Task 1: Summarization Task

Priya Navale

Step 1: Deciding (not) to Segment

1.1 Implications of Committing to Market Segmentation

Market segmentation is a major marketing strategy but isn't always the best choice for every organization. Before starting a market segmentation analysis, it's important to understand its long-term commitment and implications.

1. **Long-term Commitment**: Market segmentation requires a long-term commitment and substantial changes and investments.
2. **Costs**: Segmenting a market is expensive, involving research, surveys, focus groups, and creating multiple packages and advertisements.
3. **Profit Justification**: The increase in sales must justify the costs of segmentation. It should be more profitable than marketing without segmentation.
4. **Organizational Changes**: Changes may include new products, modified products, pricing adjustments, changes in distribution channels, and market communications.
5. **Internal Structure**: Organizations may need to adjust their internal structure, focusing on market segments rather than products. Strategic business units should manage different segments.
6. **Executive Decision**: The decision to pursue segmentation should be made at the highest executive level and communicated across all organizational levels.

1.2 Implementation Barriers

1. Books on Market Segmentation:

- Various books highlight barriers to successful market segmentation (e.g., Dibb and Simkin 2008, Croft 1994, McDonald and Dunbar 1995).

2. Senior Management Barriers:

- Lack of leadership, commitment, and involvement.
- Insufficient resources for analysis and long-term implementation.

3. Organizational Culture Barriers:

- Lack of market orientation.
- Resistance to change, lack of creativity, poor communication, and short-term thinking.
- Office politics and unwillingness to change.
- Croft (1994) created a questionnaire to assess market orientation barriers.

4. Training and Knowledge:

- Lack of understanding of market segmentation fundamentals.
- Importance of formal marketing expertise and qualified data analysts.

5. Resource Constraints:

- Limited financial resources.
- Inability to make necessary structural changes.

6. **Process-Related Barriers:**
    - Undefined objectives, poor planning, lack of structured processes, unclear responsibilities, and time pressure.
7. **Operational Barriers:**
    - Management's reluctance to use techniques they don't understand.
    - Importance of making segmentation analysis easy to understand with graphical visualizations.
8. **Proactive Identification and Removal of Barriers:**
    - Identify barriers early and remove them if possible.
    - If barriers persist, consider abandoning segmentation efforts.
9. **McDonald and Dunbar's Recommendation:**
    - Require a strong sense of purpose, patience, and willingness to address problems in implementation.

Step 2: Specifying the Ideal Target Segment

2.1 Segment Evaluation Criteria

User involvement is crucial throughout the entire market segmentation analysis, not just at the beginning or end. After committing to a segmentation strategy in Step 1, the organization must make a major conceptual contribution in Step 2, guiding subsequent steps, especially data collection (Step 3) and target segment selection (Step 8). In Step 2, the organization defines two sets of criteria: knock-out criteria, which are essential and non-negotiable features for target segments, and attractiveness criteria, which evaluate the attractiveness of segments that meet the knock-out criteria. The literature provides various segment evaluation criteria without generally distinguishing between these two types, offering detailed descriptions at different levels. Knock-out criteria automatically eliminate unsuitable segments, while attractiveness criteria are used to rank and evaluate the remaining segments.

Table 4.1 Overview:

- Lists various evaluation criteria proposed by different sources.
- Includes criteria like measurability, size, growth potential, competitive advantage, and compatibility with the company.

Two Types of Criteria:

1. Knock-Out Criteria:
    - Essential, non-negotiable features.
    - Automatically eliminate segments that do not meet these criteria.
    - Not subject to negotiation by the segmentation team.
2. Attractiveness Criteria:

- Longer and more diverse list.
- Represents a "shopping list" for the segmentation team.
- Used to evaluate the relative attractiveness of market segments that meet knock-out criteria.
- Criteria selection and importance are negotiated by the segmentation team.
- Determines overall attractiveness in Step 8.

Process:

- Step 2: Organization determines knock-out and attractiveness criteria.
- Step 8: Attractiveness criteria are applied to rank and select target segments.

Key Points:

- Knock-out criteria ensure only suitable segments are considered.
- Attractiveness criteria allow for a nuanced evaluation of potential segments.
- Team involvement is crucial in selecting and weighting attractiveness criteria.

