Enginius

Segmentation

Puja Gundu, California state university- East bay

Segmentation options

Options selected

	Option	Selection
	Clustering method	Hierarchical
	Standardization method	standard
	Segments forced	3
	Run discriminant analysis	No
	Run classification analysis	No
	Date and time	2024-12-07 00:10:39 UTC

Options selected.

Data description

	Data	Number of Rows	Number of columns	Column names
1	Segmentation data	60	7	1. What is your age group?, 2. How do you plan your Black Friday shopping?, 3. How often do you shop during Black Friday?, 4. How do you prefer to shop on Black Friday?, 9. What is your preferred device for online Black Friday shopping?, ...

Data description.

Data transformation

The segmentation data has been scaled column wise

	Mean	Standard deviation
2. How do you plan your Black Friday shopping? = Impulse buying without planning	0.3333	0.4754
2. How do you plan your Black Friday shopping? = Decide on the day	0.3833	0.4903
3. How often do you shop during Black Friday? = Every year	0.5333	0.5031
3. How often do you shop during Black Friday? = Occasionally	0.3167	0.4691
4. How do you prefer to shop on Black Friday? = In-store	0.2000	0.4034
4. How do you prefer to shop on Black Friday? = Online	0.1833	0.3902
9. What is your preferred device for online Black Friday shopping? = Smartphone	0.5500	0.5017
9. What is your preferred device for online Black Friday shopping? = Laptop/Desktop computer	0.4000	0.4940
14. Which categories do you shop for most during Black Friday? = Electronics	0.3500	0.4810
14. Which categories do you shop for most during Black Friday? = Other	0.1500	0.3601
14. Which categories do you shop for most during Black Friday? = Apparel	0.4167	0.4972
15. How much do you typically spend during Black Friday? = More than 500	0.3833	0.4903
15. How much do you typically spend during Black Friday? = Less than 100	0.1333	0.3428

Mean and standard deviation column wise.

Segment solution

3-segment solution

The ideal number of segments is a function of statistical fit (what the data say), managerial relevance (what makes the most sense from a managerial point of view), and targetability (can the segments be easily targeted).

When the three criteria do not perfectly converge, selecting the right number of segments becomes a judgment call.

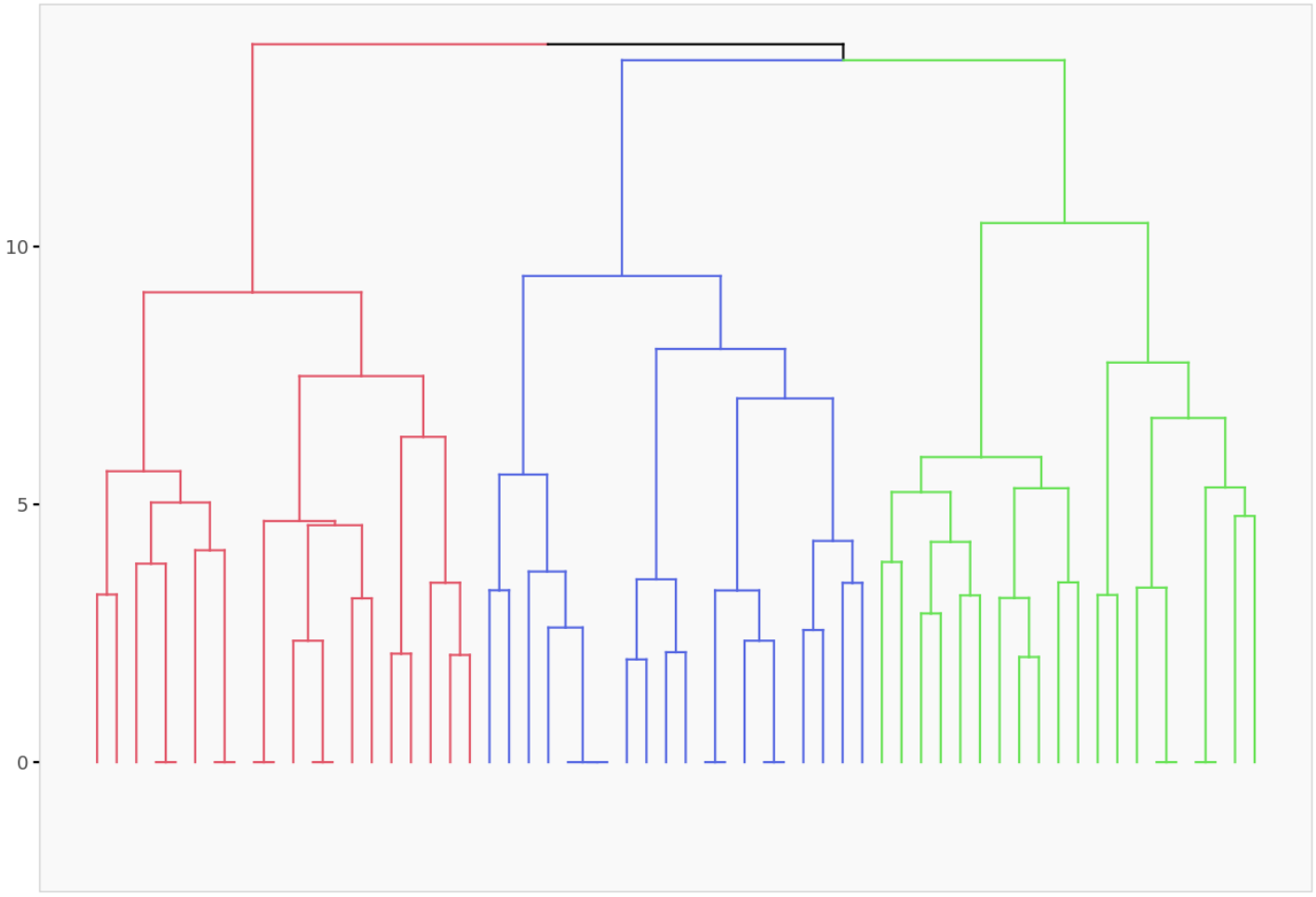
You have decided to perform the analysis with 3 segments.

