# Step 6: Profiling Segments

## Identifying Key Characteristics of Market Segments

1**. Profiling in Data-Driven Segmentation**: Profiling is crucial in data-driven segmentation because the defining characteristics of resulting market segments are unknown until after the data analysis. It involves characterizing the segments individually and in comparison, to each other. For example, while many winter tourists in Austria may engage in alpine skiing, this activity alone may not differentiate one segment from another.

2.**Inspection of Alternative Solutions**: Profiling involves examining multiple segmentation solutions, especially if no natural segments exist in the data. This is important when a reproducible or constructive segmentation approach is required. Good profiling is essential for correctly interpreting segmentation results, which in turn informs strategic marketing decisions.

3**. Challenges in Interpreting Data-Driven Segmentation**: Data-driven segmentation solutions are often challenging for managers to interpret correctly. A survey cited in the text indicates that a significant portion of marketing managers struggle to understand these solutions, with many finding them to be presented in a confusing or unhelpful manner.

**4. Presentation of Segmentation Results:** The passage provides quotes from marketing managers describing how segmentation results are typically presented to them. These quotes highlight common issues such as contradictory results, lack of clear summaries, and the presentation of information in a rushed or confusing manner.

**5. Approaches to Profiling**: Traditional and graphical statistics approaches to segment profiling are discussed. Graphical statistics approaches are noted for making profiling less tedious and prone to misinterpretation compared to traditional methods.

## Traditional Approaches to Profiling Market Segments

Interpreting such tables involves comparing the percentages across segments and with the total sample to identify defining characteristics of each segment. However, this process can be extremely tedious, especially when considering multiple segmentation solutions or comparing segments across variables. For example, interpreting just one table with six segments for 20 travel motives would require 420 comparisons. If there are multiple segmentation solutions to consider, the number of comparisons becomes overwhelmingly large.

It highlights that providing information about the statistical significance of differences between segments for each variable is not statistically correct. This is because segment membership is derived directly from the segmentation variables, and segments are created to be maximally different, making standard statistical tests inappropriate for assessing differences.

## Segment Profiling with Visualisations

**Importance of Graphics in Data Analysis**: Graphics are integral to statistical data analysis, offering insights into complex relationships between variables. They are particularly valuable in exploratory statistical analysis, such as cluster analysis, and provide a simple way to monitor developments over time, especially in the era of big data.

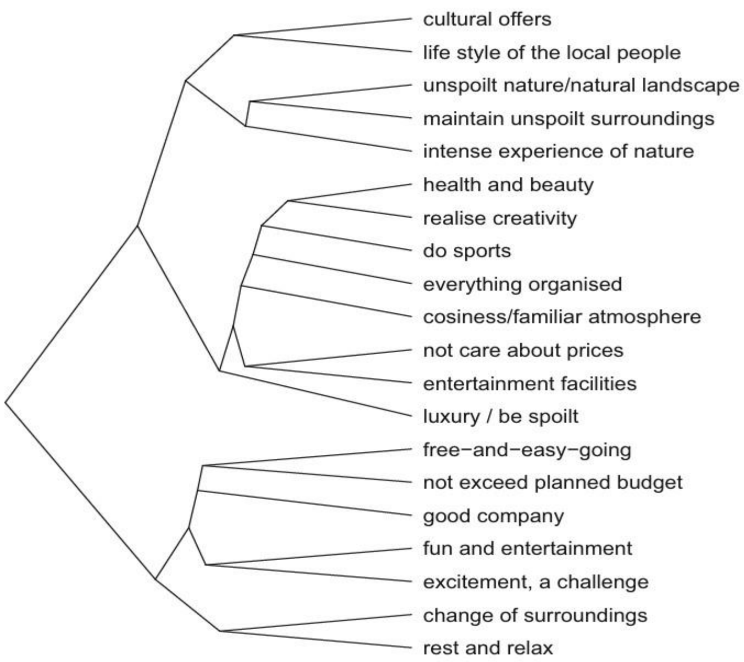
**Recommendations from Experts:** Experts in market segmentation analysis recommend the use of visualization techniques to make results easier to interpret. Visualizations offer insights that are not as readily apparent in tabular formats.

**Historical Perspective:** Even before the widespread adoption of graphical statistics, experts noted the limitations of tabular presentations compared to graphical representations for conveying insights.