2.2 Knock-Out Criteria

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.

Original Criteria by Kotler (1994):

1. Substantiality: The segment must be large enough to be profitable.
2. Measurability: The segment’s size and purchasing power can be measured.
3. Accessibility: The segment can be effectively reached and served.

Additional Criteria (Kotler and Others):

1. Homogeneity: Members of the segment must be similar to one another.

2. Distinctness: The segment must be distinctly different from other segments.

3. Size:

    - The segment must be large enough to justify customized marketing efforts.
    - The minimum viable target segment size needs to be specified.

4. Organizational Fit:

    - The segment must match the strengths and capabilities of the organization.

5. Identifiability: It must be possible to identify and spot segment members in the marketplace.

6. Reachability: There must be effective ways to reach and communicate with segment members.

Importance for Stakeholders

Senior Management, Segmentation Team, and Advisory Committee must understand and agree upon these criteria.

Specification Required: For some criteria, such as size, exact thresholds need to be determined.

These criteria help in filtering out unviable segments early in the market segmentation process, ensuring that resources are focused on the most promising and manageable segments.

2.3 Attractiveness Criteria

- **Wide Range of Criteria**: Table 4.1 provides various criteria for evaluating segment attractiveness.
- **Non-Binary Assessment**: Attractiveness criteria are not yes/no (binary) but rather a spectrum.
- **Rating Segments**: Each market segment is rated based on how attractive it is according to specific criteria.
- **Overall Attractiveness**: The combined attractiveness across all criteria helps determine which segments to target in Step 8 of the market segmentation analysis.

2.4 Implementing a Structured Process

When assessing market segments, a structured process is widely endorsed. One prominent method is the segment evaluation plot, which maps segment attractiveness against organizational competitiveness. However, the criteria for determining attractiveness and competitiveness are unique to each organization and require negotiation and consensus. It's recommended to limit these criteria to no more than six factors to maintain focus. Ideally, a diverse team from various organizational units should be involved in this process. While a core team can propose initial criteria, an advisory committee, comprising representatives from all units, should discuss and refine them. This diversity ensures that different perspectives on the organization's business are considered, as the segmentation strategy impacts all units. Selecting attractiveness criteria early in the process facilitates data collection and simplifies target segment selection later on. Each criterion should be weighted to reflect its importance, with allocations negotiated until agreement is reached. Seeking approval from the advisory committee ensures a balanced perspective, considering the interests of all organizational units involved. This structured approach ensures that segment assessment is thorough, collaborative, and aligned with organizational objectives.

Step 3: Collecting Data

3.1 Segmentation Variables

Empirical data is essential for both commonsense and data-driven market segmentation, as it helps identify or create market segments and describe them in detail. In commonsense segmentation, a single characteristic of consumers, known as the segmentation variable, is used to divide the market into segments. For example, gender can be used to create segments of men and women. Other characteristics in the data, such as age, the number of vacations taken, and the benefits people seek on vacation, serve as descriptor variables. These variables provide detailed descriptions of the segments, which is critical for developing effective marketing strategies. Descriptor variables often include socio-demographic information and media behavior, enabling marketers to target segments with specific communication messages. In contrast, data-driven market segmentation uses multiple characteristics to identify or create market segments, offering a more nuanced and comprehensive understanding of the market. While commonsense segmentation is simpler and relies on one characteristic, data-driven segmentation is more complex and provides deeper insights by considering multiple characteristics.

Table 3.1 Gender as a possible segmentation variable in commonsense market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	Nº of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	87	2	1	0	1	0	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
segmentation variable	descriptor variables						

Table 3.2 Segmentation variables in data-driven market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	Nº of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Male	87	2	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
descriptor variables			segmentation variables				

Empirical data is vital for market segmentation, allowing for the creation and detailed description of segments. It can come from surveys, observations, or experimental studies, with accurate data being essential for effective segmentation and marketing strategies. In commonsense segmentation, one characteristic (e.g., gender) is used to split the market, with other characteristics (e.g., age, vacation habits) describing the segments. For example, Table 5.1 shows segmentation by gender.

