Market Segmentation

Marketing Analytics

Professor Kamel Jedidi
Columbia University

Overview

- Segmentation
  - Basics
  - Bases
- Segmentation methods
    - Hierarchical cluster analysis
    - K-means clustering
    - Latent class clustering
- Marketing application
- Summary

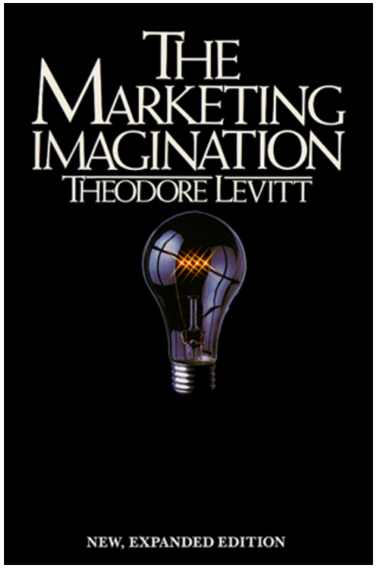
What is Segmentation?

- Market segmentation is the subdividing of a market into distinct subsets of customers.
- Segment members are different between segments but similar within.

Levitt on Segmentation

To think segments means you have to think about what drives customers, customer groups, and the choices that are or might be available to them.

—Levitt, *Marketing Imagination*



Why Segment? Designing a Barbie

- Attributes of a Barbie:
  - Face (light, ivory, tan, brown)
  - Eye color (blue, green, brown)
  - Lip color (rose, cranberry, red)
  - Hair style (short, medium, long)
  - Hair color (blond, red, brown, black)

One Barbie Appeals to None

Attribute	**Customer 1**	**Customer 2**
Face	Light	Brown
Eye Color	Blue	Green
Lip Color	Rose	Cranberry
Hair Style	Long	Short
Hair Color	Red	Black

Benefits of Segmentation

To the Firm …

- Identification of valuable customers
- More targeted promotions & marketing communications
- Higher Customer Lifetime Value (CLV)

Sustainable Profit Growth

To the Customer …

- Customized products & services
- Personalized experience
- Increased customer satisfaction

Customer Loyalty & Retention

Bases for Segmentation

- **Bases** — characteristics that tell us why segments differ (the “why”)
- **Descriptors** — characteristics that help us find and reach segments (the “who”)
- **Behaviors** — metrics that help us measure segments’ behaviors and responsiveness to marketing efforts (the “what”)

Segmentation Bases

Descriptors "Who"	Bases "Why"	Behaviors "What"
Age	**Needs**	**Usage**
Income	**Preferences**	**Loyalty**
Education	**Decision processes**	**Frequency of purchase**
Profession	**Lifestyles**	**Responsiveness to marketing mix**
Media Habits	**Attitudes**	
Gender		

Taxonomy at the Pump: Mobil's Five Types of Gasoline Buyers

Road Warriors: Generally higher income middle-aged men who drive 25,000 to 50,000 miles a year…buy premium with a credit card…purchase sandwiches and drinks from the convenience store…will sometimes wash their cars at the carwash.

True Blues: Usually men and women with moderate to high incomes who are loyal to a brand and sometimes to a particular station..frequently buy premium gasoline and pay in cash

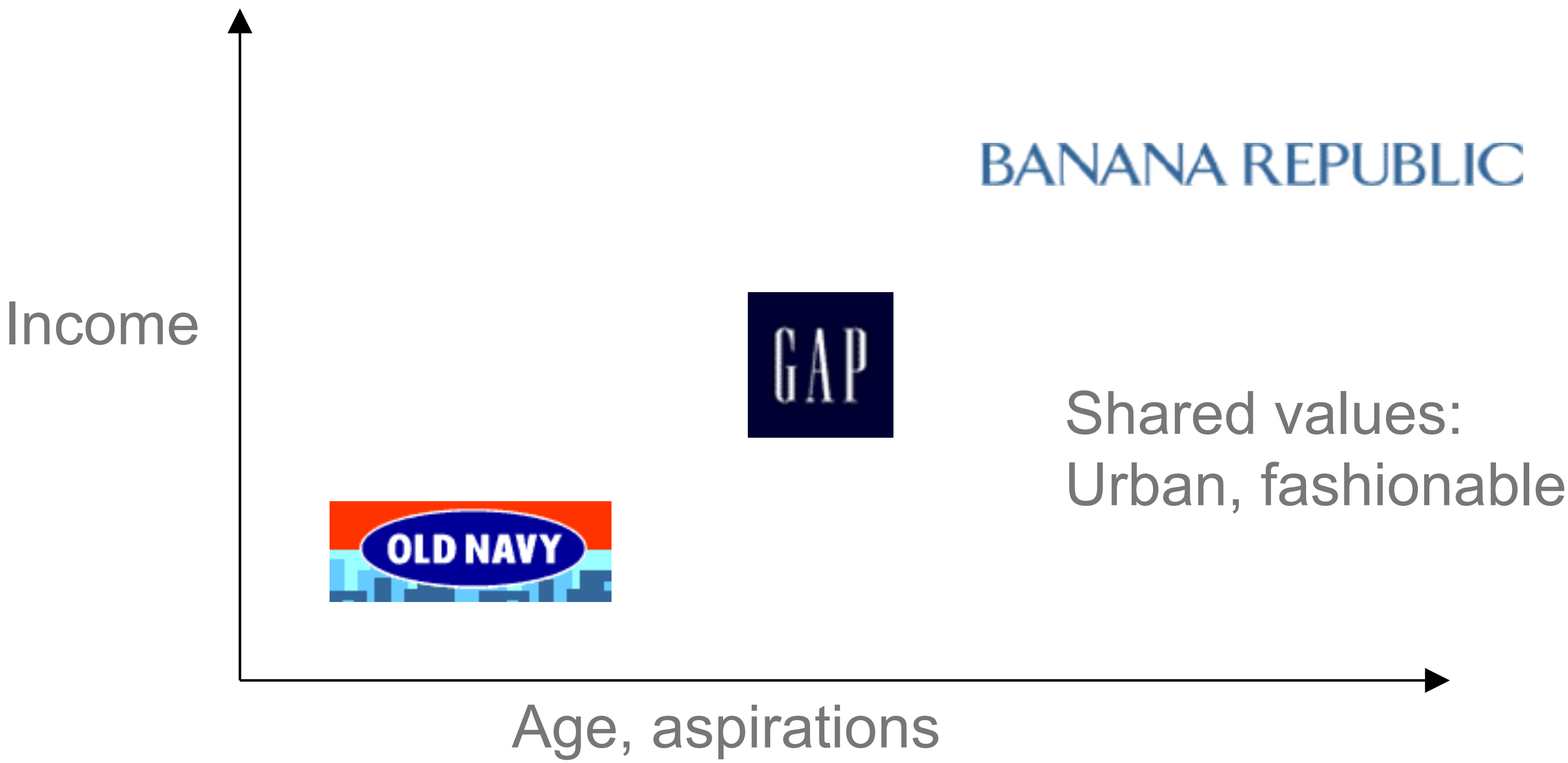
Generation F3: (for fuel, food and fast): Upwardly mobile men and women - half under 25 years of age-who are constantly on the go…drive a lot and snack heavily from the convenience store

Homebodies: Usually housewives who shuttle their children around during the day and use whatever gasoline station is based in town or along their route of travel.

Price Shoppers: Generally aren't loyal to either a brand or a particular station, and rarely buy the premium line…frequently on tight budgets…efforts to woo them have been the base of marketing strategies for years.

Source: Allanna Sullivan, "Mobil Bets Drivers Pick Cappuccino over Low Prices," The Wall Street Journal, January 30, 1995, B1

Demographic Segmentation



Behavioral Segmentation

- Suppose you have a marketing budget to target only ten (out of 20) physicians through detailing.
- Which ten physicians would you target?

Physician	Rx Volume
1	10
2	20
3	30
4	50
5	100
6	90
7	75
8	40
9	45
10	55
11	70
12	15
13	60
14	80
15	120
16	35
17	65
18	75
19	85
20	5

Behavioral Segmentation

Physician	Rx Volume
15	120
5	100
6	90
19	85
14	80
7	75
18	75
11	70
17	65
13	60
10	55
4	50
9	45
8	40
16	35
3	30
2	20
12	15
1	10
20	5

Behavioral Segmentation

- Suppose you have data on physician loyalty to different drugs.
- How would you improve your targeting?

