Managing Customer Heterogeneity

Dr. Ashutosh Singh

World Ranked — Triple Accredited — Award Winning

EFMD EQUIS ACCREDITED · AACSB ACCREDITED · ASSOCIATION AMBA ACCREDITED

- Framework

- Segmentation, Targeting, and Positioning (STP) Approach

- Analysis tools
  - Cluster analysis for segmentation
  - Factor analysis for developing perceptual maps
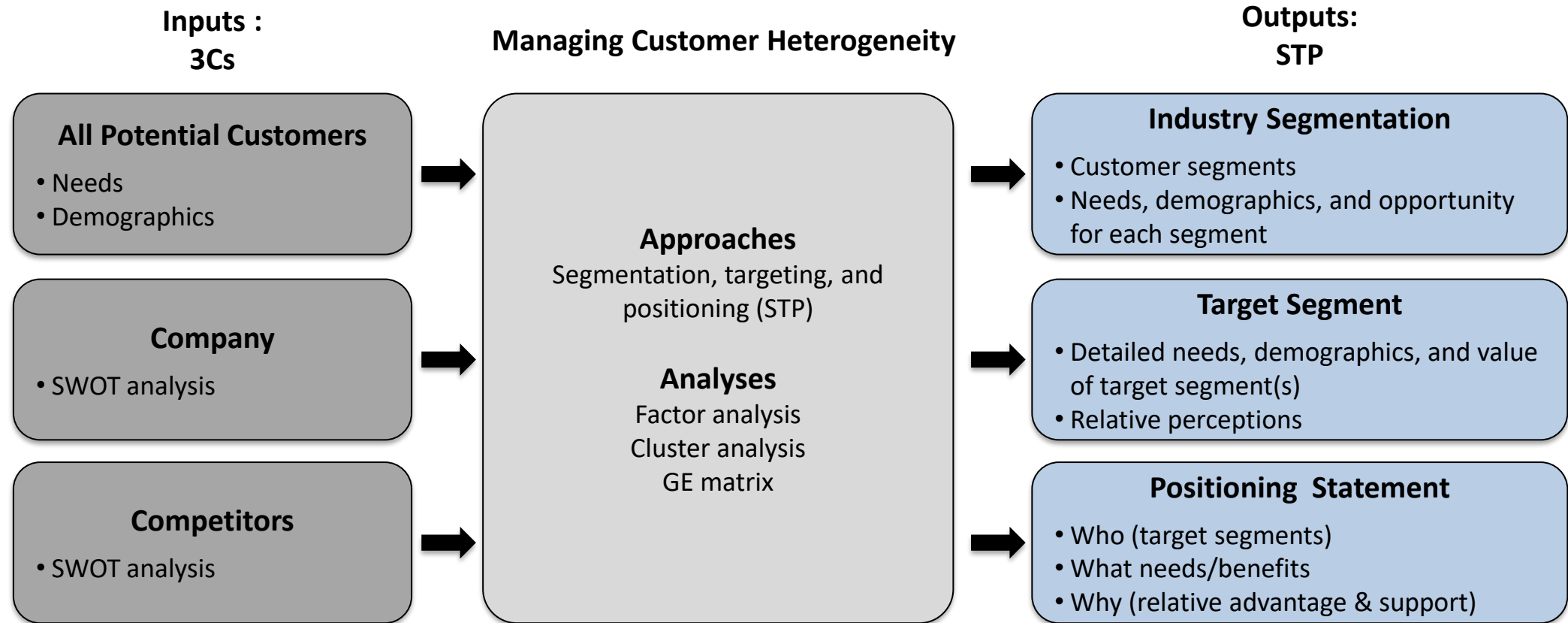
Example

How would you segment the sun care market?
Which segments do you expect to find?