The segmentation method relies on the hierarchical clustering approach. This approach generates a dendrogram that we display next.

Dendrogram

The dendrogram represents the grouping process of observations into clusters. The chart reads from bottom (all initial observations are separated) to top (all observations are clustered into one unique segment).

The height represents the distance between the two groups of observations being merged at each step. If two very distant groups are being merged, this will create a 'jump' in the dendrogram, indicating that it might be wise to stop the clustering process before.



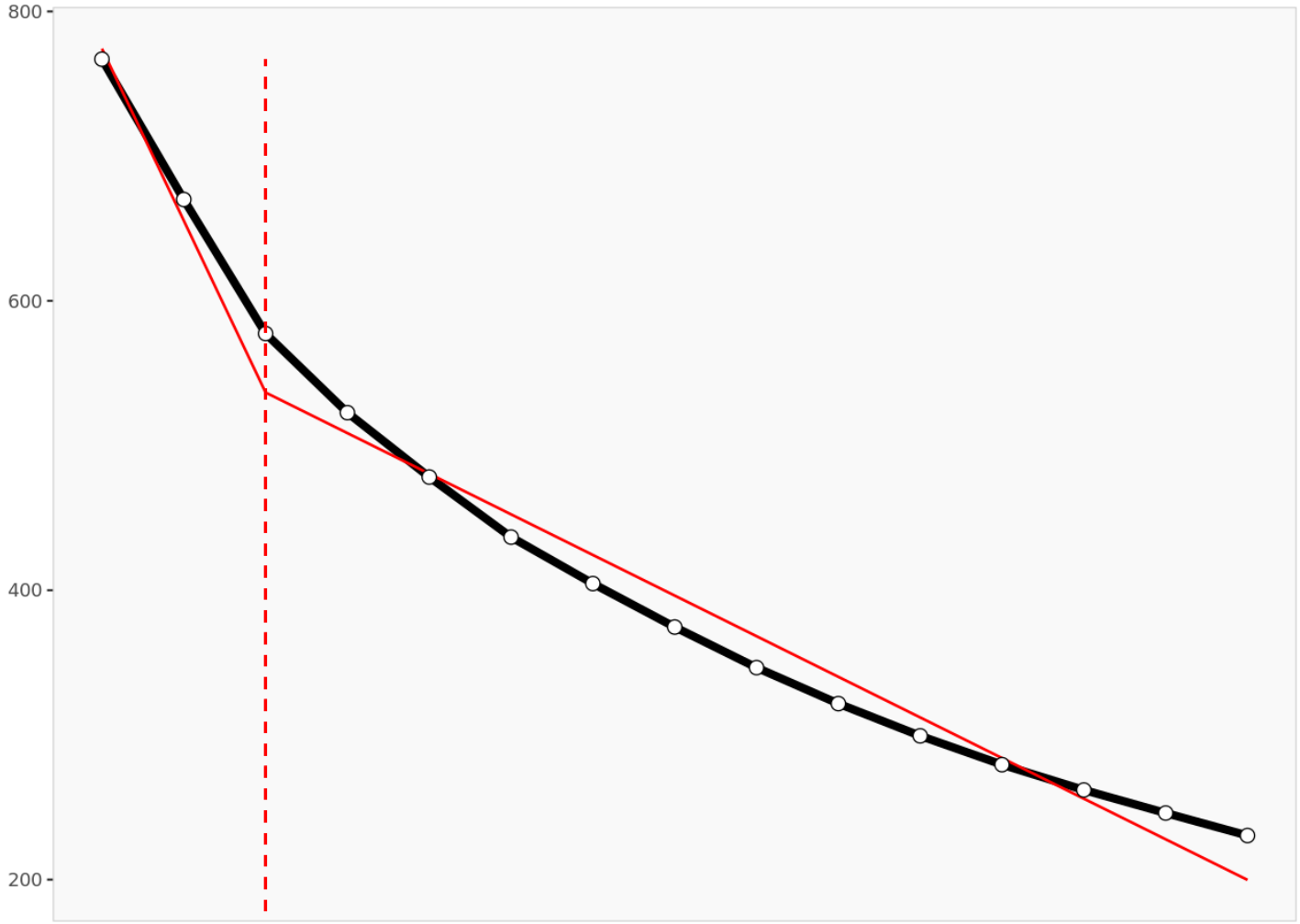
Dendrogram. The dendrogram is a tree diagram to illustrate the arrangement of clusters produced by hierarchical clustering, and how the observations are incrementally clustered together.