Data-driven segmentation uses multiple characteristics to identify segments, as shown in Table 5.2, where benefits sought on vacation are used for segmentation, and characteristics like gender and age are descriptors.

High-quality data ensures accurate segment assignment and description, aiding in tailored marketing. While surveys are common, they might not always reflect true behavior, so using the most accurate data sources is crucial.

3.2 Segmentation Criteria

Before extracting market segments, an organization must choose a segmentation criterion, which is broader than a segmentation variable and involves the type of information used for segmentation. This decision requires market knowledge and can't be outsourced. Common criteria include geographic, sociodemographic, psychographic, and behavioral factors. The best approach is the simplest effective one. If demographic or geographic segmentation works, use it. More complex methods, like psychographic segmentation, aren't necessarily better. The aim is to find the most effective, cost-efficient method for your product or service.

3.2.1 Geographic Segmentation

Geographic information is the original market segmentation criterion, often using the consumer's location of residence. This approach is simple and useful, especially when targeting different languages or regional preferences, as seen with Austria's tourism board or companies like Amazon and IKEA.

Advantages of Geographic Segmentation:

- Easy to assign consumers to geographic units.
- Simplifies targeting communication through local media.

Disadvantages:

- Same location doesn't mean shared characteristics relevant to marketers, like product benefits sought.
- Location often isn't the main reason for differences in product preference.

For example, people in luxury suburbs may share luxury car preferences due to socio-demographic factors rather than location. In tourism, people from the same country can have diverse holiday preferences based on personal factors.

Despite its limitations, geographic segmentation is experiencing a revival in international market studies, aiming to identify segments across borders. This is challenging due to the need for meaningful segmentation variables across regions and potential biases from different cultural backgrounds. An example is Haverila's study on mobile phone users among young customers across countries.

3.2.2 Socio-Demographic Segmentation

Socio-demographic segmentation uses criteria like age, gender, income, and education. It is useful in industries such as luxury goods (high income), cosmetics (gender), baby products (gender), retirement villages (age), and tourism resorts (having children).

Advantages:

- Easy to determine segment membership for each consumer.
- Sometimes explains product preferences (e.g., families choosing family vacation spots).

Disadvantages:

- Often does not explain the cause of product preferences.
- Provides limited market insight for optimal segmentation.

Research suggests that socio-demographics explain only a small part of consumer behavior variance. Values, tastes, and preferences are often more influential for market segmentation, offering better insights into consumers' buying decisions.

3.2.3 Psychographic Segmentation

Psychographic segmentation groups people based on psychological criteria like beliefs, interests, preferences, aspirations, or benefits sought. It covers all measures of the mind and includes approaches like benefit segmentation and lifestyle segmentation.

Advantages:

- Reflects the underlying reasons for consumer behavior differences.
- For example, tourists motivated by culture are likely to choose cultural destinations.

Disadvantages:

- More complex to determine segment memberships.
- Depends heavily on the reliability and validity of measures used.

Psychographic segmentation often uses multiple variables, making it more complex than geographic or socio-demographic segmentation. However, it provides deeper insights into consumer behavior.

3.2.4 Behavioural Segmentation

Behavioral segmentation involves grouping people based on their actual behaviors or reported behaviors, such as prior product experience, purchase frequency, amount spent, and information search habits. This approach can provide deeper insights than geographic or demographic criteria, as it focuses on behavior that matters most.

Advantages:

- Uses actual behavior for segmentation.
- Provides insights directly relevant to consumer preferences.
- Avoids the need for complex psychological measures.

Disadvantages:

- Behavioral data may not always be available.
- Limited to existing customers if data is not accessible for potential customers.

Examples include using consumer expenses or purchase data for segmentation. This approach can offer valuable insights without the need for developing psychological measures.

3.3 Data from Survey Studies

Market segmentation frequently hinges on survey data due to its cost-effectiveness and accessibility. However, survey data can be prone to biases, which might compromise the accuracy of segmentation outcomes. Addressing these biases is crucial for ensuring the reliability of segmentation solutions.