- Concerned Consumers
  - 'a good tan is not important'
  - conscious of the harmful effects of the sun
- Sun Avoiders
  - Avoid sunbathing
  - Unlikely to use sun care
- Conscientious Sun Lovers
  - Adore sunshine
  - Importance of brand trust
- Careless Tanners
  - Adore sunshine (but don't protect)
  - Tanning > protection
- Naïve Beauty Conscious
  - Tan matters
  - Recognize importance of sun protection (but don't understand sun protection factors)

Framework

**Inputs :
3Cs**

Managing Customer Heterogeneity

**Outputs:
STP**

All Potential Customers

- Needs
- Demographics

Company

- SWOT analysis

Competitors

- SWOT analysis

Approaches
Segmentation, targeting, and
positioning (STP)

Analyses
Factor analysis
Cluster analysis
GE matrix

Industry Segmentation

- Customer segments
- Needs, demographics, and opportunity
  for each segment

Target Segment

- Detailed needs, demographics, and value
  of target segment(s)
- Relative perceptions

Positioning Statement

- Who (target segments)
- What needs/benefits
- Why (relative advantage & support)

- Dividing the market into groups of similar customers (slice the pie into pieces) where:
  - Customers within group have very similar needs
  - Customers across groups have different needs
- Descriptors: Observable customer characteristics that help you find and classify customers (e.g., gender, age, income, size, education)
- **Clustering analysis** is the tool to aid segmentation

- Selecting best customer group to sell to (picking the slice to eat)
- A firm needs to select segments to target based on certain selection criteria
  - Market attractiveness (size, growth rate, price sensitivity, etc.)
  - Competitive strength (captures the relative strength of a firm versus competitors at securing and maintaining market share in a given segment)
- **The GE matrix** is one analysis tool designed to helps managers visualize and select target segments

STP: Positioning

- Improve the relative advantage in the minds of the targeted customers

- Nearly everything you do impacts your positioning
  - Places: Samsung dropping Kmart
  - Price: No discounts at Tiffany
  - Promotion: Tiger Woods at Nike and Starbucks
  - Product: Bose and Apple

- **Perceptual maps** is one analysis tool to aid in positioning decisions

Managing Customer Heterogeneity I
## Cluster Analysis for Segmentation

Dr Ashutosh Singh

World Ranked – Triple Accredited – Award Winning

Cluster analysis

- A data-driven partitioning technique that can be used to identify and classify a large set of heterogeneous consumers into a small number of homogeneous segments

- Cluster
  - Set of objects similar to each other and separated from other objects
  - High quality clusters are homogeneous within themselves and heterogeneous between one another
- Cluster analysis
  - Grouping individuals in unknown groups

Intra-cluster distances are minimized

Inter-cluster distances are maximized
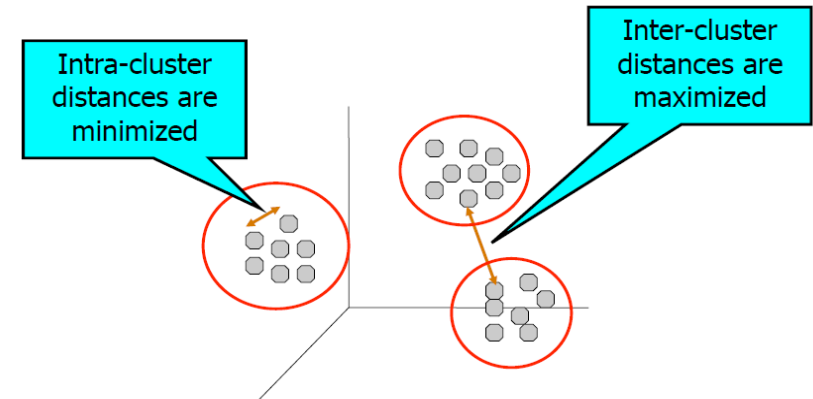
- Steps
  1. Decide on the clustering variables
     - e.g., Speed and acceleration
  2. Select a measure of similarity (for some models)
     - e.g. Euclidean distance to derive the Euclidean distance matrix
  3. Select a clustering method
     - e.g. hierarchical clustering
  4. Describe the clusters