Scree plot

The screeplot displays, for each cluster solution, a measure of within-cluster heterogeneity. If clusters group observations that are widely different (which will happen if the number of clusters is too small to capture the variability in the data), the value will be high.

A good cluster solution might be where the screeplot displays an 'elbow', that is, where increasing the number of clusters beyond a certain point does not dramatically decreases within-cluster heterogeneity.

The measure displayed in the screeplot is related, but not equivalent, to the distance reported in the dendrogram.



Scree plot. The scree plot compares the sum of squared error (SSE) for each cluster solution. A good cluster solution might be when the SSE slows dramatically, creating an 'elbow'. Such elbow does not always exist. If number of segments is equal to maxumum possible segments elbow cannot be created.

From a statistical point of view, the SSE reported in the screeplot is computed as the sum of squared error between each observation and its cluster centroid (or center), summed over all the observations.

Segment description

Segment size

	Population	Segment 1	Segment 2	Segment 3
Size	60	20	20	20
Relative size	100%	33%	33%	33%

Segment size.

Segment description

	Population	Segment 1	Segment 2	Segment 3
2. How do you plan your Black Friday shopping? = Impulse buying without planning	0.333	0.400	0.200	0.400
2. How do you plan your Black Friday shopping? = Decide on the day	0.383	0.200	0.750	0.200
3. How often do you shop during Black Friday? = Every year	0.533	0.500	0.300	0.800
3. How often do you shop during Black Friday? = Occasionally	0.317	0.350	0.500	0.100
4. How do you prefer to shop on Black Friday? = In-store	0.200	0.100	0.200	0.300
4. How do you prefer to shop on Black Friday? = Online	0.183	0.300	0.100	0.150
9. What is your preferred device for online Black Friday shopping? = Smartphone	0.550	0.900	0.550	0.200
9. What is your preferred device for online Black Friday shopping? = Laptop/Desktop computer	0.400	0.100	0.300	0.800
14. Which categories do you shop for most during Black Friday? = Electronics	0.350	0.900	0.000	0.150
14. Which categories do you shop for most during Black Friday? = Other	0.150	0.000	0.000	0.450
14. Which categories do you shop for most during Black Friday? = Apparel	0.417	0.050	1.000	0.200

15. How much do you typically spend during Black Friday? = More than 500	0.383	0.300	0.300	0.550
15. How much do you typically spend during Black Friday? = Less than 100	0.133	0.150	0.100	0.150

Segment description. Average value of each segmentation variable, overall for each segment (centroid). Segmentation variables that are statistically different from the rest of the population are highlighted in red (lower) or green (higher).

Segment differences per segment. Cell colors indicate to what extent a segment is statistically different from the rest of the population on each segmentation variable.