**Examples of Visualization Techniques:** The passage provides examples of prior use of visualizations in segmentation analysis and recommends techniques for producing segment profile plots, which visually depict how each market segment differs from the overall sample across segmentation variables.

**Advantages of Segment Profile Plots:** Segment profile plots offer a direct visual translation of tabular data, making it easier to identify defining characteristics of each segment. Marker variables, which deviate substantially from the overall mean, are highlighted to draw attention to key differences between segments.

**Research Findings:** Eye tracking studies indicate that people expend more cognitive effort and time when interpreting tabular data compared to graphical representations. Well-designed graphical visualizations facilitate interpretation and lead to better decision-making by managers.




## Assessing Segment Separation

**Description of Segment Separation Plots:** Segment separation plots depict the overlap of segments in the data space across relevant dimensions. They typically consist of scatterplots of observations coloured by segment membership, along with cluster hulls and a neighbourhood graph indicating similarity between segments.

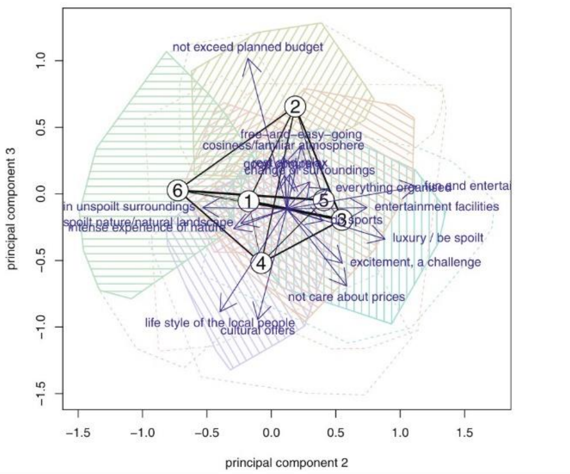
**Illustrative Examples:** The passage provides examples of segment separation plots for two different datasets, one with three distinct, well-separated segments and another with an elliptic data structure. These plots help visualize the shape and spread of segments and assess the separation between them.

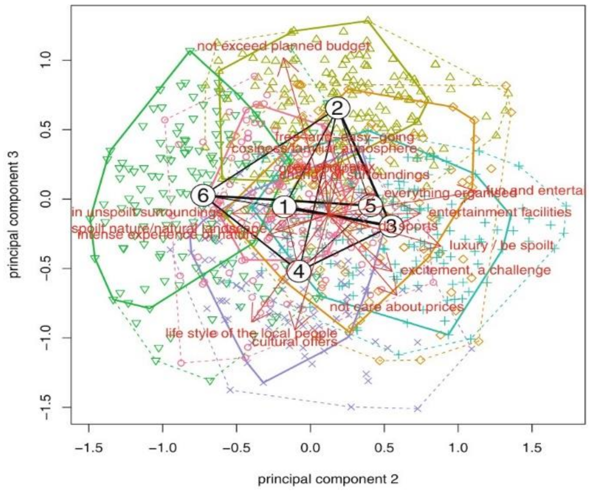
**Projection Techniques**: In cases where the segmentation variables have high dimensionality, such as in the case of 20-dimensional travel motives data, projection techniques are used to reduce the dimensionality of the data for visualization purposes. Techniques like principal components analysis (PCA) can be employed to project the data onto a smaller number of dimensions.

**Enhanced Visualization:** The passage discusses enhancements to segment separation plots, such as adding directions of the projected segmentation variables, to combine the advantages of segment separation plots with perceptual maps. These enhancements make the plots easier to interpret while retaining their effectiveness in conveying segmentation insights.

**Interpretation of Results:** The passage illustrates how segment separation plots can aid in interpreting segmentation results. For example, in one plot, segments are differentiated based on their preferences for certain travel motives, such as luxury, budget considerations, or cultural experiences.

**Limitations of Projection:** It's important to note that each segment separation plot visualizes only one possible projection, and segments that overlap in one projection may not overlap in others. However, well-separated segments in a single projection can still indicate distinct differences between them in terms of travel motives.




## Step 6 Checklist

| Task | Who is responsible? | Completed? |
|---|---|---|
| Use the selected segments from Step 5. | | ☐ |
| Visualise segment profiles to learn about what makes each segment distinct. | | ☐ |
| 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. | | ☐ |
| Pass on the remaining segments to Step 7 for describing. | | ☐ |

# Step 7: Describing Segments

## Developing a Complete Picture of Market Segments

Segment profiling and segment description are crucial steps in market segmentation analysis, focusing on understanding the differences and characteristics of market segments. While they share similarities, they also have distinct purposes and methodologies.

