density = 10)`, where `density` specifies how many lines shade the area) leads to a cleaner version (Fig. 8.6):

```
R> plot(vacmot.k6, project = vacmot.pca, which = 2:3,
+       col = flxColors(1:6, "light"),
+       points = FALSE, hull.args = list(density = 10),
+       xlab = "principal component 2",
+       ylab = "principal component 3")
R> projAxes(vacmot.pca, which = 2:3, col = "darkblue",
+       cex = 1.2)
```

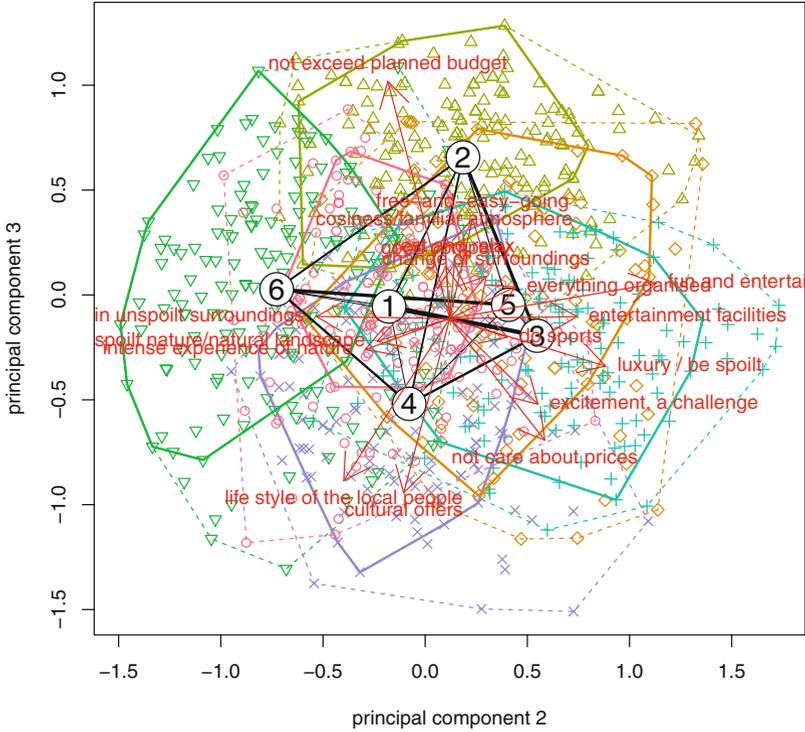


Fig. 8.5 Segment separation plot using principal components 2 and 3 for the Australian travel motives data set

The plot is still not trivial to assess, but it is easier to interpret than the segment separation plot shown in Fig. 8.5 containing additional information. Figure 8.6 is hard to interpret, because natural market segments are not present. This difficulty in interpretation is due to the data, not the visualisation. And the data used for this plot is very representative of consumer data.

Figure 8.6 shows the existence of a market segment (segment 6, green shaded area) that cares about maintaining unspoilt surroundings, unspoilt nature, and wants to intensely experience nature when on vacations. Exactly opposite is segment 3 (cyan shaded area) wanting luxury, wanting to be spoilt, caring about fun, entertainment and the availability of entertainment facilities, and not caring about prices. Another segment on top of the plot in Fig. 8.6 (segment 2, olive shaded area)

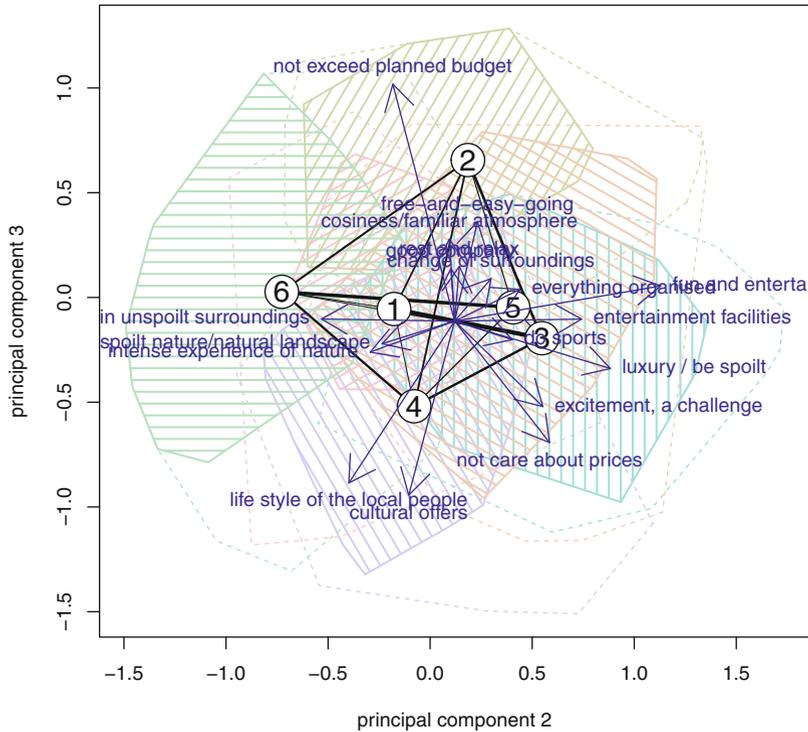


Fig. 8.6 Segment separation plot using principal components 2 and 3 for the Australian travel motives data set without observations

is characterised by one single feature only: members of this market segment do not wish to exceed their planned travel budget. Opposite to this segment, at the bottom of the plot is segment 4 (blue shaded area), members of which care about the life style of local people and cultural offers.