- Decide on the clustering variables
  - Identifying segments based on **demographic** and **behavioral** information is straightforward
  - **Psychographic** information can provide deeper insights
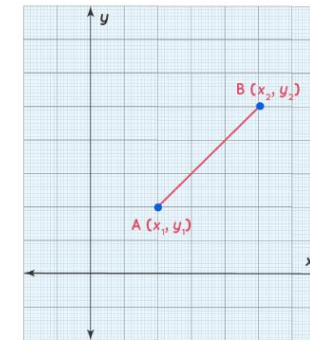 - It is important to identify clusters based on **multiple** variables.

Quantifying similarity

Euclidean Distance
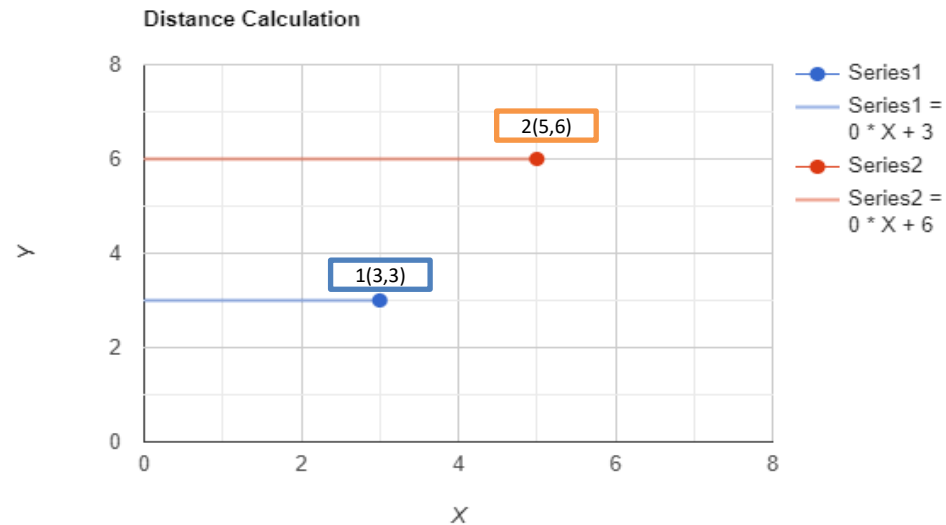
$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

id	Innov	ConstCom
1	3	3
2	5	6
3	7	7
4	4	7

Data matrix

Distance Calculation

- Series1
- Series1 = 0 * X + 3
- Series2
- Series2 = 0 * X + 6

2(5,6)

1(3,3)

Quantifying similarity

Euclidean Distance

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
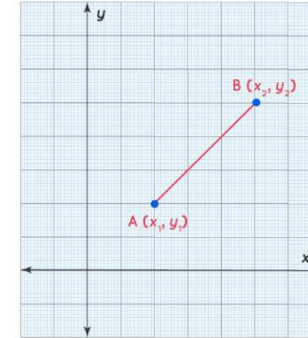
$$d_E(id1, id2) = \sqrt{(3-5)^2 + (3-6)^2} = 3.6$$

id	Innov	ConstCom
1	3	3
2	5	6
3	7	7
4	4	7

Data matrix

	id 1	id 2	id 3	id 4
id 1	0			
id 2	3.6	0		
id 3	5.7	2.2	0	
id 4	4.1	1.4	3	0

Distance matrix

World Ranked – Triple Accredited – Award Winning

Quantifying similarity

- Scaling should be the same across variables
- If this is not the case, you will have to apply **standardization** before quantifying similarity