		Brand Share by Physician			
Physician	Rx Volume	Our Brand	B	C	D
15	120	0%	0%	0%	100%
5	100	55%	0%	45%	0%
6	90	93%	0%	7%	0%
19	85	16%	69%	14%	1%
14	80	0%	0%	100%	0%
7	75	22%	5%	73%	0%
18	75	52%	48%	0%	0%
11	70	21%	13%	8%	58%
17	65	1%	99%	0%	0%
13	60	40%	60%	0%	0%
10	55	18%	0%	82%	0%
4	50	96%	0%	0%	3%
9	45	0%	0%	1%	99%
8	40	35%	0%	7%	58%
16	35	56%	17%	20%	7%
3	30	0%	1%	0%	99%
2	20	51%	0%	0%	49%
12	15	0%	0%	0%	100%
1	10	90%	0%	7%	0%
20	5	48%	52%	0%	0%

Behavioral Segmentation

		Brand Share by Physician				
Physician	Rx Volume	Our Brand	B	C	D	Segment
15	120	0%	0%	0%	100%	Lost
5	100	55%	0%	45%	0%	Competitive
6	90	93%	0%	7%	0%	Loyal
19	85	16%	69%	14%	1%	Switchable
14	80	0%	0%	100%	0%	Lost
7	75	25%	5%	70%	0%	Switchable
18	75	52%	48%	0%	0%	Competitive
11	70	21%	13%	8%	58%	Switchable
17	65	1%	99%	0%	0%	Lost
13	60	40%	60%	0%	0%	Switchable
10	55	10%	0%	90%	0%	Lost
4	50	96%	0%	0%	4%	Loyal
9	45	0%	0%	1%	99%	Lost
8	40	35%	0%	7%	58%	Switchable
16	35	56%	17%	20%	7%	Competitive
3	30	0%	1%	0%	99%	Lost
2	20	51%	0%	0%	49%	Competitive
12	15	0%	0%	0%	100%	Lost
1	10	90%	0%	7%	3%	Loyal
20	5	48%	52%	0%	0%	Competitive

Segmenting the Pregnancy Test-Kit Market

- Quidel wants to enter the B2C pregnancy test-kit market.

- How would you segment this market?

Segmenting the Pregnancy Test-Kit Market

Segments

	"Hopeful"	"Fearful"
Brand Name	Conceive	RapidVue
Package	Pink box, Smiling Baby	Brick lettering Mauve background
Shelf Location	Near ovulation testing kit	Near Condoms
Price	\$9.99	\$6.99

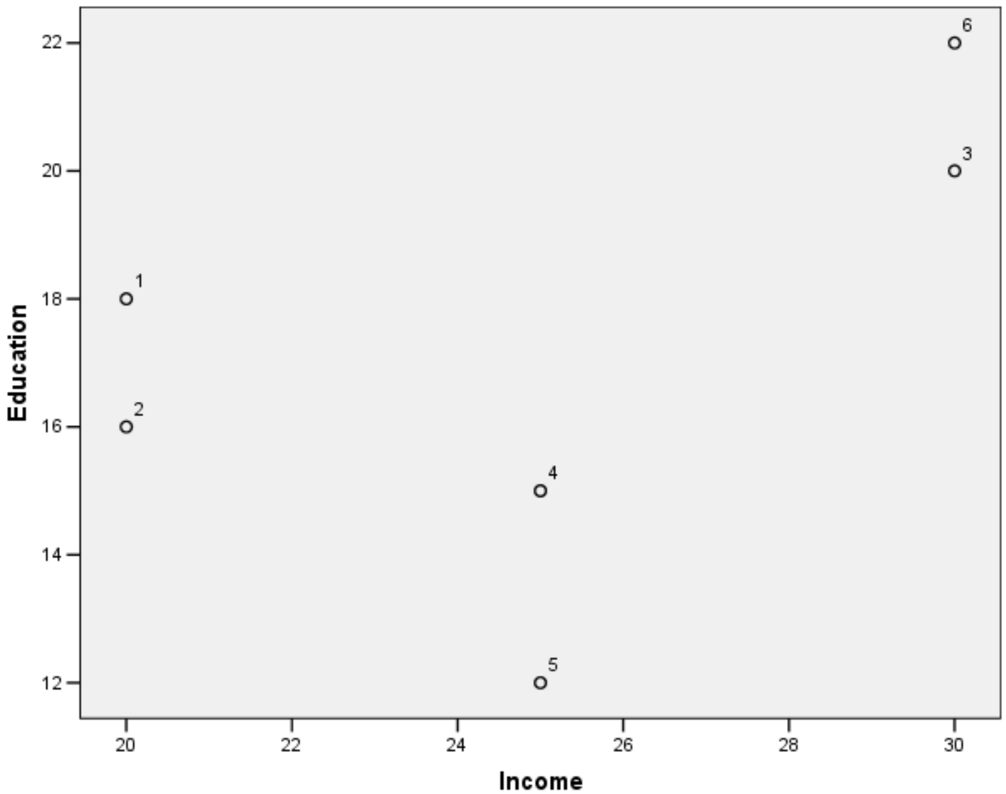
Cluster Analysis

- A class of techniques used to classify objects into groups
  - Objects within a group should be as similar as possible
  - Objects belonging to different groups should be as dissimilar as possible

Example

ID	Income in $K	Education in Yrs
1	20	18
2	20	16
3	30	20
4	25	15
5	25	12
6	30	22

Data Plot



Euclidean Distance

- $D_{12} = \sqrt{(20-20)^2 + (18-16)^2} = 2$

ID	Income in $K	Education in Yrs
1	20	18
2	20	16
3	30	20
4	25	15
5	25	12
6	30	22

The Distance Matrix D

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

Hierarchical Clustering Algorithm

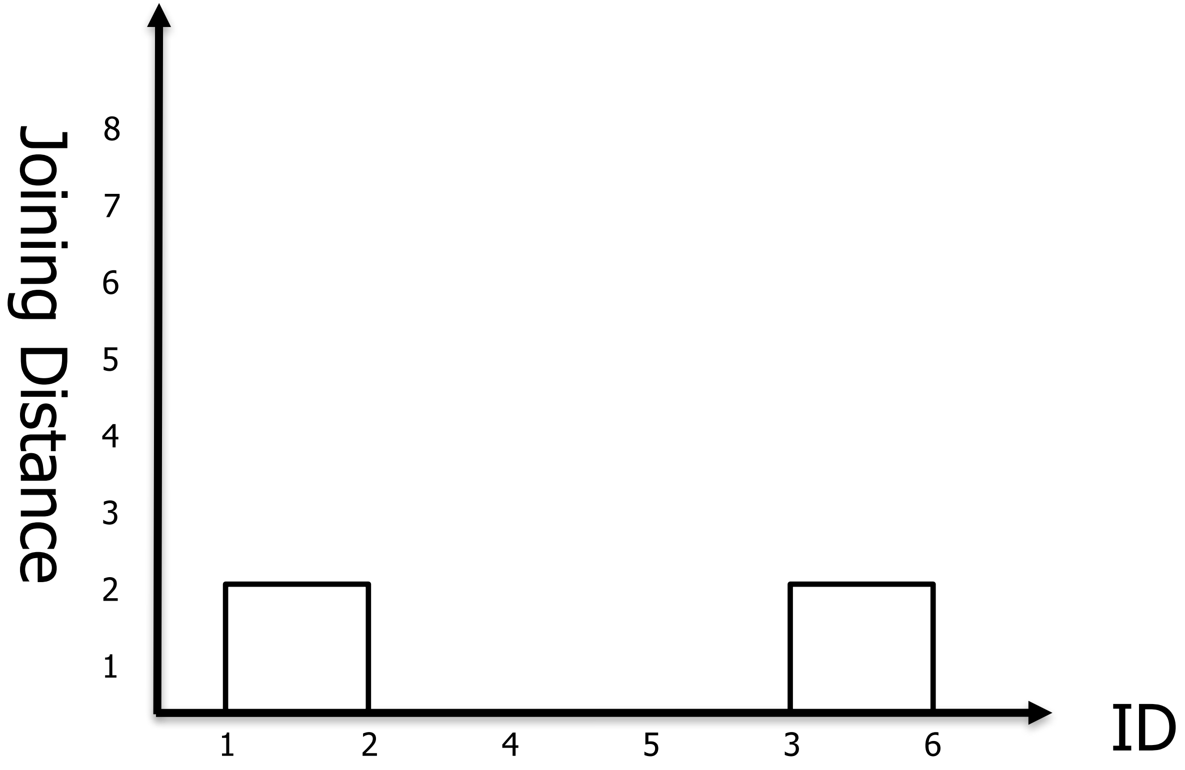
- **STEP 1:** Select Min $\{D_{ij}\}$ and join i and j at that distance