Each segment separation plot only visualises one possible projection. So, for example, the fact that segments 1 and 5 in this particular projection overlap with other segments does not mean that these segments overlap in all projections. However, the fact that segments 6 and 3 are well-separated in this projection does allow the conclusion – based on this single projection only – that they represent distinctly different tourists in terms of the travel motives.

8.4 Step 6 Checklist

Task	Who is responsible?	Completed?
Use the selected segments from Step 5.		<input type="checkbox"/>
Visualise segment profiles to learn about what makes each segment distinct.		<input type="checkbox"/>
Use knock-out criteria to check if any of the segments currently under consideration should already be eliminated because they do not comply with the knock-out criteria.		<input type="checkbox"/>
Pass on the remaining segments to Step 7 for describing.		<input type="checkbox"/>

References

- Andriotis K, Vaughan RD (2003) Urban residents' attitudes toward tourism development: the case of Crete. *J Travel Res* 42(2):172–185
- Becken S, Simmons D, Frampton C (2003) Segmenting tourists by their travel pattern for insights into achieving energy efficiency. *J Travel Res* 42(1):48–53
- Beh A, Bruyere BL (2007) Segmentation by visitor motivation in three Kenyan national reserves. *Tour Manag* 28(6):1464–1471
- Bodapati AV, Gupta S (2004) The recoverability of segmentation structure from store-level aggregate data. *J Mark Res* 41(3):351–364
- Bottomley P, Nairn A (2004) Blinded by science: the managerial consequences of inadequately validated cluster analysis solutions. *Int J Mark Res* 46(2):171–187
- Castro CB, Armario EM, Ruiz DM (2007) The influence of market heterogeneity on the relationship between a destination's image and tourists' future behavior. *Tour Manag* 28(1):175–187
- Chen CH, Härdle WK, Unwin A (2008) Handbook of data visualization. Springer handbooks of computational statistics. Springer, Heidelberg
- Cleveland W (1993) Visualizing data. Hobart Press, Summit
- Cornelius B, Wagner U, Natter M (2010) Managerial applicability of graphical formats to support positioning decisions. *Journal für Betriebswirtschaft* 60(3):167–201
- Dolnicar S (2004) Beyond commonsense segmentation: a systematics of segmentation approaches in tourism. *J Travel Res* 42(3):244–250
- Dolnicar S, Leisch F (2003) Winter tourist segments in Austria: identifying stable vacation styles for target marketing action. *J Travel Res* 41(3):281–193
- Dolnicar S, Leisch F (2014) Using graphical statistics to better understand market segmentation solutions. *Int J Mark Res* 56(2):97–120.
- Dolnicar S, Lazarevski K (2009) Methodological reasons for the theory/practice divide in market segmentation. *J Mark Manag* 25(3–4):357–373
- Haley RI (1985) Developing effective communications strategy – a benefit segmentation approach. Wiley, New York
- Hennig C (2004) Asymmetric linear dimension reduction for classification. *J Comput Graph Stat* 13(4):930–945
- Horneman L, Carter R, Wei S, Ruys H (2002) Profiling the senior traveler: an Australian perspective. *J Travel Res* 41(1):23–37

- Kastellec JP, Leoni EL (2007) Using graphs instead of tables in political science. *Perspect Polit* 5(4):755–771
- Leisch F (2008) Visualization of cluster analysis and finite mixture models. In: Chen CH, Härdle W, Unwin A (eds) *Handbook of data visualization*. Springer handbooks of computational statistics. Springer, Berlin
- Leisch F (2010) Neighborhood graphs, stripes and shadow plots for cluster visualization. *Stat Comput* 20(4):457–469
- Lilien GL, Rangaswamy A (2003) *Marketing engineering: computer-assisted marketing analysis and planning*, 2nd edn. Prentice Hall, Upper Saddle River
- McDonald M, Dunbar I (2012) *Market segmentation: how to do it and how to profit from it*, 4th edn. Wiley, Chichester
- Nairn A, Bottomley P (2003) Something approaching science? Cluster analysis procedures in the CRM era. *Int J Mark Res* 45(2):241–261
- Reinartz W, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: an empirical investigation and implications for marketing. *J Mark* 64(4):17–35
- Tufte ER (1983) *The visual display of quantitative information*. Graphics Press, Cheshire
- Tufte ER (1997) *Visual explanations*. Graphics Press, Cheshire
- Wilkinson L (2005) *The grammar of graphics*. Springer, New York

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