### 1. Segment Profiling:

- Involves investigating differences between segments based on the segmentation variables used to extract the segments.

- Segmentation variables are typically chosen early in the analysis process and form the basis for extracting market segments from empirical data.

- Profiling focuses on variables directly related to the segmentation criteria, such as travel motives in the context of the Australian travel motives dataset.

- Provides insights into the differences in preferences, behaviours, or attitudes among segments based on the chosen segmentation variables.

### 2. Segment Description:

- Like profiling but includes additional information about segment members beyond the segmentation variables.

- Utilizes descriptor variables such as age, gender, past behaviour, media use, and spending patterns to provide a more comprehensive understanding of each segment.

- Helps in developing a customized marketing mix tailored to each segment's specific characteristics and preferences.

- Enables marketers to identify tangible ways of communicating with and targeting each segment effectively.

# Using Visualisations to Describe Market Segments

A wide range of charts exist for the visualisation of differences in descriptor variables. Here, we discuss two basic approaches suitable for nominal and ordinal descriptor variables (such as gender, level of education, country of origin), or metric descriptor variables (such as age, number of nights at the tourist destinations, money spent on accommodation).

Using graphical statistics to describe market segments has two key advantages: it simplifies the interpretation of results for both the data analyst and the user, and integrates information on the statistical significance of differences, thus avoiding the over-interpretation of insignificant differences.

When analyzing nominal and ordinal descriptor variables in the context of market segmentation, cross-tabulations are a fundamental tool. They allow for a visual representation of the relationship between segment membership and the descriptor variable. Here's a breakdown of how this analysis is conducted and interpreted using the Australian travel motives dataset:

## Nominal and Ordinal Descriptor Variables

### Cross-Tabulation Visualization:

Cross-tabulations of segment membership with nominal or ordinal descriptor variables provide insights into the distribution of characteristics across segments. Visualizations such as stacked bar charts or mosaic plots are used to represent these cross-tabulations. Stacked bar charts display segment sizes and the distribution of the descriptor variable within each segment. However, comparing proportions across segments can be challenging, especially when segment sizes are unequal.

Mosaic plots offer a solution by visualizing absolute segment sizes along with the proportion of the descriptor variable within each segment. They also allow for the integration of inferential statistics by highlighting differences between observed and expected frequencies.

### Interpretation:

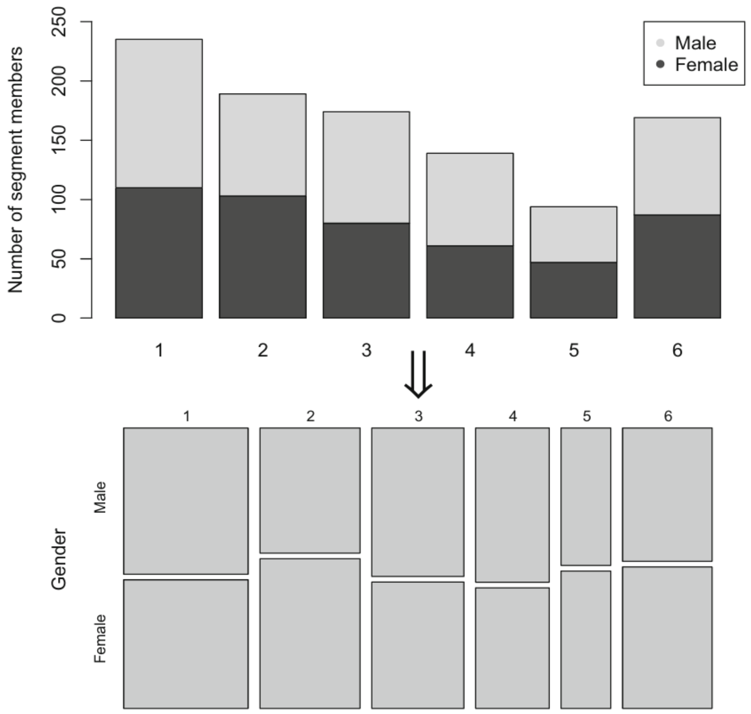
In the mosaic plot, the width of each column represents the segment size, while the height of the rectangles within each column represents the proportion of the descriptor variable (e.g., gender, income bracket, moral obligation level) within that segment.

