

R^4H_2O : R for Water Professionals: Session 4

Dr Peter Prevos

Session 4 Program

- ▶ Recap
- ▶ Survey reliability and validity
- ▶ Correlations
- ▶ Clustering
- ▶ Further Study

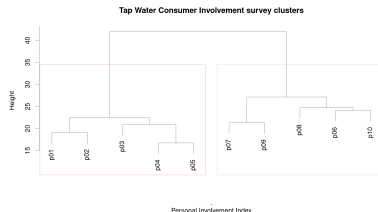


Figure 1: Hierarchical clustering of involvement data.

Recap

- ▶ Read CSV file
- ▶ Filter data
- ▶ Group data
- ▶ Arithmetic
- ▶ Descriptive statistics
- ▶ Visualise data
- ▶ Export to PowerPoint
- ▶ Clean data
- ▶ Join data
- ▶ Pivot data
- ▶ Writing clean code (pipes)

To me, tap water is:

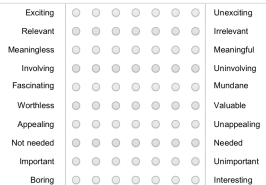


Figure 2: Personal Involvement Index differential semantic scale.

Survey validity & reliability

- ▶ Validity: Am I actually measuring consumer involvement?
 - ▶ Content reliability
 - ▶ Construct reliability
- ▶ Reliability: Is my measurement accurate?

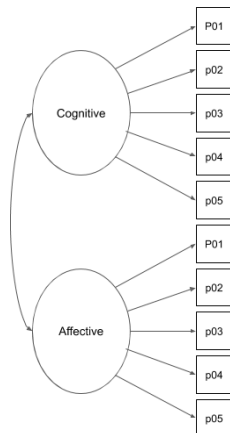


Figure 3: Structural model of Involvement construct.

Hierarchical Clustering Example

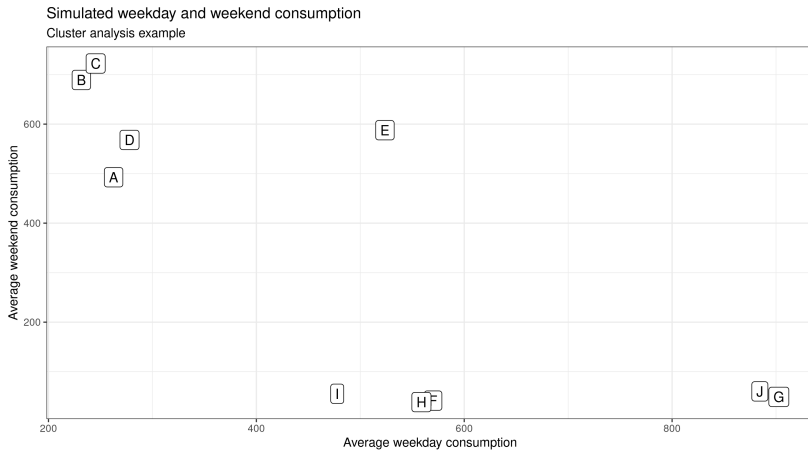


Figure 4: Synthetic customer data.

Hierarchical Clustering Example

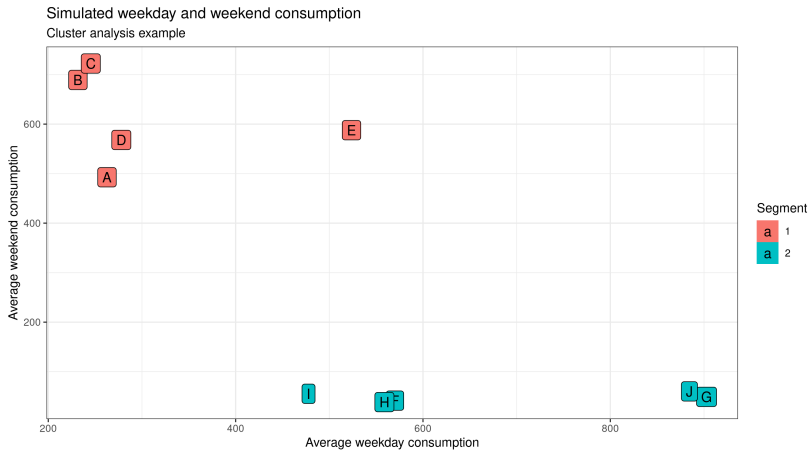


Figure 5: Synthetic customer data.

Hierarchical Clustering Example

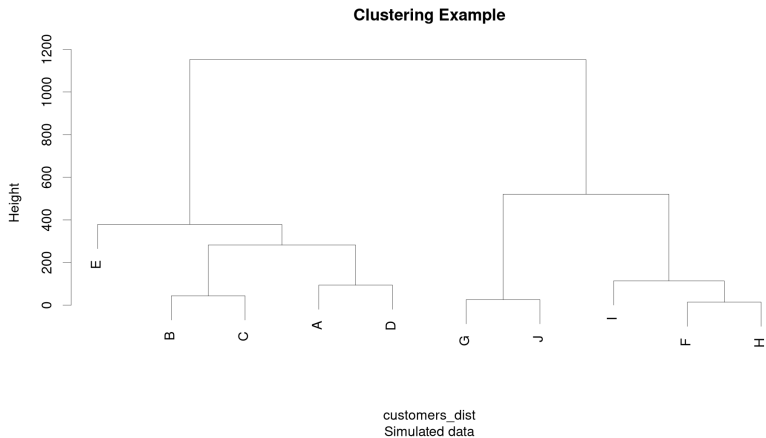


Figure 6: Dendrogram of customer segments.

Assignment

- ▶ Load the survey data and separate:
 - ▶ hardship
 - ▶ contact
 - ▶ Items starting with “f”
- ▶ What is the correlation between experienced hardship and perceived contact frequency?
- ▶ Create a correlation plot of the functional quality items.
- ▶ How many clusters do the functional quality items fall into?

Further Study

- ▶ Help files
- ▶ DataCamp
- ▶ MOOCs
- ▶ Online forums



DataCamp

Figure 7: datacamp.com