3.3.1 Choice of Variables

The selection of variables in both commonsense and data-driven segmentation is crucial for quality segmentation outcomes. In data-driven segmentation, all relevant variables must be included, while unnecessary ones should be avoided to prevent respondent fatigue and improve analysis accuracy. Noisy variables, which don't contribute useful information, can disrupt segmentation algorithms. Avoiding noisy variables requires careful questionnaire development and variable selection.

Conducting exploratory research helps identify important variables and ensures comprehensive questionnaire design for accurate segmentation.

3.3.2 Response Options

The options given to respondents in surveys determine the type of data available for analysis. Binary responses (e.g., yes/no) are straightforward for analysis, represented as 0s and 1s. Nominal responses (e.g., occupation) can be transformed into binary data. Metric data (e.g., age) allow for statistical analysis and are ideal for segmentation. Ordinal responses (e.g., agreement on a scale) are common but pose challenges as the distance between options isn't clearly defined. It's best to use binary or metric options to avoid complications in segmentation analysis. Visual analogue scales, like slider scales, are useful for capturing nuanced responses and generate metric data. Binary options often perform better than ordinal ones, especially when formulated without specific levels.

3.3.3 Response Styles

Survey data can be influenced by biases, including response biases where respondents consistently answer in a certain way regardless of the question. Response styles, such as always agreeing or using extreme options, can affect segmentation results. For example, if a segment consistently agrees with all questions about spending habits on a vacation, it may seem like an attractive market segment, but it could just be a response style. It's important to minimize response styles in segmentation data to ensure accurate results. If segments with potentially biased response patterns emerge, further analysis or removing affected respondents may be necessary before targeting those segments.

3.3.4 Sample Size

Market segmentation analysis does not have specific sample size recommendations, unlike many other statistical analyses. Insufficient sample size can make it difficult for segmentation algorithms to identify the correct number of market segments, even with simple segmentation variables. If the sample size is large enough, identifying segments becomes easier. Formann (1984) suggests a sample size of at least 2p2^p2p (preferably five times 2p2^p2p), where ppp is the number of segmentation variables, for latent class analysis with binary variables. However, this recommendation is specific and may not apply universally to all segmentation contexts.

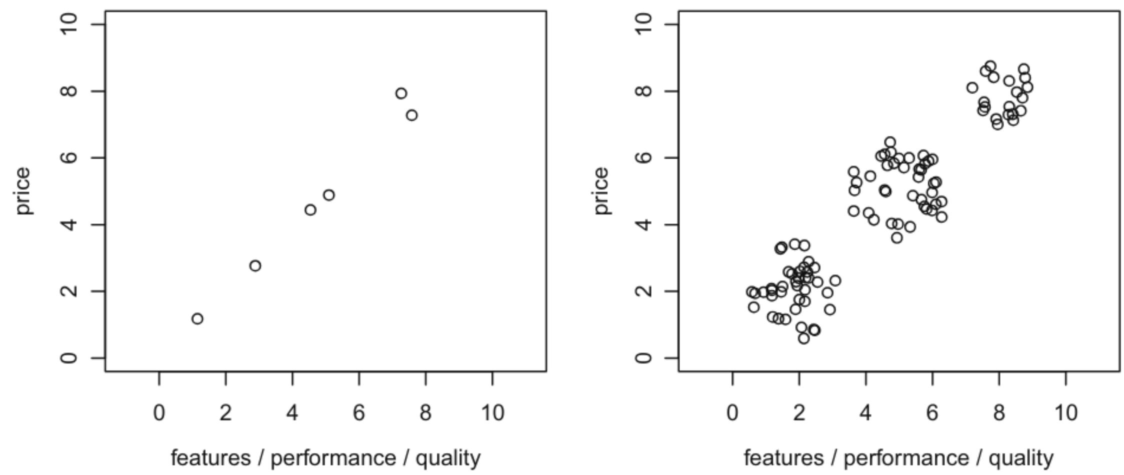

Fig. 3.1 Illustrating the importance of sufficient sample size in market segmentation analysis

Sample size is crucial for accurate market segmentation analysis. Formann (1984) suggests a sample size of at least 2^p for latent class analysis with binary variables, but this may not apply universally. Qiu and Joe (2015) recommend a sample size of at least ten times the number of segmentation variables times the number of segments (10.p.k) . If segments are uneven, the smallest segment should have at least 10.p samples. Dolnicar et al. (2014) found that larger sample sizes improve segment identification, measured by the adjusted Rand index, which shows the alignment between true and extracted segments.