Differences between observed and expected frequencies are indicated by cell colouring, with negative differences coloured in red and positive differences in blue. White cells signify statistically insignificant differences.By analyzing these visualizations, insights can be gained into how different segments vary in terms of the descriptor variable.

## Association Analysis:

These visualizations can also reveal associations between segment membership and the descriptor variable.

Visualizing cross-tabulations of nominal and ordinal descriptor variables with segment membership provides valuable insights into segment characteristics and associations, aiding marketers in developing targeted strategies tailored to each segment's preferences and behaviours




# Metric Descriptor Variables

The R package lattice is useful for creating conditional plots, which divide visualizations into sections or facets, each presenting results for a subset of the data (e.g., different market segments). Here's how you can use lattice to visualize differences between market segments using metric descriptor variables:

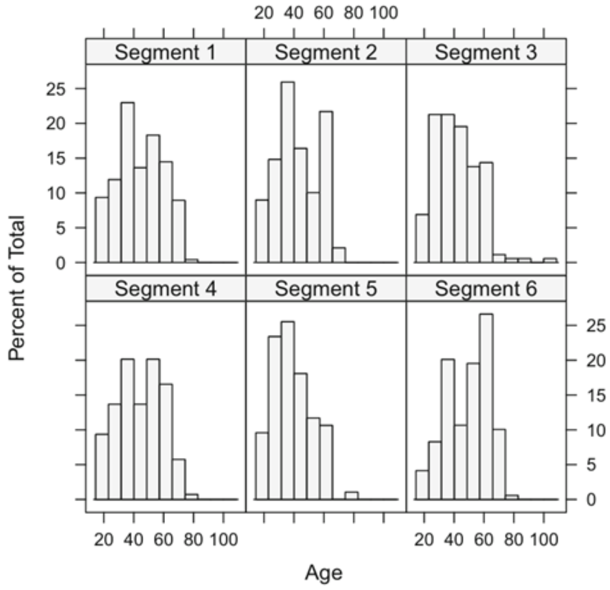
## Histograms for Age and Moral Obligation:

You can create histograms to compare the age distribution or the distribution of the original metric moral obligation scores across different market segments. The histogram function in lattice allows you to specify the variable of interest (Age or Obligation) conditioned on segment membership. These histograms provide a visual overview of the distributions within each segment but might not reveal significant differences between segments easily.

## Parallel Box-and-Whisker Plots:

Another approach is to use parallel box-and-whisker plots, which show the distribution of the variable separately for each segment. The boxplot function in R can create these plots, with options to customize axis labels (xlab and ylab).By setting Var width = TRUE and notch = TRUE, you can incorporate elements of statistical hypothesis testing into the plot.

These plots allow for a more detailed comparison of medians and confidence intervals between segments.



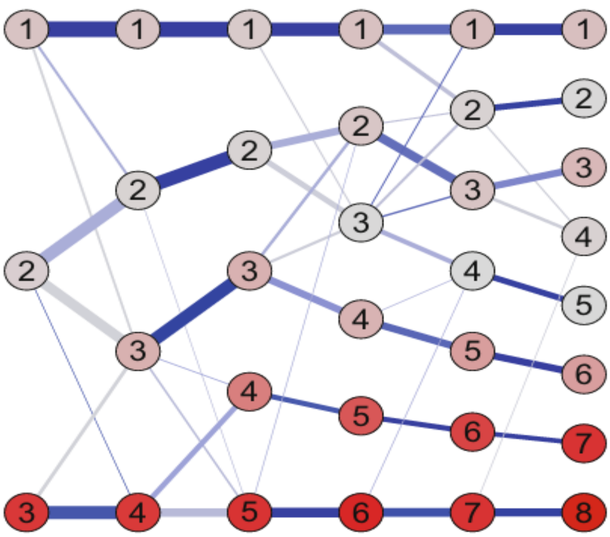

## Segment Level Stability Across Solutions (SLSA) Plot:

You can modify the SLSA plot to trace the value of a metric descriptor variable over a series of market segmentation solutions In this modified version, different colors are used to represent the value of the descriptor variable for each segment.

The slsa plot function can be used to create this plot, with the nodecol argument specifying the color based on the moral obligation score in this case.

This plot provides a visual representation of how the moral obligation varies across different segments and segmentation solutions.