  $D_{12} = 2.0 \rightarrow$ join subjects 1 and 2 in one group (cluster)

  $D_{36} = 2.0 \rightarrow$ join 3 and 6 in another group

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

Dendogram Construction



Algorithm—Continued

- **STEP 2:** Update D using minimum distance rule
  $D_{[ij]k} = \text{Min}\ [D_{ik}, D_{jk}]$
- Example:
  $D_{[12]3} = \text{Min}\ [10.2, 10.8] = 10.2$
  $D_{[12]6} = \text{Min}\ [10.8, 11.6] = 10.8$
  $D_{[12],[3,6]} = \text{Min}\ [10.2, 10.8] = 10.2$

ID	1	2	3	4	5	6
1	0	2.0	10.2	5.8	7.8	10.8
2		0	10.8	5.1	6.4	11.6
3			0	7.1	9.4	2.0
4				0	3.0	8.6
5					0	11.2
6						0

Linkage Rules

- Minimum (single) linkage
- Average linkage
- Maximum (complete) linkage
- Ward (based on minimizing the within-cluster variability of distances)

Updated Distance Matrix

	[1,2]	[3,6]	4	5
[1,2]	0	10.2	5.1	6.4
[3,6]		0	7.1	9.4
4			0	3.0
5				0

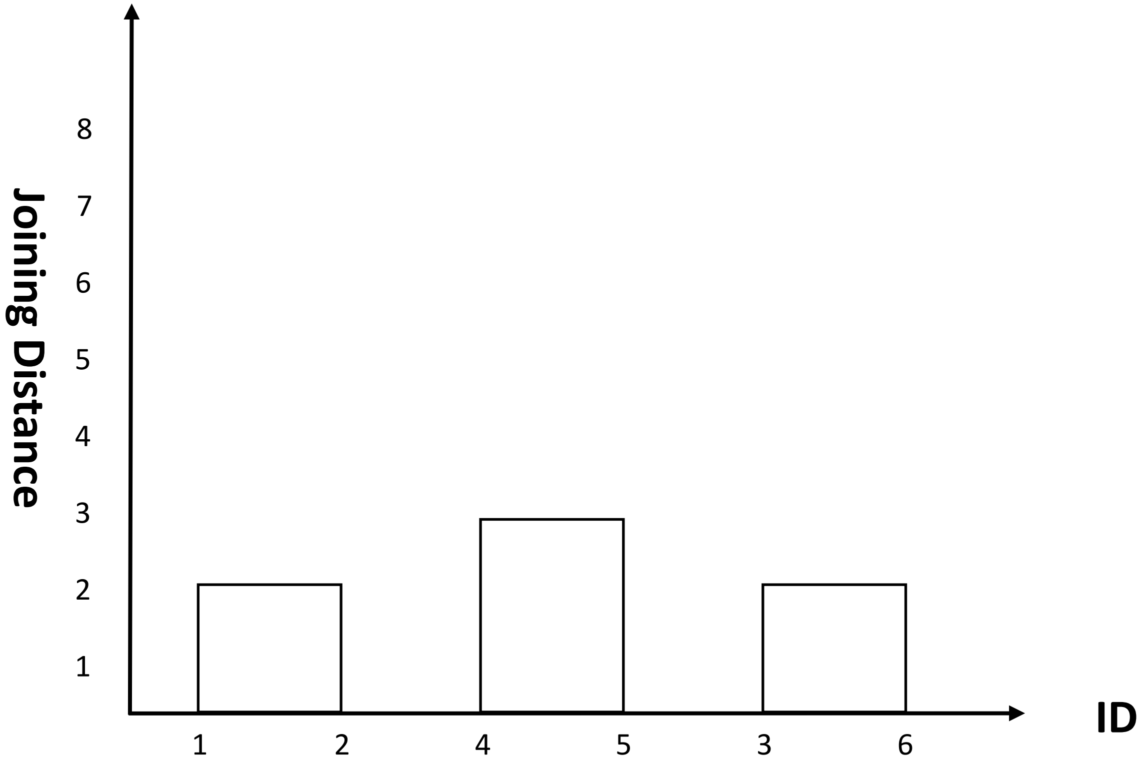
Algorithm—Continued

- **STEP 3:** Pick min distance D_{ij} and join i and j.