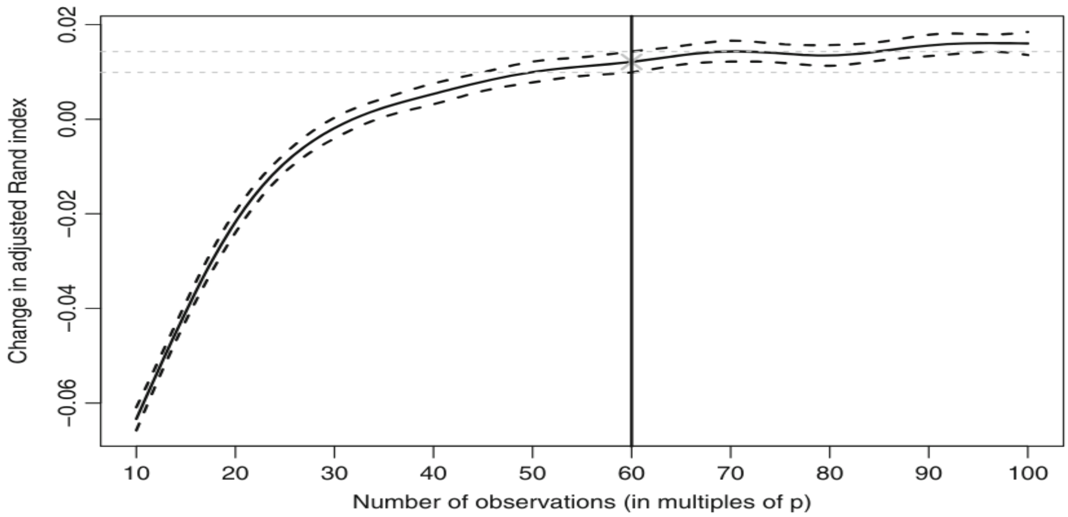

Fig. 3.2 Effect of sample size on the correctness of segment recovery in artificial data. (Modified from Dolnicar et al. 2014)

Increasing sample size enhances the accuracy of market segmentation, especially with small samples. Beyond a certain size, additional increases yield fewer benefits. A sample size of at least 60.p is advised for standard data, and $70 \cdot p$ for more complex data sets.

Dolnicar et al. (2016) found that various factors affect sample size needs, such as the number, size, and overlap of market segments, as well as survey data characteristics like sampling error, response biases, low data quality, varied response options, irrelevant items, and item correlation. Unequal and overlapping segments particularly challenge accurate segment extraction.