By using these visualization techniques, you can gain insights into the differences in metric descriptor variables across market segments, aiding in the segmentation analysis and decision-making process.



# Testing for Segment Differences in Descriptor Variables

In segment profiling, simple statistical tests can be used to formally test for differences in descriptor variables across market segments. Here's a summary of the testing methods described:

## 1. Chi-squared Test:

- The chi-squared test can be used to test for differences in nominal variables like gender distribution across market segments.
- This test compares observed and expected frequencies and provides a p-value indicating the likelihood of observing the data if there is no association between the variables.
- In R, you can conduct the chi-squared test using the chisq.test function.

## 2. Analysis of Variance (ANOVA):

- ANOVA can be used to test for differences in means of metric variables (e.g., moral obligation scores) across multiple market segments.
- ANOVA compares the variance between segment means to the variance within segments and provides an F-value and p-value.
- A significant p-value indicates at least two segments differ in their mean values.
- In R, you can perform ANOVA using the aov function and examine the results with summary.

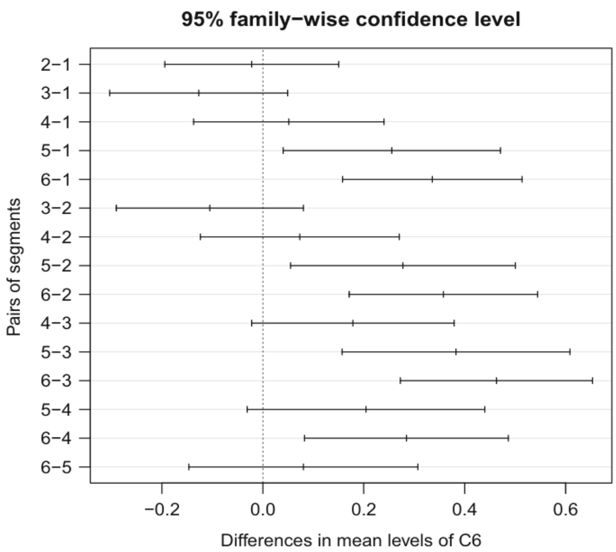
## 3. Pairwise Comparisons:

- If ANOVA indicates significant differences, pairwise t-tests can be conducted to identify which segments differ from each other.
- These tests compare the means of two segments at a time and provide p-values adjusted for multiple testing.
- In R, pairwise t-tests can be conducted using the pairwise. Test function.

## 4. Tukey's Honest Significant Differences:

- Tukey's HSD test provides adjusted pairwise comparisons with confidence intervals to identify significantly different segment pairs.
- It adjusts for multiple comparisons and provides a clearer understanding of the direction and significance of differences.
- In R, you can visualize Tukey's HSD results using the plot function on the output of TukeyHSD.

These tests help identify statistically significant differences in descriptor variables between market segments, providing valuable insights for segmentation analysis and decision-making. Adjustments for multiple testing, such as Holm's method or the false discovery rate procedure, ensure appropriate control of the overall error rate.




## Predicting Segments from Descriptor Variables

Predicting market segment membership from descriptor variables involves using regression models with segment membership as the categorical dependent variable and descriptor variables as independent variables. Here's a breakdown of the process and methods:

**Linear Regression Model:** The basic regression model assumes a linear relationship between the dependent variable (segment membership) and independent variables (descriptor variables).

In R, the lm() function fits a linear regression model. The formula specifies the dependent variable on the left side of ~ and the independent variables on the right side.

For categorical variables like segment membership, R automatically creates dummy variables for each category, and the intercept or one category can be dropped for identifiability.

Regression coefficients indicate the mean difference in the dependent variable for each category compared to a reference category (or the intercept).

Linear regression assumes a normal distribution for the dependent variable.

### Generalized Linear Models (GLMs):

GLMs extend the linear regression framework to accommodate a wider range of distributions for the dependent variable. GLMs are suitable for cases where the normal distribution assumption of linear regression does not hold, such as categorical dependent variables.

GLMs introduce a link function to transform the mean value of the dependent variable to an unlimited range, allowing for modelling with a linear function.