$D_{45} = 3.0 \rightarrow$ 4 and 5 are joined

	[1,2]	[3,6]	4	5
[1,2]	0	10.2	5.1	6.4
[3,6]		0	7.1	9.4
4			0	3.0
5				0

Updated Dendogram



Algorithm—Continued

- **STEP 4:** Update the distance matrix as in step 2

	[1,2]	[3,6]	[4,5]
[1,2]	0	10.2	5.1
[3,6]		0	7.1
[4,5]			0

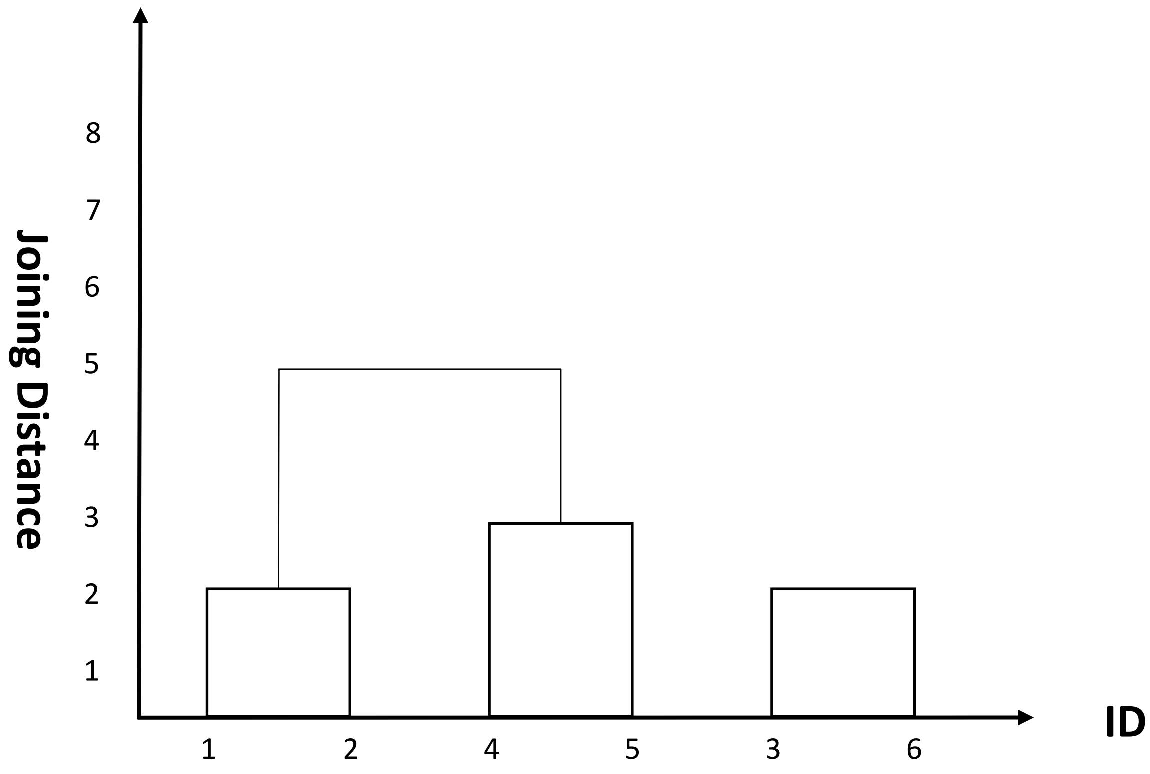
Algorithm

- **STEP 5:** Select min $\{D_{ij}\}$ and join i and j

[1,2] and [4,5] are joined

	[1,2]	[3,6]	[4,5]
[1,2]	0	10.2	5.1
[3,6]		0	7.1
[4,5]			0

Updated Dendogram



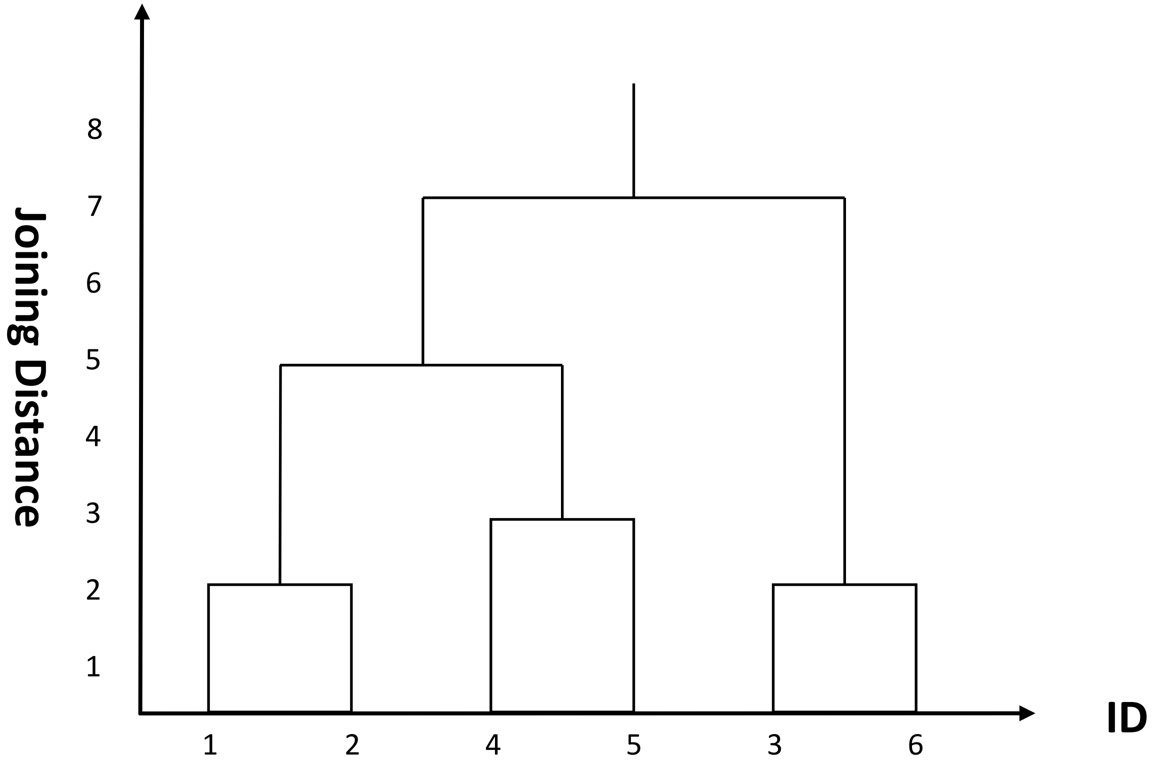
Algorithm—Continued

- **<u>STEP 6:</u>** Update distance matrix

	[1,2,4,5]	[3,6]
[1,2,4,5]	0	7.1
[3,6]		0

- **<u>STEP 7:</u>** Join [1,2,4,5] and [3,6] <u>and</u> Stop

Final Dendogram



Summary
Hierarchical Cluster Analysis is:

- A numerical procedure which attempts to separate a set of observations into groups/clusters
- Members of the same group/cluster are more similar than members of different clusters
- <u>Agglomerative</u> – seeks to join objects sequentially until get one large cluster
- Obtain a tree or “dendogram” representation
- One of the most popular technique used for market segmentation

A Marketing Application

- Attribute importance data collected from 72 students (24 MBAs and 49 undergrad) using the constant-sum method

Attribute	Allocation
Trendy/innovative	
Styling	
Reliability	
Sportiness	
Performance	
Comfort	
Total	100 Points

- Are there different benefit segments? How many segments? How are they different?

Excerpt from the Dataset

```
seg_data <- read.csv(file = "SegmentationData.csv",row.names=1)
head(seg_data)
```

Trendy <int>	Styling <int>	Reliability <int>	Sportiness <int>	Performance <int>	Comfort <int>	MBA <fctr>	Choice <fctr>
10	20	35	5	20	10	MBA	Lexus
25	5	25	5	25	15	MBA	BMW
10	20	30	10	10	20	MBA	Lexus
10	15	30	10	20	15	MBA	BMW
20	10	40	1	14	15	MBA	Mercedes
20	30	10	20	10	10	MBA	Lexus

6 rows

Data Pre-processing

- It is wise to standardize the variables, especially if they are measured with different scales
- To perform hierarchical cluster analysis in R, use hclust()
- hclust() requires the distance matrix as data input

Data Pre-processing

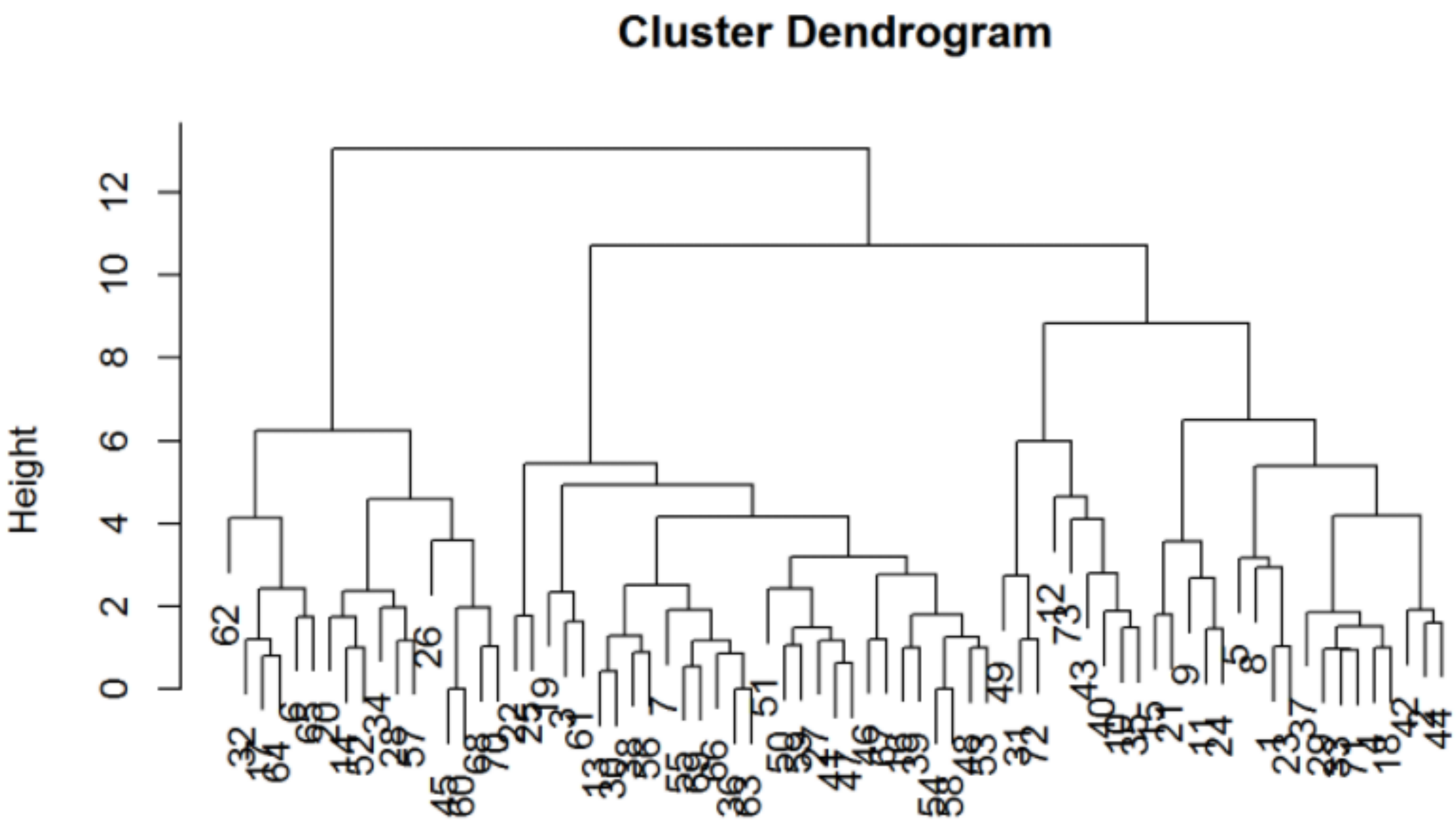
```
std_seg_data <- scale(seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance", "Comfort")])
dist <- dist(std_seg_data, method = "euclidean")
as.matrix(dist)[1:5, 1:5]
```