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$d_E(id1, id2) = \sqrt{(3-5)^2 + (200 - 350)^2} = 150$$

id	Innov	Price
1	3	200
2	5	350
3	7	220
4	4	320

Data matrix

id	id 1	id 2	id 3	id 4
id 1	0			
id 2	150	0		
id 3	20	130	0	
id 4	120	30	100	0

Distance matrix

World Ranked – Triple Accredited – Award Winning

Quantifying similarity

- Standardization $\dfrac{x - \bar{x}}{\hat{\sigma}_x}$

$$\frac{3 - 4.75}{1.71} = -1.02$$

id	Innov	Price
1	3	200
2	5	350
3	7	220
4	4	320

Data matrix

id	Innov_std	Price_std
1	-1.02	-0.98
2	0.15	1.05
3	1.32	-0.71
4	-0.44	0.64

Standardized data matrix

STANDARDIZATION

$$z = \frac{x - \mu}{\sigma}$$

$\sigma = 1$

$\mu = 0$
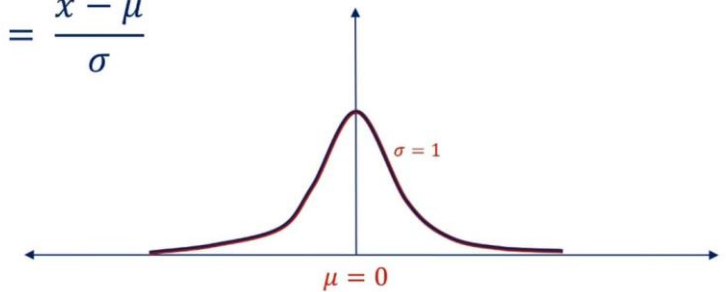
z ~ N(0,1)

Hierarchical clustering

- From N clusters to 1 cluster: From many single-object clusters to 1 cluster
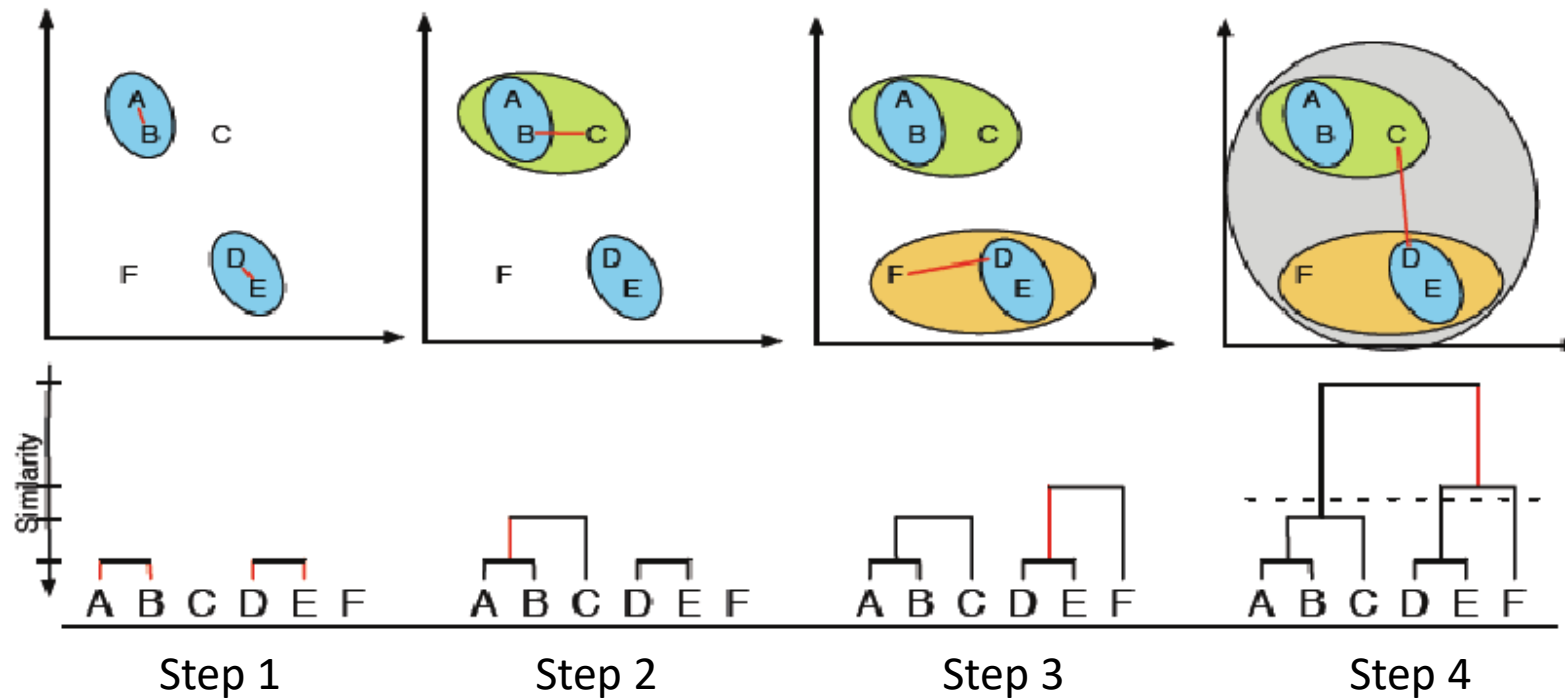
- Linkage criteria:

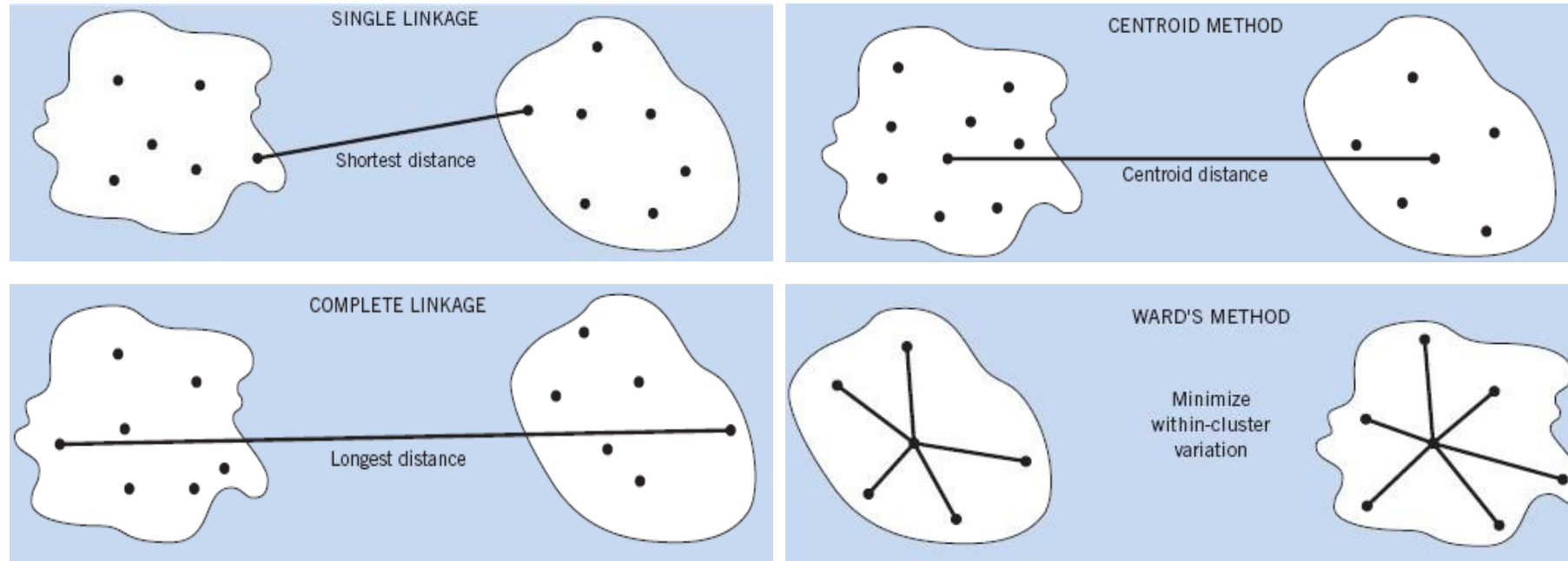# Hierarchical clustering: A complete linkage example