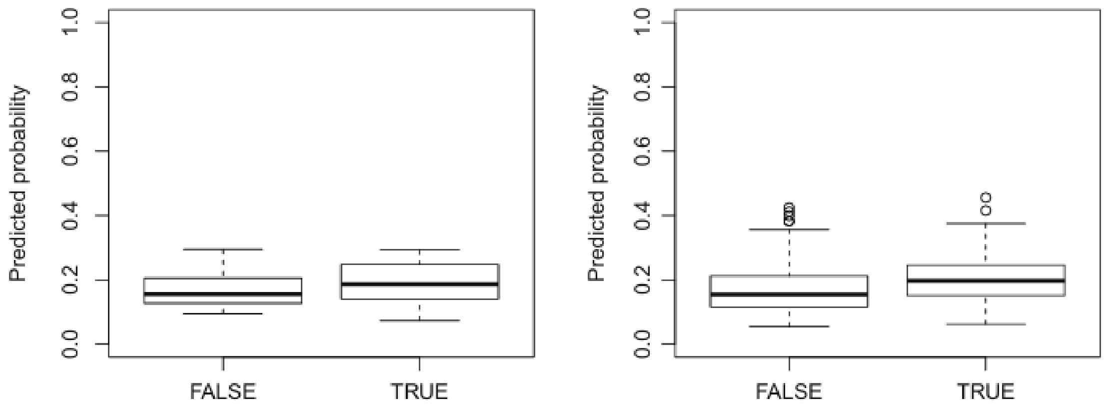
Common distributions used in GLMs include normal, Poisson, binomial, and multinomial distributions.

The choice of distribution and link function depends on the nature of the dependent variable and the problem at hand.

### Logistic Regression:

Logistic regression is a type of GLM used for binary classification problems where the dependent variable follows a binary distribution.

The logistic function (or logit function) is used as the link function to model the probability of belonging to a particular segment.




Logistic regression estimates the odds of being in one segment versus another based on the values of the independent variables.

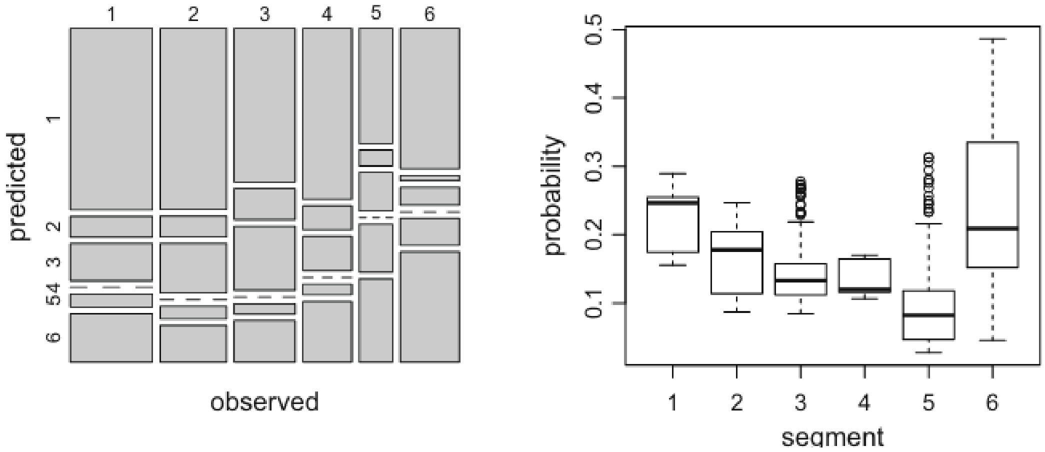
## Multinomial Logistic Regression:

Multinomial logistic regression extends logistic regression to handle multinomial (more than two) categories for the dependent variable.

It models the probability of membership in each segment relative to a reference category using the softmax function.

Multinomial logistic regression estimates separate sets of coefficients for each category of the dependent variable relative to the reference category.

These methods allow for the prediction of market segment membership based on descriptor variables and provide insights into which variables are critical for identifying segment membership.




Classification and Regression Trees (CARTs) offer an alternative approach to modeling and predicting binary or categorical dependent variables based on independent variables. Here's a summary of how CARTs work and their characteristics:

## Tree-Based Methods

### 1.Supervised Learning Technique:

- CARTs are a supervised learning technique from machine learning used for predictive modeling.
- They predict the value of a dependent variable by learning simple decision rules inferred from the data.

### 2.Advantages:

- Ability to perform variable selection: CARTs automatically select the most relevant variables for prediction.

- Ease of interpretation: CARTs generate decision trees that are easy to interpret and visualize, making them suitable for explaining the model to non-technical stakeholders.

- Incorporation of interaction effects: CARTs can capture interaction effects between variables in the form of decision rules.