Euclidean distances between the first five respondents

```
##          1        2        3        4        5
## 1 0.000000 3.730216 2.802191 1.775616 2.746615
## 2 3.730216 0.000000 4.218662 3.017462 2.984534
## 3 2.802191 4.218662 0.000000 1.974683 3.331082
## 4 1.775616 3.017462 1.974683 0.000000 2.924141
## 5 2.746615 2.984534 3.331082 2.924141 0.000000
```

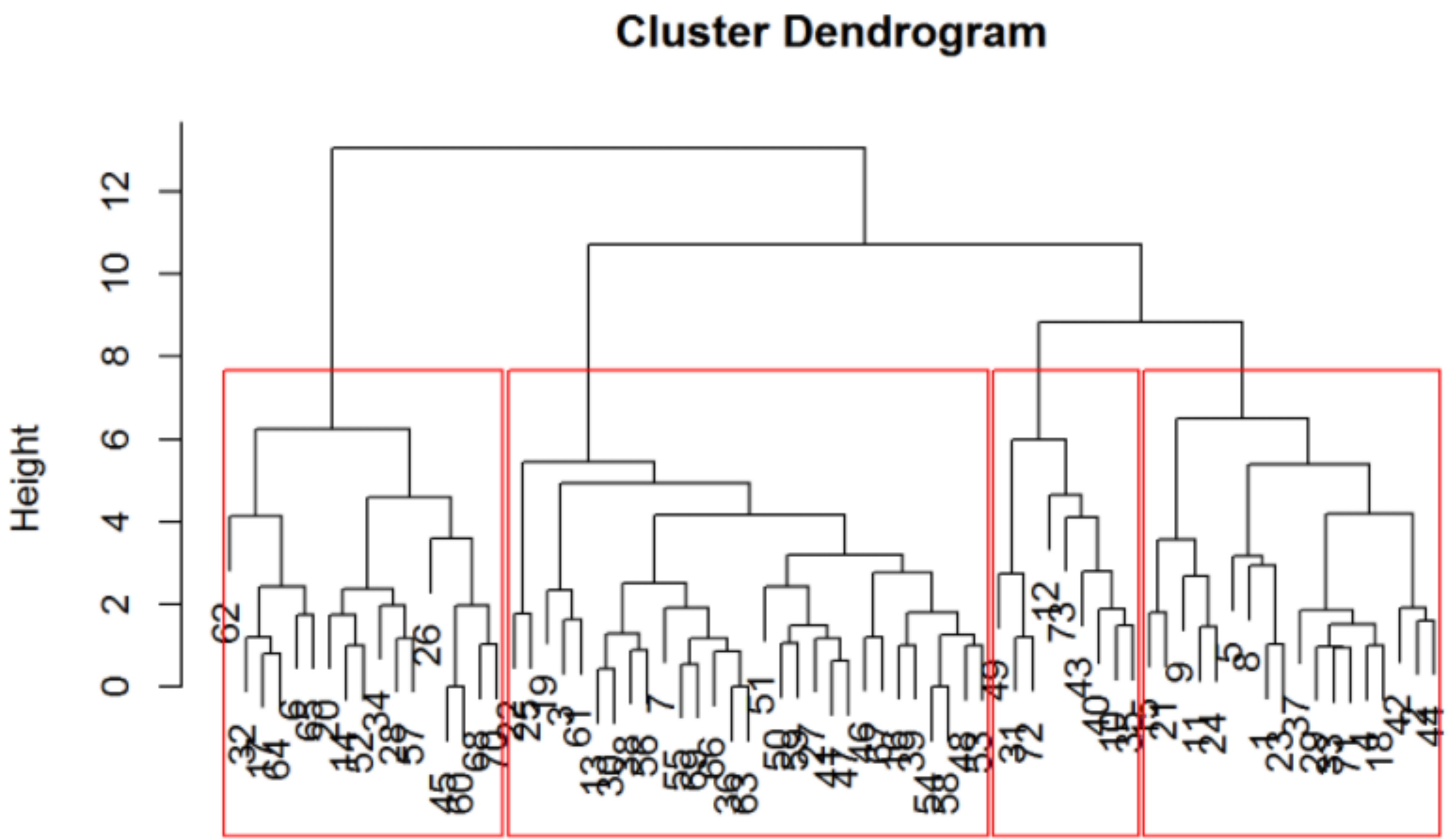
hclust() output

```
clust <- hclust(dist, method = "ward.D2")
plot(clust)
```



Four Clusters?

```r
clust <- hclust(dist, method = "ward.D2")
plot(clust)
h_cluster <- cutree(clust, 4)
rect.hclust(clust, k=4, border="red")
```



The Four-Cluster Solution

```
table(h_cluster)
```

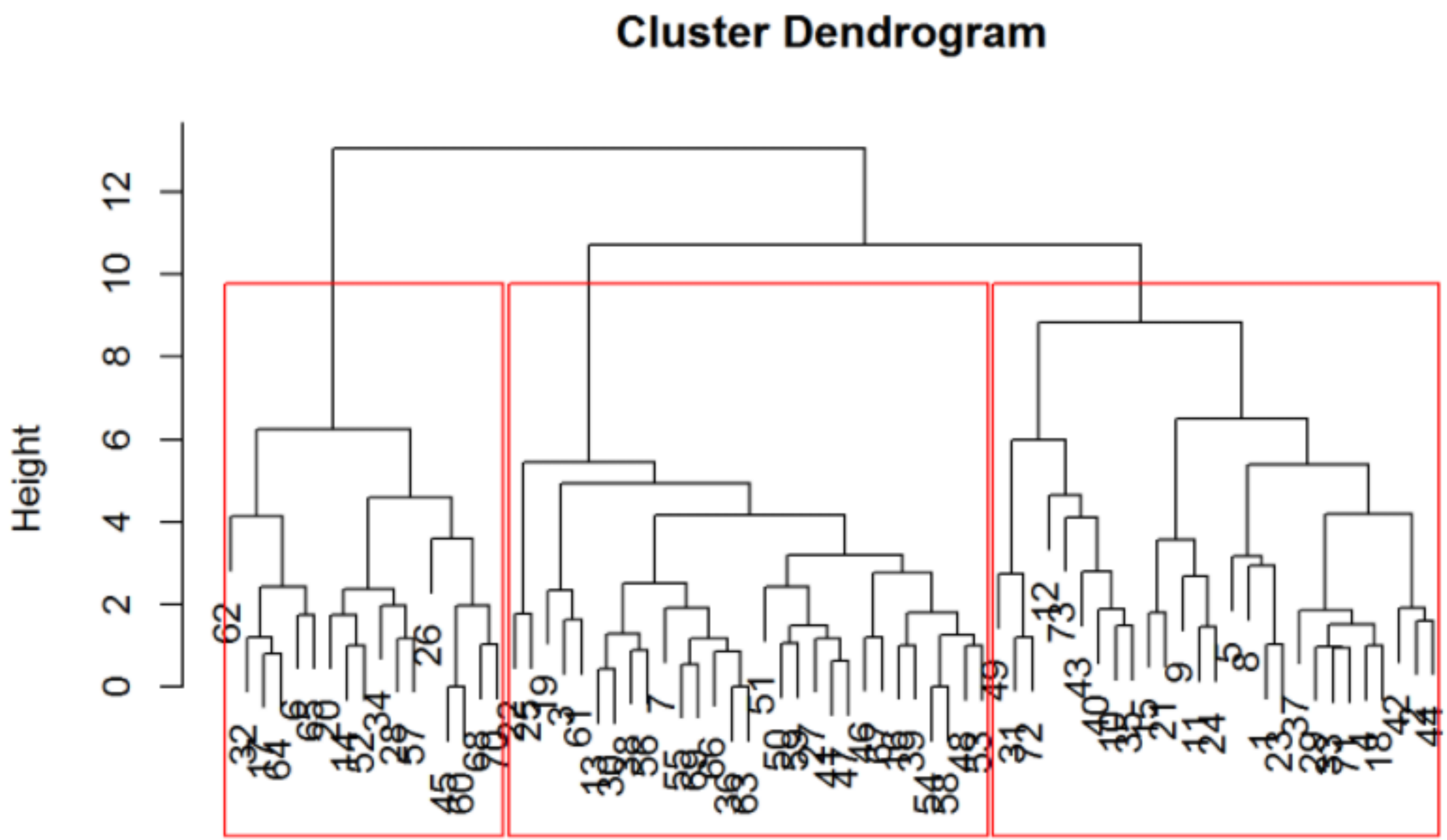
```
## h_cluster
##  1  2  3  4
## 18 29 17  9
%  25 40 23 12
```

```
hclust_summary <- aggregate(std_seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance",
"Comfort")],by=list(h_cluster),FUN=mean)
hclust_summary
```

