Hierarchical clustering

David M. Blei

COS424
Princeton University

February 28, 2008

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.
- The idea is to build a binary tree of the data that successively merges similar groups of points

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.
- The idea is to build a binary tree of the data that successively merges similar groups of points
- Visualizing this tree provides a useful summary of the data

Hierarchical clusering vs. k-means

- Recall that k-means or k-medoids requires

Hierarchical clusering vs. k-means

- Recall that k-means or k-medoids requires
  - A number of clusters k

Hierarchical clusering vs. k-means

- Recall that k-means or k-medoids requires
  - A number of clusters k
  - An initial assignment of data to clusters

Hierarchical clusering vs. k-means

- Recall that k-means or k-medoids requires
  - A number of clusters k
  - An initial assignment of data to clusters
  - A distance measure between data $d(x_n, x_m)$

Hierarchical clusering vs. k-means

- Recall that k-means or k-medoids requires
  - A number of clusters k
  - An initial assignment of data to clusters
  - A distance measure between data $d(x_n, x_m)$
- Hierarchical clustering only requires a measure of similarity between *groups* of data points.

Agglomerative clustering

- We will talk about *agglomerative clustering*.

Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:

Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group

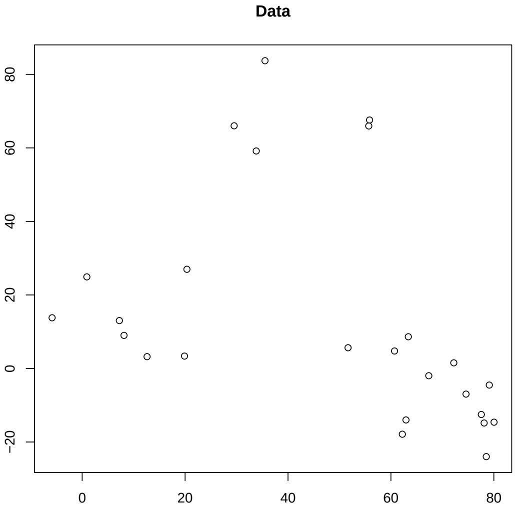
Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group
  2. Repeat: iteratively merge the two closest groups

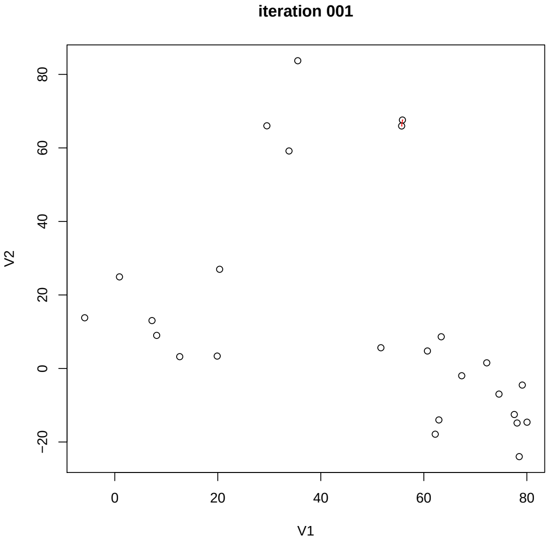
Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
  1. Place each data point into its own singleton group
  2. Repeat: iteratively merge the two closest groups
  3. Until: all the data are merged into a single cluster