Distance matrix

	id 1	id 2	id 3	id 4
id 1	0			
id 2	3.6	0		
id 3	5.7	2.2	0	
id 4	4.1	1.4	3	0

	id 1	id 2, id 4	id 3
id 1	0		
id 2, id 4	4.1	0	
id 3	5.7	3	0

The criterion to get the distance between the new cluster (id2, id4) and others is based on complete linkage (i.e. the larger one from the original distances with id2 and with id4)

Hierarchical clustering: A complete linkage example

Distance matrix

	id 1	id 2, id 4	id 3
id 1	0		
id 2, id 4	4.1	0	
id 3	5.7	3	0

	id 1	id 2, id 4, id 3
id 1	0	
id 2, id 4, id 3	5.7	0

	id 2, id 4, id 3, id 1
id 2, id 4, id 3, id 1	0

id	Innov	ConstCom
1	3	3
2	5	6
3	7	7
4	4	7

Data matrix

Distance Calculation

- Series1
- Series1 = 0 * X + 3
- Series2
- Series2 = 0 * X + 6

2(5,6)

1(3,3)
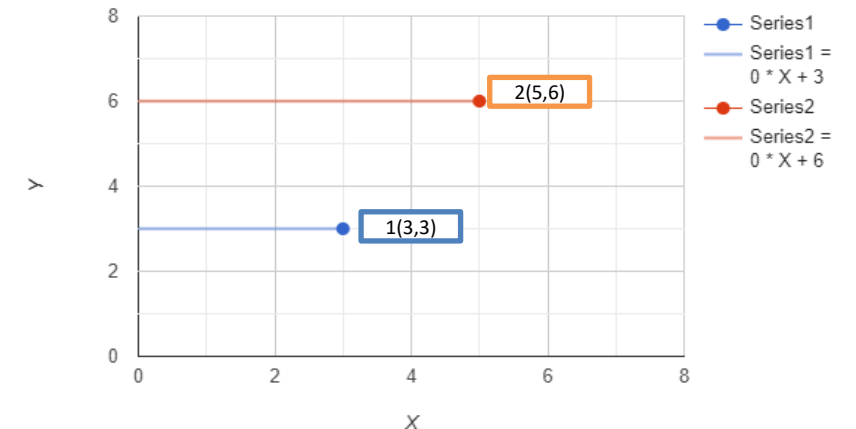
Hierarchical clustering: Dendrogram

Cluster Dendrogram

(2 clusters)

(3 clusters)