Group.1 <int>	Trendy <dbl>	Styling <dbl>	Reliability <dbl>	Sportiness <dbl>	Performance <dbl>	Comfort <dbl>
1	-0.50357227	-0.6837159	1.09976574	-0.94569654	0.6548024	0.08642535
2	-0.01577854	-0.4249072	-0.28158545	0.50052114	-0.0989237	0.58621035
3	1.14725137	0.8552172	-0.65660558	0.16346240	-0.9192806	-0.69794311
4	-1.10904387	1.1211667	-0.05194561	-0.03015957	0.7455682	-0.74341374

Three Clusters?

```r
clust <- hclust(dist, method = "ward.D2")
plot(clust)
h_cluster <- cutree(clust, 3)
rect.hclust(clust, k=3, border="red")
```



The Three-Cluster Solution

```
table(h_cluster)
```

```
## h_cluster
##  1  2  3
## 27 29 17
 %  37 40 23
```

```
hclust_summary <- aggregate(std_seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance",
"Comfort")],by=list(h_cluster),FUN=mean)
hclust_summary
```

Group.1 <int>	Trendy <dbl>	Styling <dbl>	Reliability <dbl>	Sportiness <dbl>	Performance <dbl>	Comfort <dbl>
1	-0.70539614	-0.08208834	0.7158620	-0.6405175	0.6850577	-0.1901877
2	-0.01577854	-0.42490717	-0.2815854	0.5005211	-0.0989237	0.5862104
3	1.14725137	0.85521724	-0.6566056	0.1634624	-0.9192806	-0.6979431

The Three-Cluster Solution

```r
table(h_cluster)
```

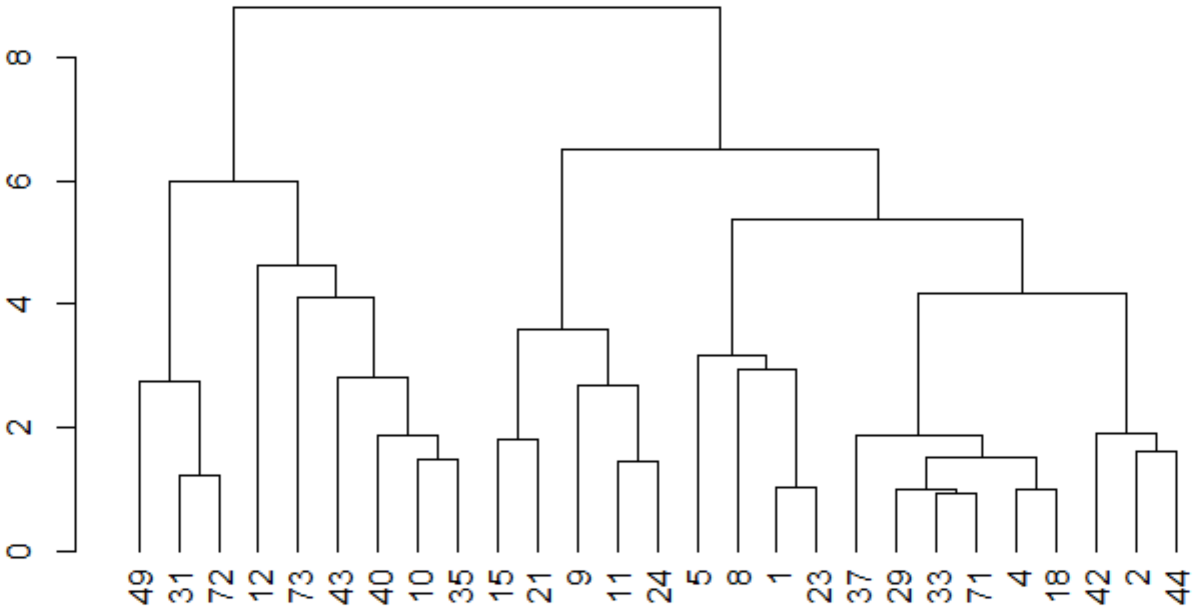
```
## h_cluster
##   1  2  3
## 27 29 17
 % 37 40 23
```

```r
hclust_summary <- aggregate(std_seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance",
"Comfort")],by=list(h_cluster),FUN=mean)
hclust_summary
```

Group.1 <int>	Trendy <dbl>	Styling <dbl>	Reliability <dbl>	Sportiness <dbl>	Performance <dbl>	Comfort <dbl>
Performance	-0.70539614	-0.08208834	0.7158620	-0.6405175	0.6850577	-0.1901877
Comfort	-0.01577854	-0.42490717	-0.2815854	0.5005211	-0.0989237	0.5862104
Appearance	1.14725137	0.85521724	-0.6566056	0.1634624	-0.9192806	-0.6979431

Cluster 3: Appearance-Driven Segment

```
plot(cut(as.dendrogram(clust), h=9)$lower[[3]])
```



NbClust() in R Uses 26 Criteria to Determine the Number of Clusters

```
NbClust(data=std_seg_data[,1:5], min.nc=3, max.nc=15, index="all", method="ward.D2")
```

```
## $Best.nc
##                     KL      CH Hartigan     CCC   Scott  Marriot   TrCovW
## Number_clusters 3.0000  3.0000   6.0000 15.0000  5.0000        9   4.0000
## Value_Index     4.2356 19.1708   1.5295 -2.8227 43.5959 32900086 821.1635
##                 TraceW Friedman   Rubin Cindex      DB Silhouette   Duda
## Number_clusters 6.0000  13.0000 12.0000  5.000 12.0000      9.000 3.0000
## Value_Index     5.1859   2.7154 -0.0482  0.356  1.0742      0.255 0.6935
##                 PseudoT2  Beale Ratkowsky    Ball PtBiserial Frey McClain
## Number_clusters   3.0000 3.0000    3.0000  4.0000     9.0000    2   3.000
## Value_Index       6.6287 1.2967    0.3432 26.1091     0.5828   NA   1.236
##                  Dunn Hubert SDindex Dindex    SDbw
## Number_clusters 12.00      0  9.0000      0 15.0000
## Value_Index      0.23      0  1.7508      0  0.3038
```

```
* According to the majority rule, the best number of clusters is  3
```

Are the Segments Identifiable?

```
CrossTable(seg_data$MBA,h_cluster,prop.chisq = FALSE, prop.r = T, prop.c = T,
          prop.t = F,chisq = T)
```

```
             | h_cluster
seg_data$MBA |       Perf. |     Comfort |  Appearance |   Row Total |
-------------|-------------|-------------|-------------|-------------|
         MBA |          14 |           6 |           4 |          24 |
             |       0.583 |       0.250 |       0.167 |       0.329 |
             |       0.519 |       0.207 |       0.235 |             |
-------------|-------------|-------------|-------------|-------------|
   Undergrad |          13 |          23 |          13 |          49 |
             |       0.265 |       0.469 |       0.265 |       0.671 |
             |       0.481 |       0.793 |       0.765 |             |
-------------|-------------|-------------|-------------|-------------|
Column Total |          27 |          29 |          17 |          73 |
             |       0.370 |       0.397 |       0.233 |             |
-------------|-------------|-------------|-------------|-------------|


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  7.03013     d.f. =  2     p =  0.02974588
```