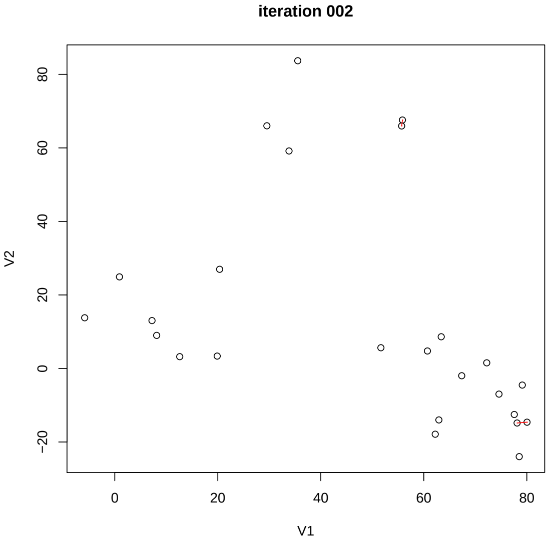
Example



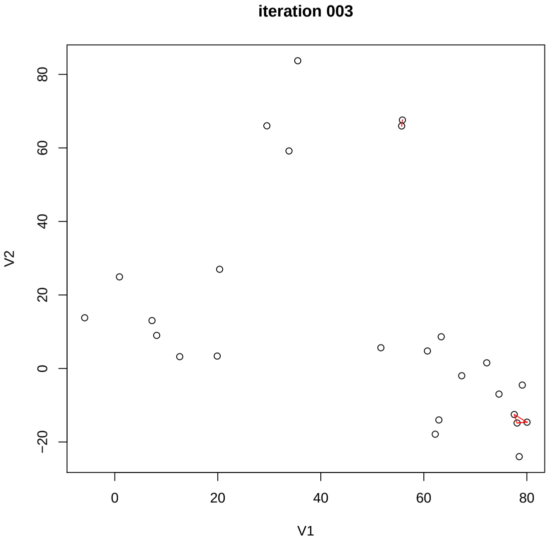
Example



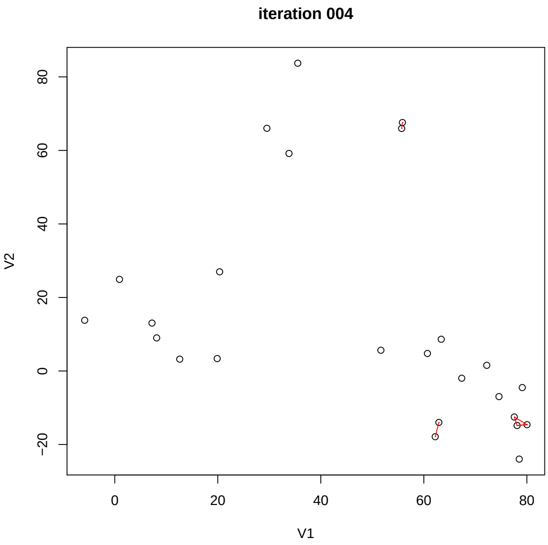
Example



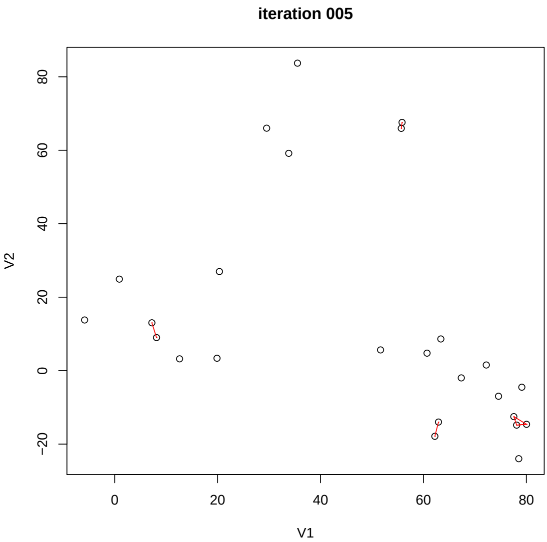
Example



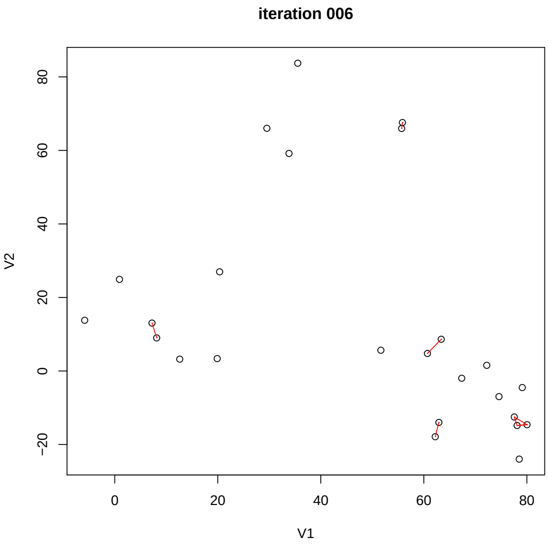
Example



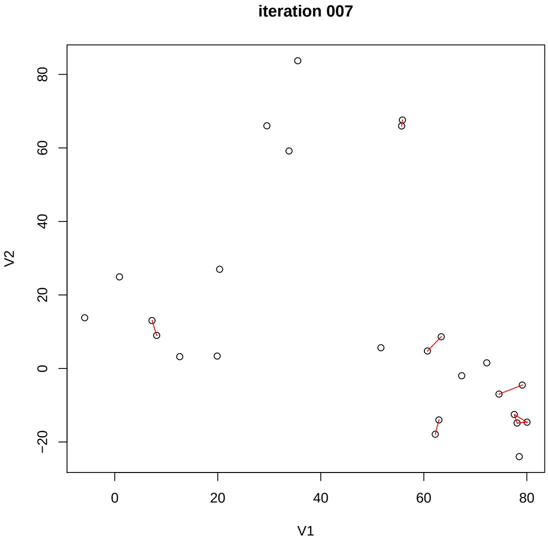
Example



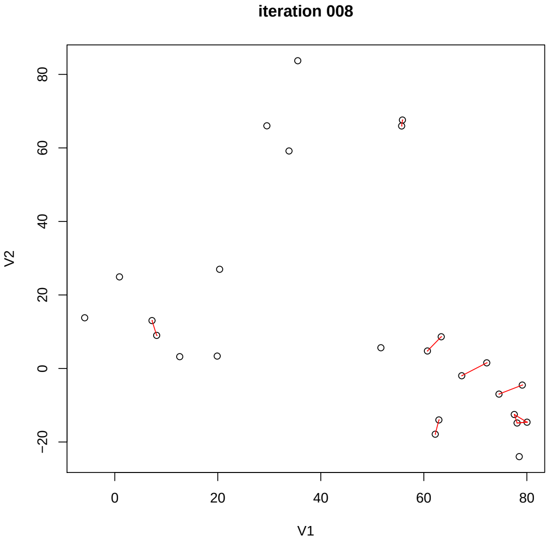
Example



