MSBA 324. Assignment 3: R and Cluster Analysis For Acme Shopping
Stephan Sorger

Ainazik Amiraeva

You are the data scientist for Acme Shopping, a large (fictitious) shopping mall that incorporates a balanced tenancy of different types of stores along with eating areas. To understand your customers better, you conduct a survey to study beliefs on shopping. You test those beliefs by asking shoppers the six questions listed below. Shoppers then answer using a 7-point Likert scale, from 1 (strongly disagree) to 7 (strongly agree). Your goal is to group the shoppers at Acme Shopping into different clusters (segments) using Ward's method. The questions are listed below:
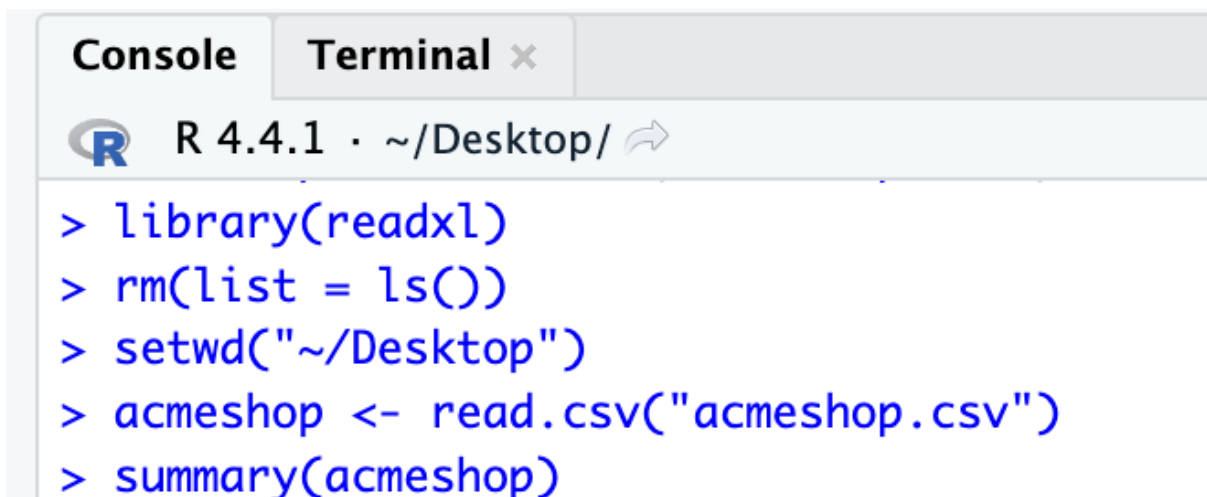
V1. Shopping is fun
V2. Shopping is bad for your budget
V3. I combine shopping with eating out
V4. I try to get the best buys while shopping
V5. I don't care about shopping
V6. You can save a lot of money by comparing prices

You also collect other data, such as income and number of mall visits.

1. Cluster Analysis Data Preparation
Prepare the data by eliminating any portion of the dataset that you do not need, such as the descriptive material in the spreadsheet that is not part of the data. Save the data in comma separated values (CSV) format using the filename "acmeshop". Read the data into R using the read.csv command. Use the summary() function in R to check the summary statistics of the "acmeshop" file after you have read it into R. Check for any problems or outlier values. Show a screenshot of the output of the summary() function.

We cleared the environment in R and uploaded a CSV dataset to start cluster analysis. Here are the screenshots of Rstudio and the outputs:


```
Console    Terminal ×

R    R 4.4.1 · ~/Desktop/
> library(readxl)
> rm(list = ls())
> setwd("~/Desktop")
> acmeshop <- read.csv("acmeshop.csv")
> summary(acmeshop)
```

Environment | History | Connections | Tutorial

Import Dataset ▾ | 13 MiB ▾

R ▾ | Global Environment ▾

Data

▶ acmeshop   40 obs. of 7 variables

Summary output and interpretation:

Min and Max in the Respondent column indicate that we have 40 respondents. We are interested in the distribution of answers for the questions. Most average responses are around 4, indicating a generally neutral stance. However, V4 (getting the best buys) and V6 (comparing prices) show a slight tendency toward agreement, suggesting a practical approach to shopping. In contrast, V5 (I don't care about shopping) tends to lean more toward disagreement, implying that most respondents do care about shopping. Also, higher 3rd quartile for V3 may suggest that combining shopping and dining is a more frequent or habitual behavior among Acme Shopping customers than some of the other beliefs surveyed. No obvious outliers are present.