### 3. Disadvantages:

- Instability: Results from CARTs can be sensitive to small changes in the data, leading to different trees with similar performance metrics.

### 4. Recursive Partitioning:

- CARTs use a stepwise procedure called recursive partitioning to build the model.

- At each step, the algorithm splits the data into groups based on the values of independent variables.

- The goal of the split is to maximize the purity of the resulting groups with respect to the dependent variable.

### 5. Tree Construction Algorithms:

- CART algorithms differ in several aspects:

  - Splits into two or more groups at each node (binary vs. multi-way splits)

  - Selection criterion for the independent variable for the next split

  - Selection criterion for the split point of the independent variable

  - Stopping criterion for the stepwise procedure

  - Final prediction at the terminal node

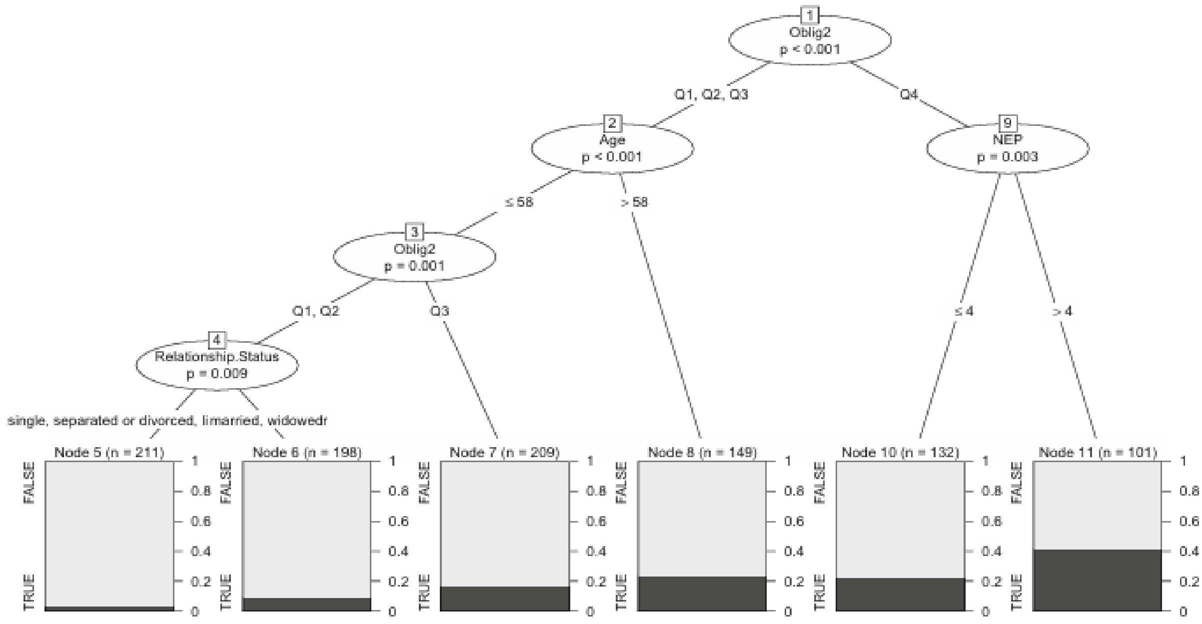



**Fig. 9.16** Conditional inference tree using membership in segment 6 as dependent variable for the Australian travel motives data set

## Step 7 Checklist

| Task | Who is responsible? | Completed? |
| --- | --- | --- |
| Bring across from Step 6 (profiling) one or a small number of market segmentation solutions selected on the basis of attractive profiles. | | ☐ |
| Select descriptor variables. Descriptor variables are additional pieces of information about each consumer included in the market segmentation analysis. Descriptor variables have not been used to extract the market segments. | | ☐ |
| Use visualisation techniques to gain insight into the differences between market segments with respect to descriptor variables. Make sure you use appropriate plots, for example, mosaic plots for categorical and ordinal descriptor variables, and box-and-whisker plots for metric descriptor variables. | | ☐ |
| Test for statistical significance of descriptor variables. | | ☐ |
| If you used separate statistical tests for each descriptor variable, correct for multiple testing to avoid overestimating significance. | | ☐ |
| "Introduce" each market segment to the other team members to check how much you know about these market segments. | | ☐ |
| Ask if additional insight into some segments is required to develop a full picture of them. | | ☐ |