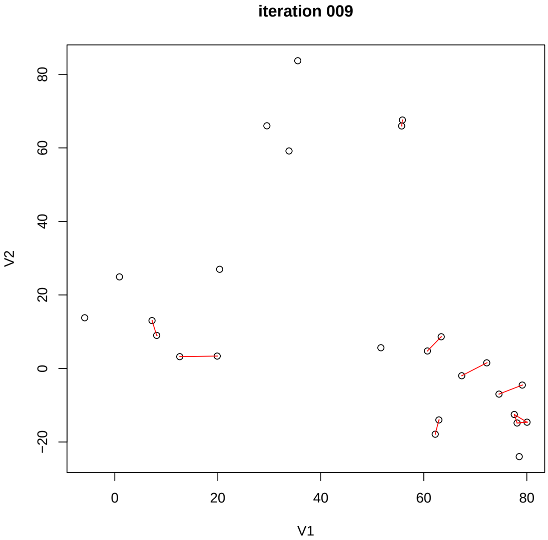
Example



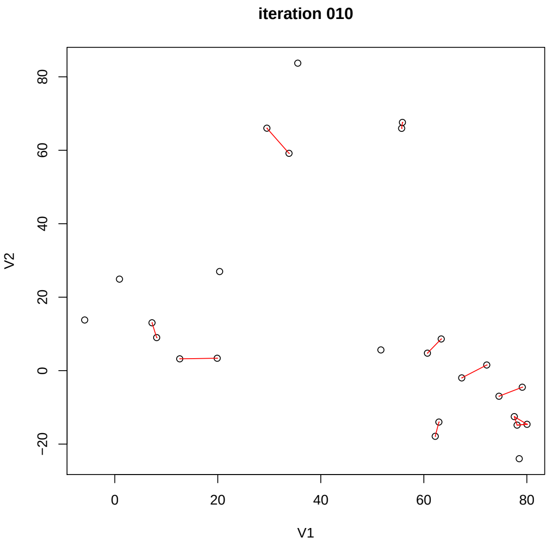
Example



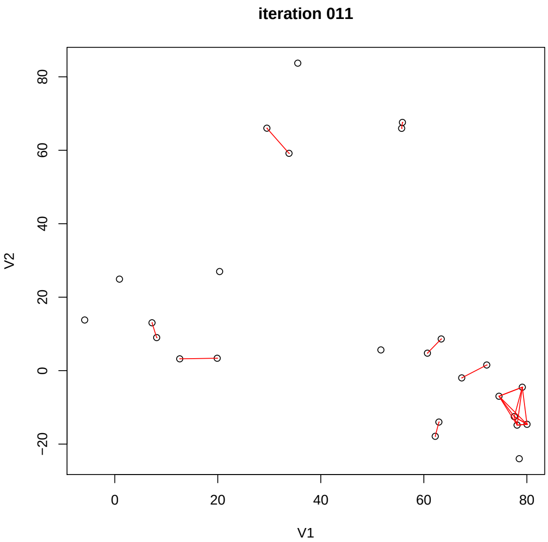
Example



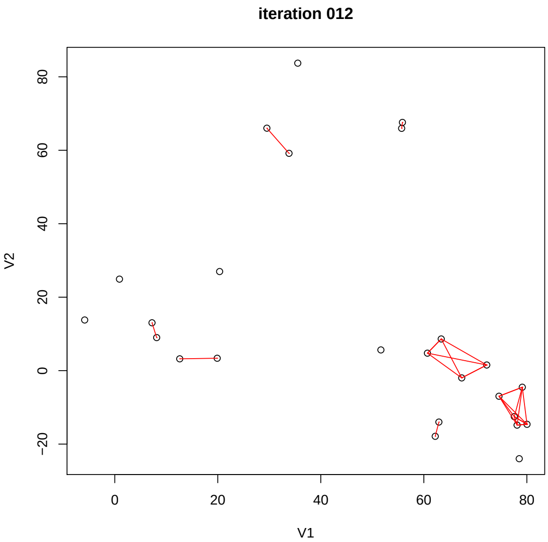
Example



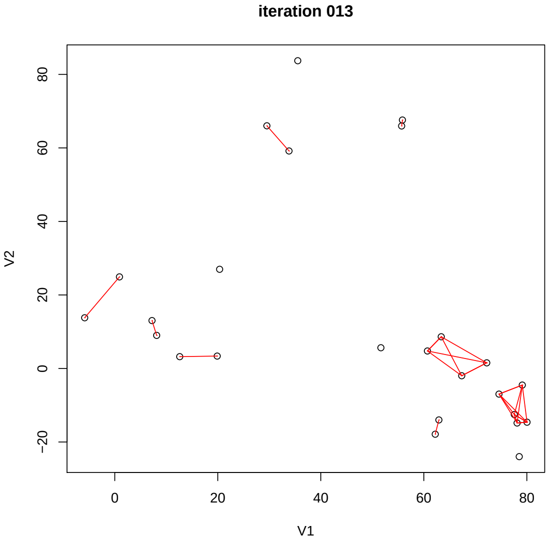
Example



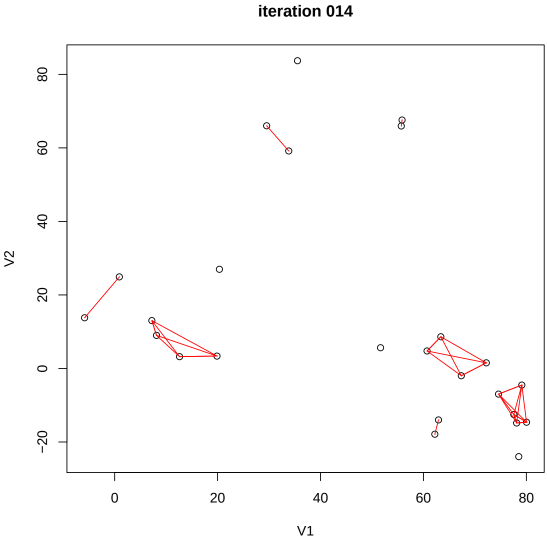
Example



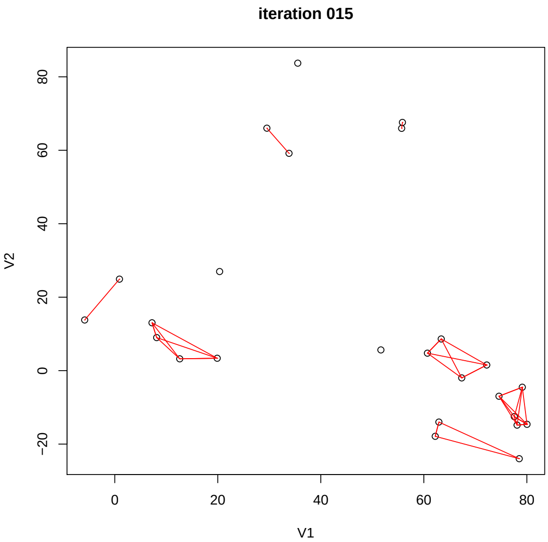
Example



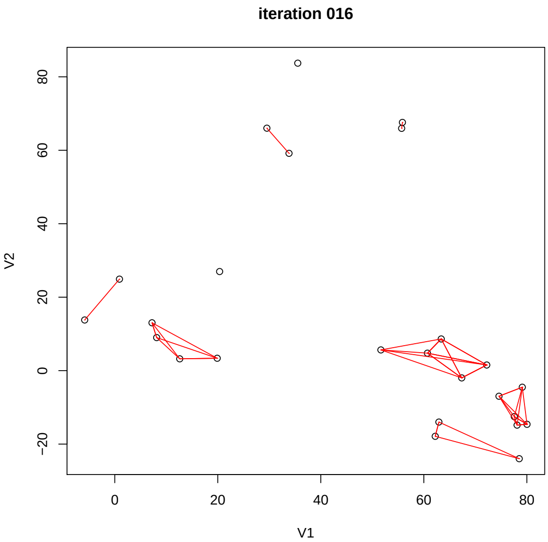