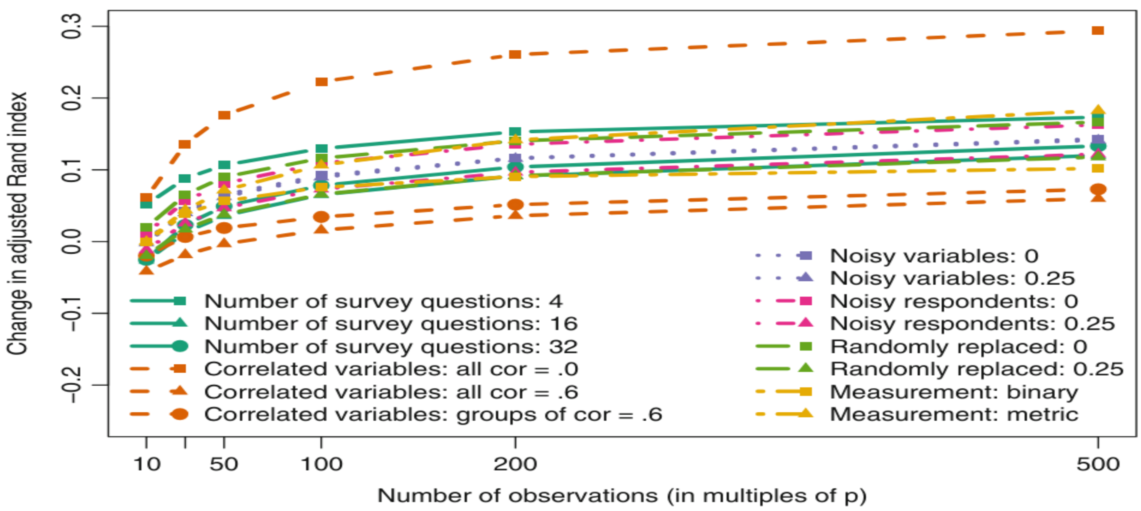

Fig. 3.3 Sample size requirements in dependence of market and data characteristics. (Modified from Dolnicar et al. 2016)

A large-scale simulation study using artificial data found that increasing sample size improves the accuracy of market segmentation, but the extent varies based on market and data characteristics. Correlated segmentation variables pose significant challenges that can't be fully offset by larger sample sizes, unlike uncorrelated variables, which lead to better segment recovery. The study recommends having at least 100 respondents per segmentation variable to ensure accuracy.

To achieve optimal market segmentation, data should:

- Include all necessary items.
- Exclude unnecessary and correlated items.
- Contain high-quality, unbiased responses.
- Use binary or metric responses.
- Be free of response styles.
- Have a sufficient sample size, ideally 100 times the number of segmentation variables.

3.4 Data from Internal Sources

Organizations can use internal data, such as scanner data from grocery stores, airline booking data, and online purchase data, for market segmentation. This data reflects actual consumer behavior, avoiding biases and memory errors common in self-reported data. Additionally, this data is automatically generated, reducing collection effort. However, internal data may be biased towards existing customers and lack information about potential future customers with different consumption patterns.

3.5 Data from Experimental Studies

Experimental data, from field or lab experiments, can be used for market segmentation. For instance, consumer responses to ads or preferences in choice experiments can serve as segmentation criteria. Conjoint analyses reveal how different product attributes influence consumer choices, providing valuable data for segmentation.

Step 4: Exploring Data

4.1. A First Glimpse at the Data

After gathering data, the first step is to explore and clean it. This process is known as Exploratory Data Analysis (EDA). EDA helps in understanding the data better and preparing it for further analysis, such as identifying market segments. Here's what EDA involves:

1. Identify Measurement Levels: Determine if variables are nominal, ordinal, interval, or ratio.
2. Univariate Analysis: Examine the distribution of each variable (e.g., mean, median, range).
3. Assess Dependencies: Explore relationships between variables (e.g., correlations).

Preprocessing:

- Handle missing values
- Remove duplicates
- Normalize/standardize data
- Encode categorical variables

Application Example

Using a travel motives data set with 20 motives from 1000 Australian residents about their last vacation, you can:

- Load the Data: Use the R package MSA to get the CSV file.
- Explore and Clean: Identify variable types, investigate distributions, and preprocess the data.

This prepares the data for effective market segmentation analysis.

The Australian travel motives dataset comprises responses from 488 women and 512 men. The age distribution spans from 18 to 105 years, with a median age between 32 and 57 years. Income and Income2 variables reflect different income categorizations, with Income2 likely representing a modified version of Income. Both income variables contain missing data, with 66 respondents not providing income information. NA values in R denote these missing entries.

4.2 Data Cleaning

Before diving into data analysis, it's essential to clean the data to ensure accuracy and consistency. This involves checking for correct recording of values and consistent labels for categorical variables. Metric variables like age should fall within expected ranges, while categorical variables like gender should have permissible values. In the Australian travel motives dataset, no cleaning is needed for gender and age variables. However, the Income2 variable's summary reveals unordered categories due to how non-numeric data is handled in R. This can be corrected by reordering the categories, ensuring accurate representation of the data.

4.3 Descriptive Analysis

In R, the summary() function provides a quick overview of numeric variables like AGE, including range, quartiles, mean, and frequency counts for categorical variables, highlighting missing values. Histograms are useful for visualizing the distribution of numeric data like AGE, showing patterns and outliers. In R, you can use the lattice package for segment-wise histogram creation. Mosaic plots help descriptive statistics and graphical methods offer insights for informed analysis and decision-making.