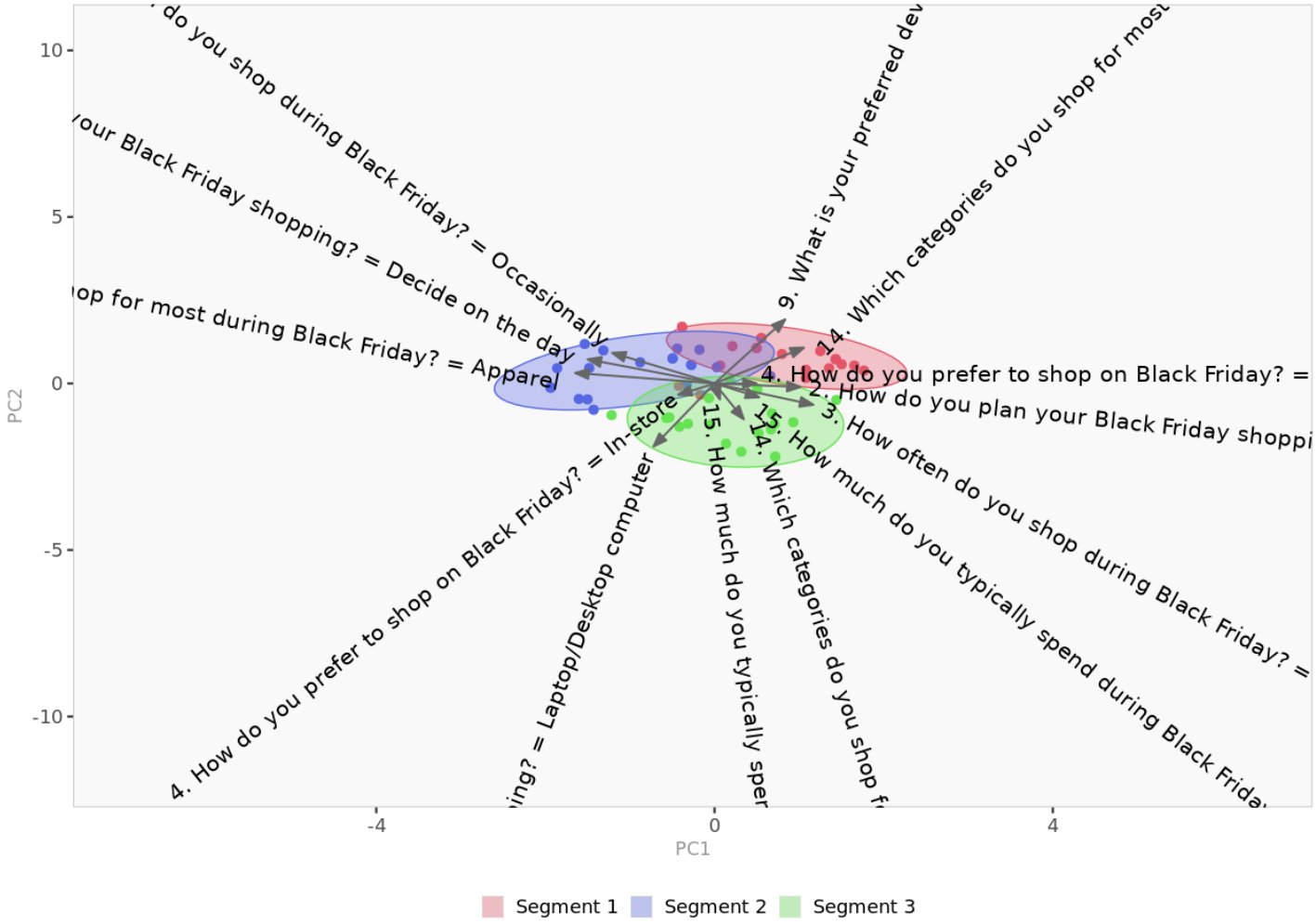
Segmentation space

The chart below is a graphical representation of the various segments, segment members, and segmentation variables. It is obtained by plotting the first two dimensions of a principal component analysis performed on the (standardized) segmentation data, on top of which segment information has been overlaid.

Because only the first two dimensions of the PCA are displayed, and these two dimensions capture only 36.3% of the variance in the data, some differences between segments might not appear here. Note that segmentation

variables with no variance, if any, have been excluded.

Two clusters that appear to overlap on the first two dimensions might be distinct on other dimensions. Consequently, this chart is a useful guide, for checking which variables are correlated, but may be misleading if used to select the optimal number of segments.



Segment space. Spatial representation of segments and segmentation variables, using principal component analysis.

Segment membership

	Segment
1	1
2	3
3	3
4	3
5	3
6	2
7	2
8	3
9	3
10	3

Segment membership (excerpt). Segment to which each member of the population belongs to. The complete membership list is only

available in the Excel formatted output.