Example



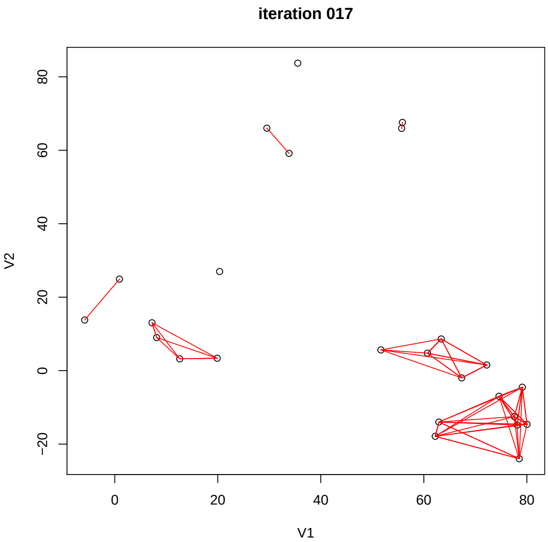
Example



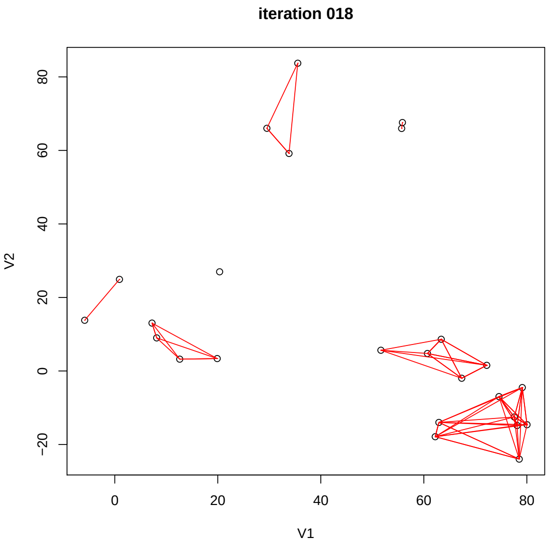
Example



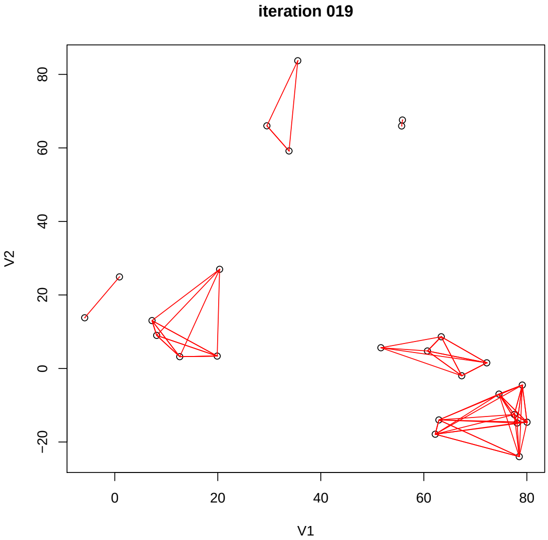
Example



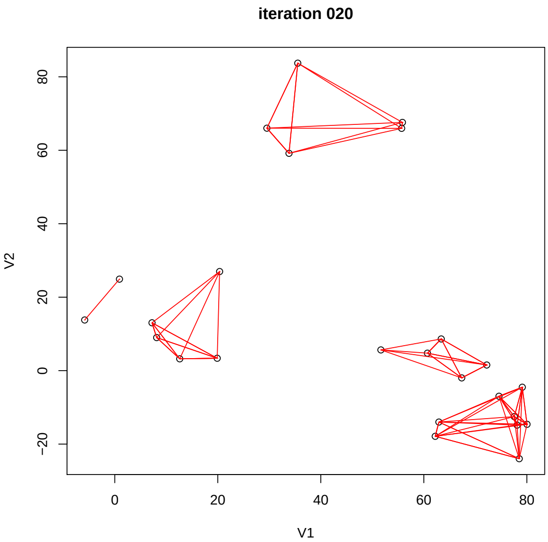
Example



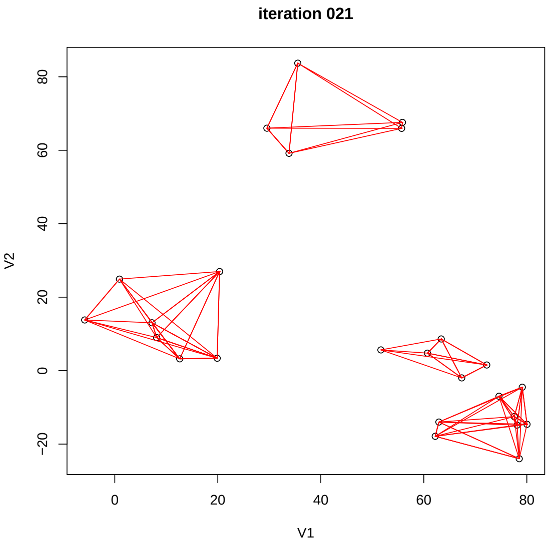
Example



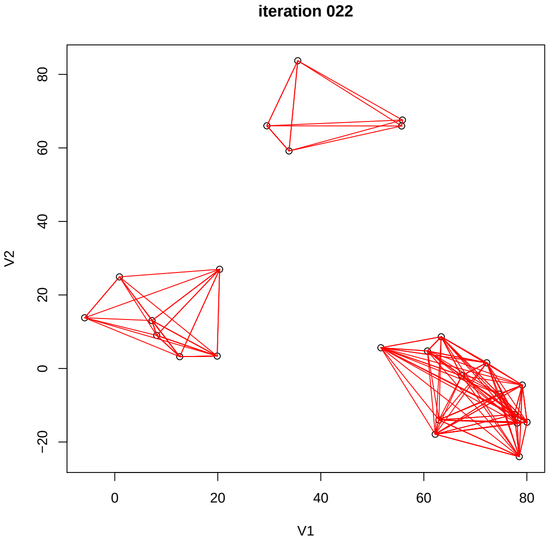
Example



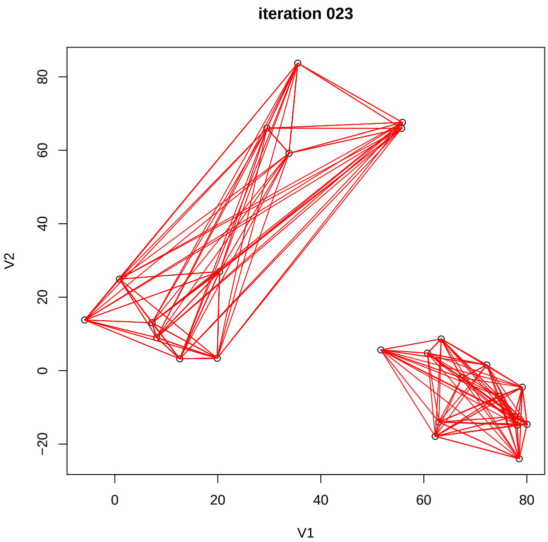