Are the Segments Meaningful?

```
CrossTable(h_cluster,seg_data$Choice,prop.chisq = FALSE, prop.r = T, prop.c = T,
          prop.t = F,chisq = T)
```

```
             | seg_data$Choice
   h_cluster |       BMW |     Lexus |  Mercedes | Row Total |
-------------|-----------|-----------|-----------|-----------|
       Perf. |        14 |         9 |         4 |        27 |
             |     0.519 |     0.333 |     0.148 |     0.370 |
             |     0.438 |     0.409 |     0.211 |           |
-------------|-----------|-----------|-----------|-----------|
     Comfort |        10 |         8 |        11 |        29 |
             |     0.345 |     0.276 |     0.379 |     0.397 |
             |     0.312 |     0.364 |     0.579 |           |
-------------|-----------|-----------|-----------|-----------|
  Appearance |         8 |         5 |         4 |        17 |
             |     0.471 |     0.294 |     0.235 |     0.233 |
             |     0.250 |     0.227 |     0.211 |           |
-------------|-----------|-----------|-----------|-----------|
Column Total |        32 |        22 |        19 |        73 |
             |     0.438 |     0.301 |     0.260 |           |
-------------|-----------|-----------|-----------|-----------|

Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  4.095664     d.f. =  4     p =  0.393214
```

K-Means Clustering

- K-means requires the specification of the number of clusters in advance, say S=3.
- K-Means algorithm:
  1. Start by randomly assigning each subject to a cluster, s=1,…,S
  2. Compute the centroid of each cluster and the distance of each subject to each of the clusters centroids
  3. Reassign each subject to the cluster with closest centroid
  4. Repeat steps 2 and 3 until no further reassignment is possible (i.e., when the within-cluster variance is minimized)

K-Means Clustering

```
car_Cluster3 <-kmeans(std_seg_data, 3, iter.max=100,nstart=100)
car_Cluster3
```

```
K-means clustering with 3 clusters of sizes 18, 32, 23

Cluster means:
        Trendy    Styling Reliability Sportiness Performance     Comfort
1 -0.637247817 -0.6837159   1.1781135 -1.0328905   0.7785740  0.08642535
2 -0.003271873 -0.3788069  -0.3496669  0.4977728  -0.0445069  0.53615835
3  0.503267855  1.0621176  -0.4355087  0.1157956  -0.5473961 -0.81359668

Clustering vector:
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
 1  1  2  1  1  3  2  1  1  3  1  2  2  3  1  2  3  2  2  3  1  2  1  1  2
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
 3  2  3  1  2  3  3  1  3  3  2  1  2  2  3  2  1  3  1  2  1  2  2  3  2
51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
 2  3  2  2  2  2  3  2  2  2  2  3  2  3  3  2  2  3  3  3  2  3  1

Within cluster sum of squares by cluster:
[1]  81.39207  83.90060 111.49649
 (between_SS / total_SS =  35.9 %)
```

NbClust() in R Uses 26 Criteria to Determine the Number of Clusters

```
NbClust(data=std_seg_data[,1:5], min.nc=3, max.nc=15, index="all", method="kmeans")
```

```
$Best.nc
                      KL      CH Hartigan     CCC   Scott   Marriot
Number_clusters   8.0000  3.0000   6.0000  3.0000  7.0000         4
Value_Index     115.6198 23.0027  11.7591 -2.3384 47.2286 154280514
                  TrCovW  TraceW Friedman   Rubin  Cindex      DB
Number_clusters   7.0000  7.0000  12.0000 11.0000 10.0000 15.0000
Value_Index     723.4098 16.8164   5.0156 -0.1911  0.3716  1.0293
                Silhouette   Duda PseudoT2  Beale Ratkowsky    Ball
Number_clusters     8.0000 3.0000   3.0000 3.0000    3.0000  4.0000
Value_Index         0.2537 0.7619   6.8769 0.9339    0.3612 24.8902
                PtBiserial Frey McClain   Dunn Hubert SDindex Dindex
Number_clusters     8.0000    2  3.0000 8.0000      0  8.0000      0
Value_Index         0.5323   NA  1.2726 0.2019      0  1.8598      0
                   SDbw
Number_clusters 15.0000
Value_Index      0.2517
```

```
According to the majority rule, the best number of clusters is  3
```

Concordance between kmeans() and hclust() cluster memberships

```
CrossTable(h_cluster,Kmean_Cluster,prop.chisq = FALSE, prop.r = T, prop.c = T,
          prop.t = F,chisq = T)
```

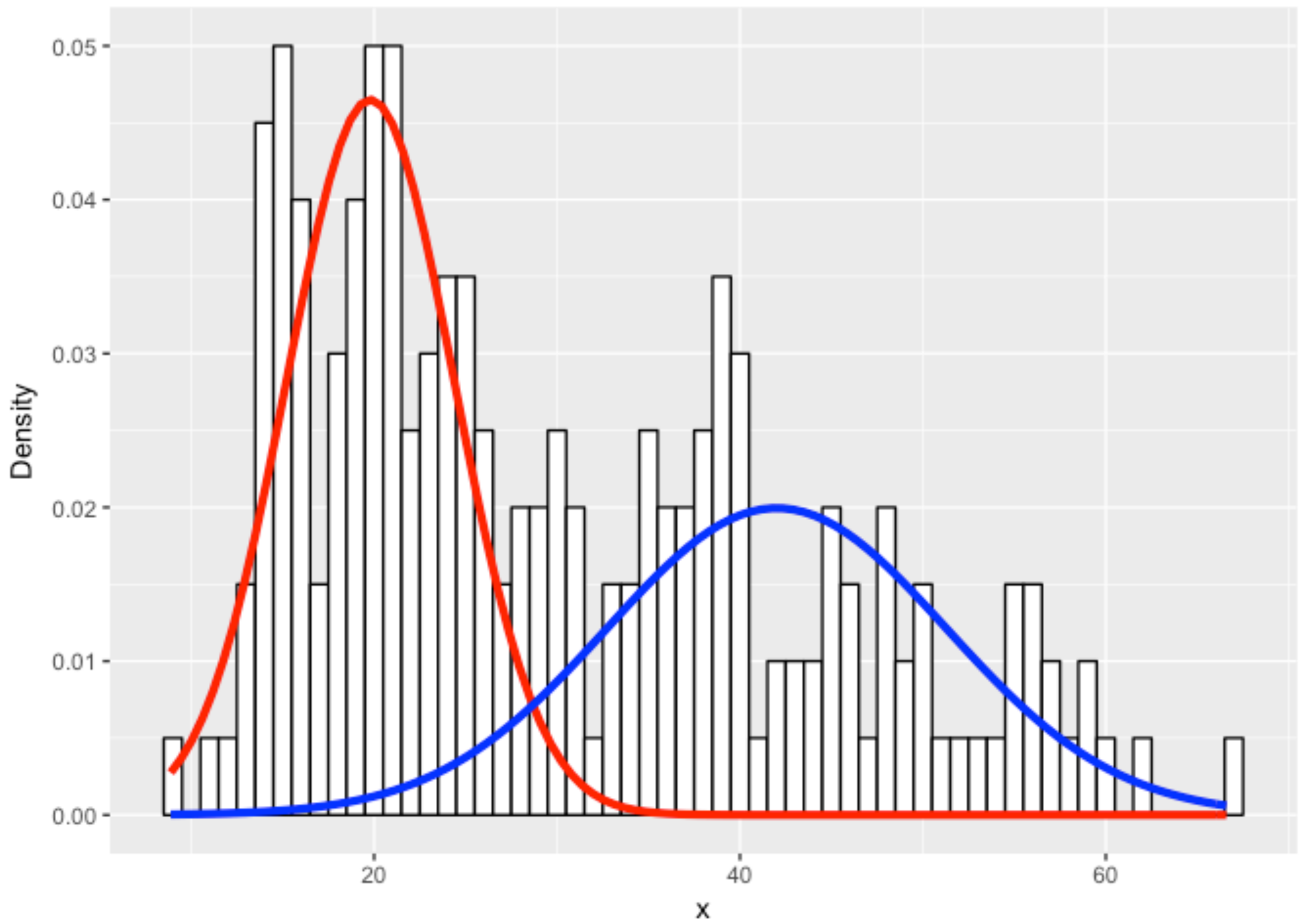
```
             | Kmean_Cluster
   h_cluster |       Perf. KM |     Comfort KM |  Appearance KM |      Row Total |
-------------|----------------|----------------|----------------|----------------|
       Perf. |             17 |              3 |              7 |             27 |
             |          0.630 |          0.111 |          0.259 |          0.370 |
             |          0.944 |          0.094 |          0.304 |                |
-------------|----------------|----------------|----------------|----------------|
     Comfort |              1 |             27 |              1 |             29 |
             |          0.034 |          0.931 |          0.034 |          0.397 |
             |          0.056 |          0.844 |          0.043 |                |
-------------|----------------|----------------|----------------|----------------|
  Appearance |              0 |              2 |             15 |             17 |
             |          0.000 |          0.118 |          0.882 |          0.233 |
             |          0.000 |          0.062 |          0.652 |                |
-------------|----------------|----------------|----------------|----------------|
Column Total |             18 |             32 |             23 |             73 |
             |          0.247 |          0.438 |          0.315 |                |
-------------|----------------|----------------|----------------|----------------|


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  77.06958     d.f. =  4     p =  7.26997e-16
```