Height

distance.matrix
hclust (*, "complete")
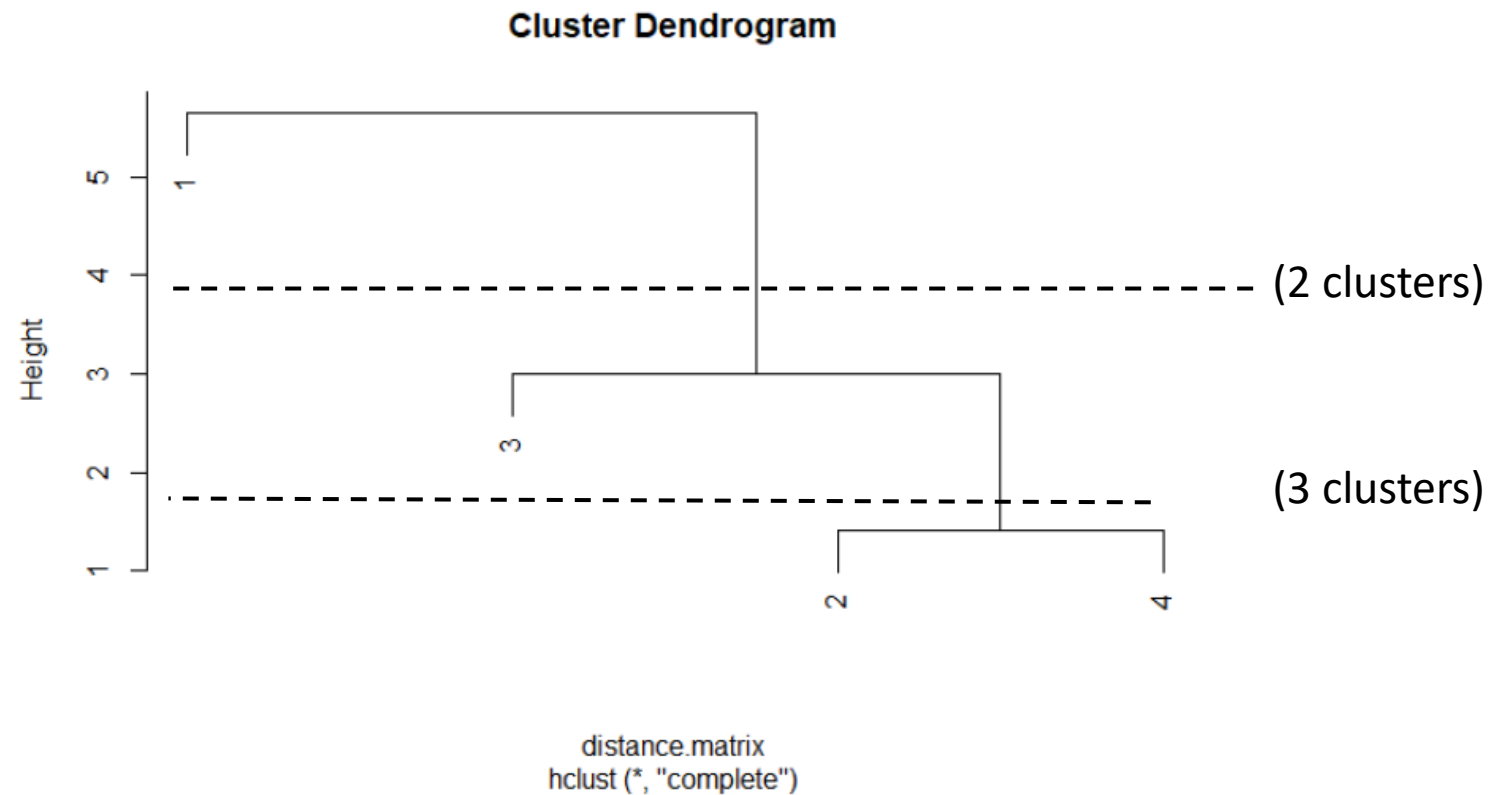
How many clusters?

- Often based on practical concerns and aims of the segmentation

- Heuristic approach
  - Estimate different solutions (e.g. 2, 3, 4... clusters)
  - Compare quality of results for different solutions

- Statistical ways to deal with it

- But in reality it's all about interpretation and about having an actionable segmentation solution

- **Interpretation matters!**
- Single best solution does not exist
- Depends on a company's goal and practical limitations
- 'Optimal' solution requires the researcher's skills **and** market knowledge

- Number of clusters (k) must be specified in advance!