Example



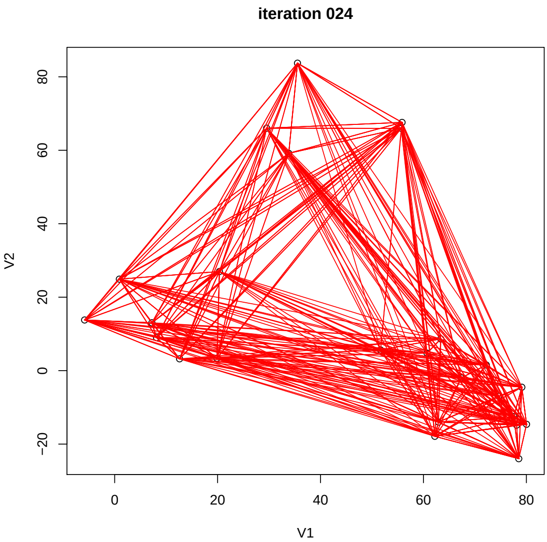
Example



Example



Example



Agglomerative clustering

- Each level of the resulting tree is a segmentation of the data

Agglomerative clustering

- Each level of the resulting tree is a segmentation of the data
- The algorithm results in a *sequence* of groupings

Agglomerative clustering

- Each level of the resulting tree is a segmentation of the data
- The algorithm results in a *sequence* of groupings
- It is up to the user to choose a "natural" clustering from this sequence

Dendrogram

- Agglomerative clustering is *monotonic*

Dendrogram

- Agglomerative clustering is *monotonic*
  - The similarity between merged clusters is monotone decreasing with the level of the merge.