Hit Rate:
(17+27+15)/73
= 81%

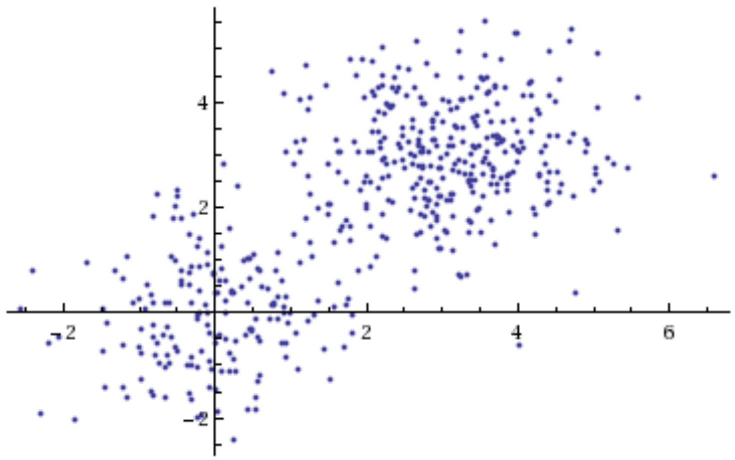
Latent Class Analysis—Mclust()

- Uses a statistical model (vs. numerical algorithm) to form clusters
- Assumes that data follow a finite mixture of normal distributions
- Estimates a family of models and selects the best based on the Bayesian Information Criterion (BIC)
- Outputs cluster means and cluster membership for each subject
- Requires a large sample size

Finite Mixture of Two Univariate Normals



Finite Mixture of Two Bivariate Normals



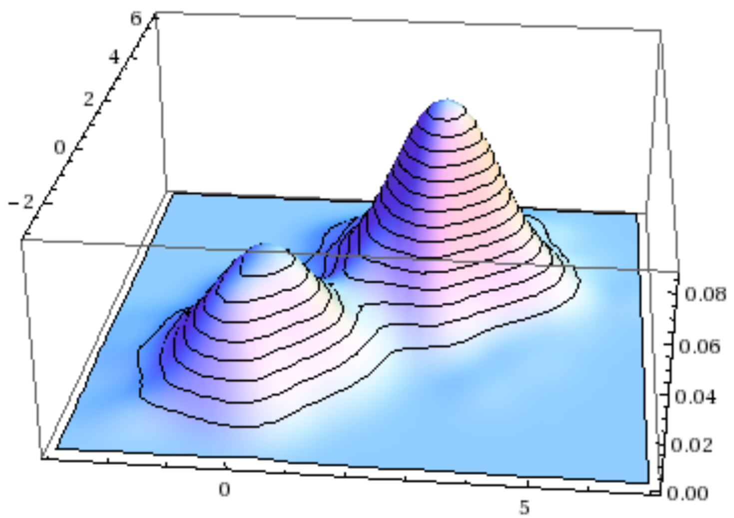


Figure is generated by Mathematica.

Latent Class Analysis

```
lca_clust <- Mclust(std_seg_data[,1:5],verbose = FALSE,modelNames="VEE")
summary(lca_clust)
```

```
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust VEE (ellipsoidal, equal shape and orientation) model with 2 components:

 log.likelihood  n df       BIC       ICL
      -431.8859 73 27 -979.6142 -990.9486

Clustering table:
 1  2
47 26
```

Segments Interpretation

```
lca_clusters <- lca_clust$classification
lca_clust_summary <- aggregate(std_seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance",
 "Comfort")],by=list(lca_clusters),FUN=mean)
lca_clust_summary
```

Group.1	Trendy	Styling	Reliability	Sportiness	Performance	Comfort
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.02627812	0.2424889	0.1364223	-0.2787552	-0.04411188	-0.1929288
2	-0.04750275	-0.4383453	-0.2466095	0.5039037	0.07974071	0.3487559

2 rows

Segments Interpretation

```
lca_clusters <- lca_clust$classification
lca_clust_summary <- aggregate(std_seg_data[,c("Trendy", "Styling", "Reliability", "Sportiness", "Performance",
 "Comfort")],by=list(lca_clusters),FUN=mean)
lca_clust_summary
```

Group.1 <dbl>	Trendy <dbl>	Styling <dbl>	Reliability <dbl>	Sportiness <dbl>	Performance <dbl>	Comfort <dbl>
Styling/Reliab	0.02627812	0.2424889	0.1364223	-0.2787552	-0.04411188	-0.1929288
Sport/Comfort	-0.04750275	-0.4383453	-0.2466095	0.5039037	0.07974071	0.3487559

Are the Segments Identifiable?

```
CrossTable(seg_data$MBA,lca_clusters,prop.chisq = FALSE, prop.r = T, prop.c = T,
          prop.t = F,chisq = T)
```

```
             | lca_clusters
seg_data$MBA | Reliability LCA |     Comfort LCA |       Row Total |
-------------|-----------------|-----------------|-----------------|
         MBA |              20 |               4 |              24 |
             |           0.833 |           0.167 |           0.329 |
             |           0.426 |           0.154 |                 |
-------------|-----------------|-----------------|-----------------|
   Undergrad |              27 |              22 |              49 |
             |           0.551 |           0.449 |           0.671 |
             |           0.574 |           0.846 |                 |
-------------|-----------------|-----------------|-----------------|
Column Total |              47 |              26 |              73 |
             |           0.644 |           0.356 |                 |
-------------|-----------------|-----------------|-----------------|
```

```
Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  5.599129     d.f. =  1     p =  0.01796941
```

Are the Segments Meaningful?

```
CrossTable(lca_clusters,seg_data$Choice,prop.chisq = FALSE, prop.r = T, prop.c = T,
          prop.t = F,chisq = T)
```

```
               | seg_data$Choice
  lca_clusters |       BMW |     Lexus |  Mercedes | Row Total |
---------------|-----------|-----------|-----------|-----------|
Reliability LCA |        21 |        17 |         9 |        47 |
               |     0.447 |     0.362 |     0.191 |     0.644 |
               |     0.656 |     0.773 |     0.474 |           |
---------------|-----------|-----------|-----------|-----------|
    Comfort LCA |        11 |         5 |        10 |        26 |
               |     0.423 |     0.192 |     0.385 |     0.356 |
               |     0.344 |     0.227 |     0.526 |           |
---------------|-----------|-----------|-----------|-----------|
   Column Total |        32 |        22 |        19 |        73 |
               |     0.438 |     0.301 |     0.260 |           |
---------------|-----------|-----------|-----------|-----------|


Pearson's Chi-squared test
------------------------------------------------------------
Chi^2 =  4.014183     d.f. =  2     p =  0.134379
```

Summary

- Market segmentation is a core concept in marketing
  - Customized offers to different groups of customers
  - Better targeted resources and higher CLV
- Be clear about which criteria to use for segmentation
- Use multiple methods for clustering (Hierarchical, K-Means, Latent Class)
  - Use judgment and statistical criteria to decide on number of clusters
  - Make sure that clusters are interpretable, identifiable, and managerially meaningful