Using finer bins in a histogram reveals detailed patterns, like the bimodal distribution seen around ages 35-40 and 60. Specifying type = "density" in R scales the y-axis to show density estimates, useful for overlaying probability density functions. Boxplots offer a concise summary of distributions without manual bin selection, showing minimum, quartiles, median, and maximum values. They're widely used in natural sciences, less so in business and social sciences. R defaults to displaying the five-number summary and mean in numeric summaries, aiding in understanding central tendency and spread.
The summary of the Australian travel motives study reveals key insights about the age distribution of respondents. The youngest participant is 18 years old, while a quarter are under 32, half are under 42, and three-quarters are under 57. The oldest respondent is either 105 years old or an outlier. The boxplot generated from this data illustrates the distribution, showing quartiles and outliers. In this case, the skewness of the age distribution is evident, with the median not centered in the box. The outlier, the 105-year-old respondent, significantly influences the whisker length. To address this, statistical packages like R often limit whisker length to avoid outlier dominance, displaying outliers as separate points. This ensures outlier information is preserved without distorting the overall distribution depiction.
The standard box-and-whisker plot for the variable AGE in R can be generated using the boxplot() function, specifying horizontal = TRUE for horizontal alignment. This plot visually displays the distribution of ages in the dataset. To showcase the value of graphical methods further, we visualize the percentage of agreement with travel motives from columns 13 to 32 in the Australian travel dataset. By calculating the mean percentage of "yes" responses for each motive, then sorting and

plotting them using a dot chart, we get an intuitive overview of the importance attributed to each motive. This chart reveals varying levels of agreement with different motives among respondents, highlighting the heterogeneity in preferences and confirming the suitability of these variables for segmentation analysis. In essence, graphical representations offer quick insights into complex data structures, aiding in understanding and decision-making.

4.4 Pre-Processing

4.4.1 Categorical Variables

Pre-processing for categorical data involves two main steps: merging categories that are too specific and turning categories into numbers. Merging helps when there are very few people in certain groups, making data more balanced. Converting categories to numbers works if we assume equal distances between them, like with income ranges. Likert scales, common in surveys, can also be treated this way, but it's important to note that sometimes people don't answer consistently. Using simple "yes" or "no" answers can avoid this issue. In R, we can easily convert these answers to numbers, making analysis easier. These steps make data more manageable for analysis but may change it slightly.

4.4.2 Numeric Variables

The range of values in segmentation variables can affect their impact in segmentation methods. For instance, if one variable is binary (like whether a tourist likes dining out), and another ranges from $0 to $1000 (like daily expenditure), small differences in spending may be considered as significant as the binary choice. To balance this influence, variables can be standardized, meaning they're transformed to a common scale. The usual method subtracts the average and divides by the standard deviation, making the mean 0 and the standard deviation 1. In R, this is easily done with the scale() function. However, if the data has outliers, different methods, like using the median and interquartile range, may be more appropriate.

4.5 Principal Components Analysis

Principal Components Analysis (PCA) is a technique that transforms a dataset with multiple variables into a new set of variables called principal components. These components are uncorrelated and ordered by importance, with the first one capturing the most variability in the data, and so on. PCA keeps the relative positions of observations unchanged but views the data from a different perspective. It typically works with the covariance or correlation matrix of numeric variables. In R, you can use the prcomp() function to perform PCA.

PCA is often used to reduce high-dimensional data into lower dimensions for visualization. It helps to identify which variables contribute most to the variation in the data. The rotation matrix obtained from PCA shows how the original variables contribute to each principal component. By plotting the principal components, we can visualize the data in a lower-dimensional space.

In PCA, the proportion of variance explained by each principal component is important. If only a small subset of components explains a large proportion of variance, it suggests that the original variables are not redundant. However, using a subset of principal components as segmentation variables is not recommended, as it may lead to loss of information. Instead, PCA can be used to identify redundant variables, which can then be removed to reduce the dimensionality of the dataset while retaining the original variables. This helps in exploratory analysis and identifying highly correlated variables.