Dendrogram

- Agglomerative clustering is *monotonic*
  - The similarity between merged clusters is monotone decreasing with the level of the merge.
- *Dendrogram*: Plot each merge at the (negative) similarity between the two merged groups

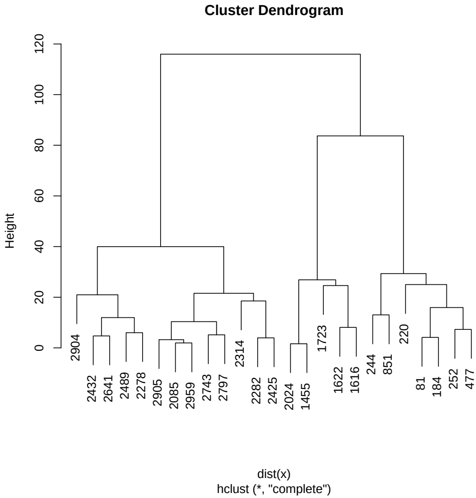
Dendrogram

- Agglomerative clustering is *monotonic*
  - The similarity between merged clusters is monotone decreasing with the level of the merge.
- *Dendrogram*: Plot each merge at the (negative) similarity between the two merged groups
- Provides an interpretable visualization of the algorithm and data

Dendrogram

- Agglomerative clustering is *monotonic*
  - The similarity between merged clusters is monotone decreasing with the level of the merge.
- *Dendrogram*: Plot each merge at the (negative) similarity between the two merged groups
- Provides an interpretable visualization of the algorithm and data
- Useful summarization tool, part of why hierarchical clustering is popular

Dendrogram of example data



Groups that merge at high values relative to the merger values of their subgroups are candidates for natural clusters. (Tibshirani et al., 2001)

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

  - *Complete linkage*: the similarity of the furthest pair

$$d_{CL}(G, H) = \max_{i \in G, j \in H} d_{i,j}$$

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
  - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

  - *Complete linkage*: the similarity of the furthest pair

$$d_{CL}(G, H) = \max_{i \in G, j \in H} d_{i,j}$$

  - *Group average*: the average similarity between groups

$$d_{GA} = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d_{i,j}$$

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups
- Complete linkage has the opposite problem. It might not merge close groups because of outlier members that are far apart.

Properties of intergroup similarity

- Single linkage can produce "chaining," where a sequence of close observations in different groups cause early merges of those groups
- Complete linkage has the opposite problem. It might not merge close groups because of outlier members that are far apart.
- Group average represents a natural compromise, but depends on the scale of the similarities. Applying a monotone transformation to the similarities can change the results.

Caveats

- Hierarchical clustering should be treated with caution.

Caveats

- Hierarchical clustering should be treated with caution.
- Different decisions about group similarities can lead to vastly different dendrograms.

Caveats

- Hierarchical clustering should be treated with caution.
- Different decisions about group similarities can lead to vastly different dendrograms.
- The algorithm *imposes* a hierarchical structure on the data, even data for which such structure is not appropriate.

Examples

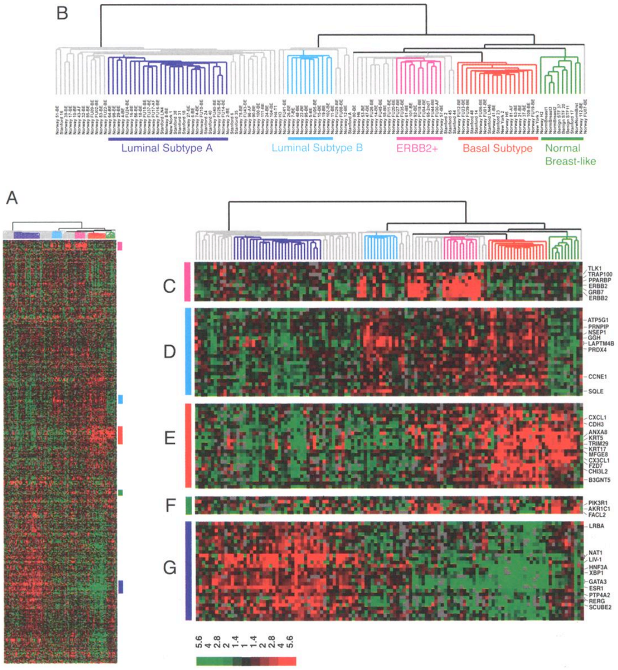
- "Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets" (Sorlie et al., 2003)

Examples

- "Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets" (Sorlie et al., 2003)
- Hierarchical clustering of gene expression data lead to new theories

Examples

- “Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets” (Sorlie et al., 2003)
- Hierarchical clustering of gene expression data lead to new theories
- Later, theories tested in the lab.