Console | Terminal ×

R 4.4.1 · ~/Desktop/

```
Mean   :4.55
3rd Qu.:5.25
Max.   :7.00
> summary(acmeshop)
   Respondent          V1             V2             V3             V4             V5
 Min.   : 1.00   Min.   :1.00   Min.   :2.0   Min.   :1.00   Min.   :2.00   Min.   :1.00
 1st Qu.:10.75   1st Qu.:2.00   1st Qu.:3.0   1st Qu.:2.00   1st Qu.:3.00   1st Qu.:2.00
 Median :20.50   Median :4.00   Median :4.0   Median :4.00   Median :4.00   Median :3.50
 Mean   :20.50   Mean   :3.85   Mean   :4.1   Mean   :3.95   Mean   :4.10   Mean   :3.45
 3rd Qu.:30.25   3rd Qu.:5.25   3rd Qu.:5.0   3rd Qu.:6.00   3rd Qu.:5.25   3rd Qu.:4.25
 Max.   :40.00   Max.   :7.00   Max.   :7.0   Max.   :7.00   Max.   :7.00   Max.   :7.00
       V6
 Min.   :2.00
 1st Qu.:3.00
 Median :4.00
 Mean   :4.35
 3rd Qu.:5.25
 Max.   :7.00
>
```

2. Cluster Analysis Distance Computation

Compute the distances between points in the dataset by using the "dist" function. Use the Euclidean method. Next, ask R to compute the hierarchical clusters (hclust), based on the distance matrix you found in the previous step. Type out the text used in the R session and place it in the document to be submitted for the assignment.

```
> acmeshop_df <- acmeshop[, -1]
> acmeshop_dist <- dist(acmeshop_df, method = "euclidean")
> acmeshop_clust <- hclust(acmeshop_dist, method = "ward.D2")
```

3. Cluster Analysis Dendogram Plot

Plot the results of the "hclust" operation into a Dendogram tree diagram. Add boxes around clusters you identified in the Dendogram. Include a screenshot of the dendogram in the document you submit for the assignment.

```
> plot(acmeshop_clust, main = "Dendrogram of acmeshop Respondents", xlab = "Respondents", ylab = "Height")
> rect.hclust(acmeshop_clust, k = 3, border = "red")
```
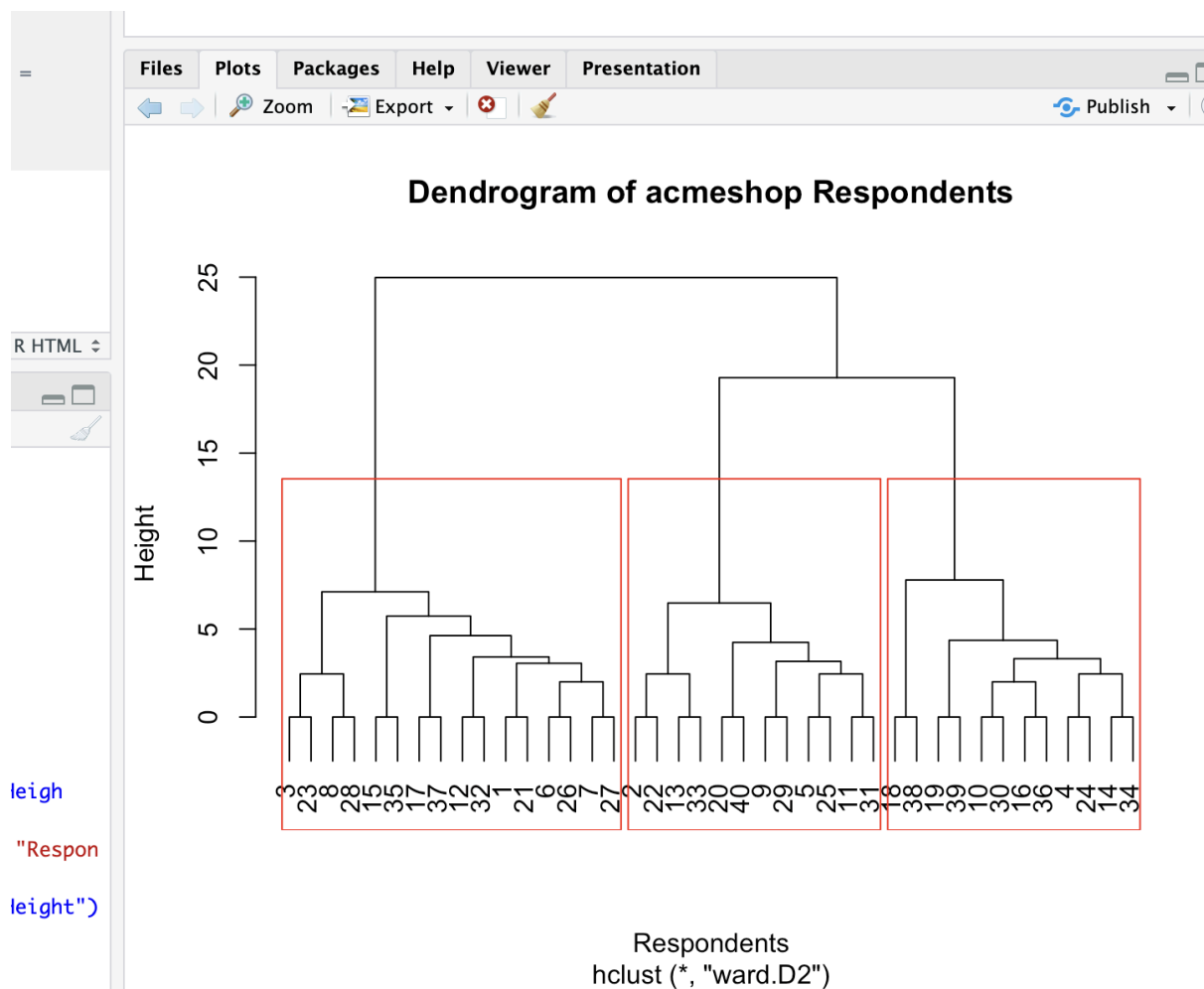
4. CRITICAL THINKING: Cluster Analysis
Provide critical thinking about the results of the cluster analysis. Use the Findings, Conclusions, Recommendations, References format and methodology described in the Instructions.
In particular, discuss expected characteristics for the clusters (segments) you identified in your cluster analysis. What types of behaviors would you expect from the different clusters of shoppers?
Include research: What types of segments are typically found by market research such as this? (Hint: Search for the term "shopping mall customer segments" or similar using an Internet search engine such as Google)
What types of behaviors would you expect from the different clusters of shoppers?