- Initialize k cluster centres
  - Assign each object to the cluster with the closest centroid (Euclidean distance)
  - Mean vectors of obtained clusters taken as new cluster centres
  - Re-compute centroid of each cluster until centroids don't change
  - Minimize sum of squared within-cluster distances

K-means clustering

Model-based clustering
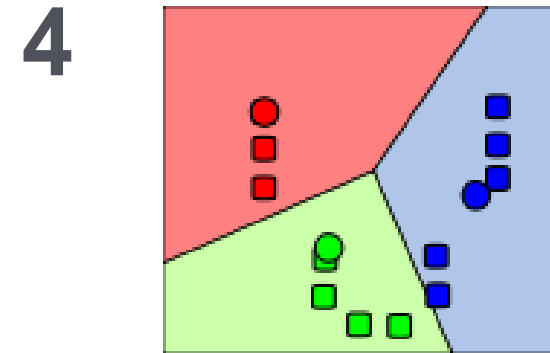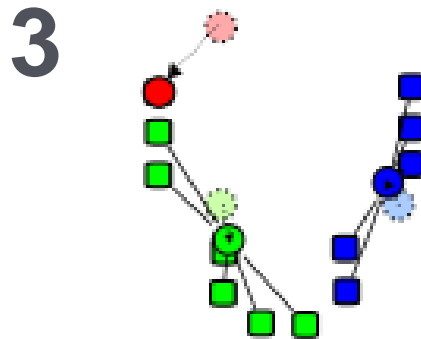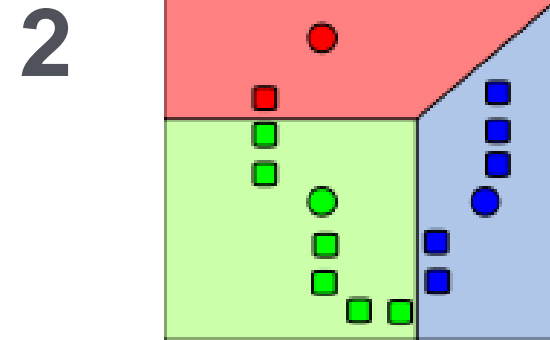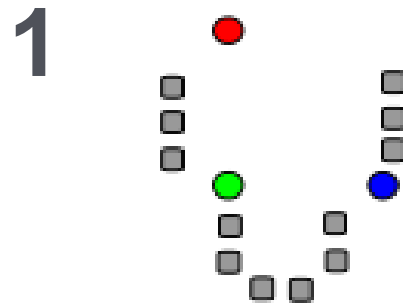
- The key idea is that observations come from unobservable groups with different statistical distributions (such as means and variances)

- The algorithms try to find the best set of such underlying distributions to explain the observed data.

Leeds University Business School

World Ranked – Triple Accredited – Award Winning

EQUIS
ACCREDITED

AACSB
ACCREDITED

AMBA
ACCREDITED

Comparison of methods

Hierarchical methods	Non-hierarchical methods (k-means)	Model-based methods
• "What is the closest thing using distance?" • Any desired number of clusters can be obtained by cutting dendrogram at proper level • At each step they require computation of the full similarity distance matrix	• "What is the closest thing using distance?" • Faster, more reliable • Need to specify the number of clusters • Need to set the initial seeds	• "what is the most similar patterns using probability?" • Instead of using distance, it uses probabilities • More flexible, such as not as sensitive to missing data as traditional ones

Practical tips

- Scaling matters
  - Standardize where necessary

- Seek a stable solution
  - Use different techniques and check results

- Easy to estimate, challenge lays in interpretation
  - Single best solution does usually not exist
  - Usefulness of the segmentation will often dominate optimal number of cluster solution
  - Researcher's skills + market knowledge = Better solution

- Once you have your clustering solution:
  - Describe clusters (attitude, behaviour, demographics)
  - Ideally they should differ as much as possible

- Explain how STP approach is used to manage customer heterogeneity

- Explain how cluster analysis is used to group similar customers together into segments

- Explain the different clustering analysis methods

Workshops

Costco is a subscription-based retailer and offers an individual membership for customers. They have collected data from N=300 potential or existing customers on some demographic variables, including: age, gender, income, number of children, whether they own or rent their homes, and whether they currently subscribe to the offered membership or not. The company is interested the insights from this data on how to segment the market and to target the promising groups.

- Data_Segmentation.csv
- Clustering Analysis:
  - Hierarchical clustering
  - K-means clustering
  - Model-based clustering