Findings

Using hierarchical cluster analysis in R, we identified three customer segments for Acme Shopping. To better understand how they responded and to potentially create a customer profile for each segment, we calculated the mean responses to the questions for each cluster:

```
> clusters <- cutree(acmeshop_clust, k = k)
> acmeshop$Cluster <- clusters
> cluster_means <- aggregate(acmeshop, by = list(Cluster = acmeshop$Cluster), FUN = mean)
> print(cluster_means)
  Cluster Respondent      V1       V2       V3    V4    V5       V6 Cluster
1       1     18.625 5.750000 3.625000 6.000000 3.125 1.875 3.875000       1
2       2     20.000 1.666667 3.000000 1.833333 3.500 5.500 3.333333       2
3       3     23.500 3.500000 5.833333 3.333333 6.000 3.500 6.000000       3
>
```

Now, let's break down each cluster. Cluster 1 has higher average values for V1 (Shopping is fun), V3 (I combine shopping with eating out), and moderate value for V6 (You can save a lot of money by comparing prices), with a low average for V5 (I don't care about shopping). This suggests that customers in this segment are generally satisfied and enjoy the shopping. In different studies, this is referred to as recreational shopping - "... the other (the recreational shoppers) is looking for fun, leisure, and social interaction" (Ruiz et al., 2003)

Cluster 2 has the highest average value for V5 (I don't care about shopping), compared to Clusters 1 and 3, and moderately agrees with V2 (Shopping is bad for your budget), V4 (I try to get the best buys while shopping), and V6 (You can save a lot of money by comparing prices). It has the lowest values among all clusters in V1 (Shopping is fun) and V3 (I combine shopping with eating out). These might be functional shoppers who are "looking for products/services" (Ruiz et al., 2003). Considering these results, we can suggest that customers from this cluster are less engaged in shopping and possibly shop out of necessity, rather than for enjoyment.

Cluster 3 has the highest values for V4 (I try to get the best buys while shopping), V6 (You can save a lot of money by comparing prices), and V2 (Shopping is bad for your budget). This suggests that they look for the best quality products while also being extremely concerned about price. We can assume that these customers are price-sensitive and possibly interested in deals, discounts, and sales.

Conclusions

We identified three customer segments:
- Recreational shoppers who enjoy shopping and are not highly price conscious.
- Functional shoppers who likely shop out of necessity.
- Price-sensitive customers who are interested in deals, discounts, and sales.

Recommendations

This cluster analysis does not include information about respondents' age, gender or other factors related to shopping behavior. To better understand Acme Shopping customers, we recommend conducting a deeper analysis of each segment. For example, why are Cluster 2 customers less engaged? There may be specific factors or pain points contributing to their disengagement. Addressing these issues could encourage more engagement and shopping, especially since they are not as price-sensitive as Cluster 3.

Based on the findings at this stage, we can assume that Cluster 2 customers are functional shoppers who spend less time shopping and don't enjoy it. They are not regular visitors. For these customers, the shopping mall could consider online shopping with free delivery, targeting their preference for convenience and functionality.

For Cluster 3 customers, the shopping mall could develop special seasonal promotions and discounts.

The most stable and frequent shoppers in Cluster 1 are important. The mall could consider special loyalty programs and events for these customers, as they like the overall idea of shopping experience and enjoy it. Their expected behavior includes regular visits and a prioritization of experience over cost.

References

Guiry, M., Mägi, A. W., & Lutz, R. J. (n.d.). *Defining and measuring recreational shopper identity*. SpringerLink. https://link.springer.com/article/10.1177/0092070305282042

Ishantha, A. (2021, March). *Mall customer segmentation using clustering algorithms.* https://www.researchgate.net/publication/349714847_MALL_CUSTOMER_SEGMENTATION_USING_CLUSTERING_ALGORITHM

Joberri. (2023, July 26). *Mall customers segmentation and clustering*. Medium. https://medium.com/@iwojojo54/mall-customers-segmentation-and-clustering-57b0e5fbdb81

Ruiz, J.-P., Chebat, J.-C., & Hansen, P. (2003). *Another trip to the mall: a segmentation study of customers based on their activities*. Journal of Retailing and Consumer Services.