B
Luminal Subtype A
Luminal Subtype B
ERBB2+
Basal Subtype
Normal
Breast-like
A
C
D
E
F
G
TLK1
TRAP100
PPARBP
ERBB2
GRB7
ERBB2
ATP5G1
PRNPIP
NSEP1
GGH
LAPTM4B
PRDX4
CCNE1
SQLE
CXCL1
CDH3
ANXA8
KRT5
TRIM29
KRT17
MFGE8
CX3CL1
FZD7
CHI3L2
B3GNT5
PIK3R1
AKR1C1
FACL2
LRBA
NAT1
LIV-1
HNF3A
XBP1
GATA3
ESR1
PTP4A2
RERG
SCUBE2
5.6 4 2.8 2 1.4 1 1.4 2 2.8 4 5.6

Examples

- "The Balance of Roger de Piles" (Studdert-Kennedy and Davenport, 1974)

Examples

- "The Balance of Roger de Piles" (Studdert-Kennedy and Davenport, 1974)
- Roger de Piles rated 57 paintings along different dimensions.

Examples

- "The Balance of Roger de Piles" (Studdert-Kennedy and Davenport, 1974)
- Roger de Piles rated 57 paintings along different dimensions.
- These authors cluster them using different methods, including hierarchical clustering

Examples

- "The Balance of Roger de Piles" (Studdert-Kennedy and Davenport, 1974)
- Roger de Piles rated 57 paintings along different dimensions.
- These authors cluster them using different methods, including hierarchical clustering
- They discuss the different clusters. (They are art critics.)

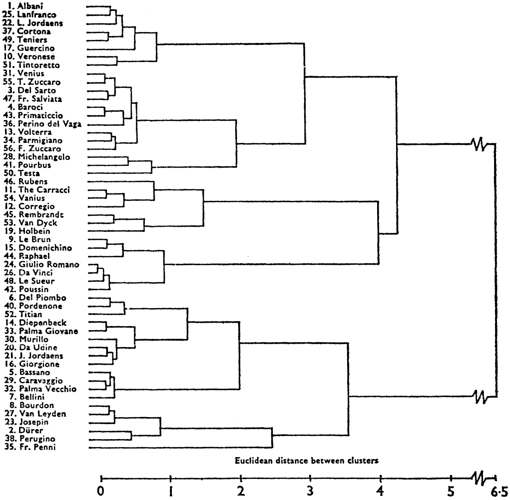


FIG. 1.

Good: They are cautious. “The value of this analysis...will depend on any interesting speculation it may provoke.”

Examples

- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)

Examples

- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities

Examples

- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities
- Use features such as

Examples

- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities
- Use features such as
    - # of staff in different departments

Examples

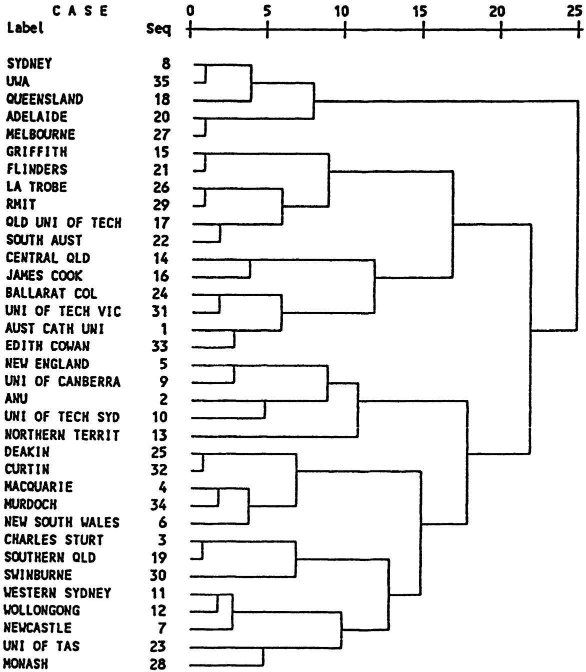
- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities
- Use features such as
    - # of staff in different departments
    - entry scores

Examples

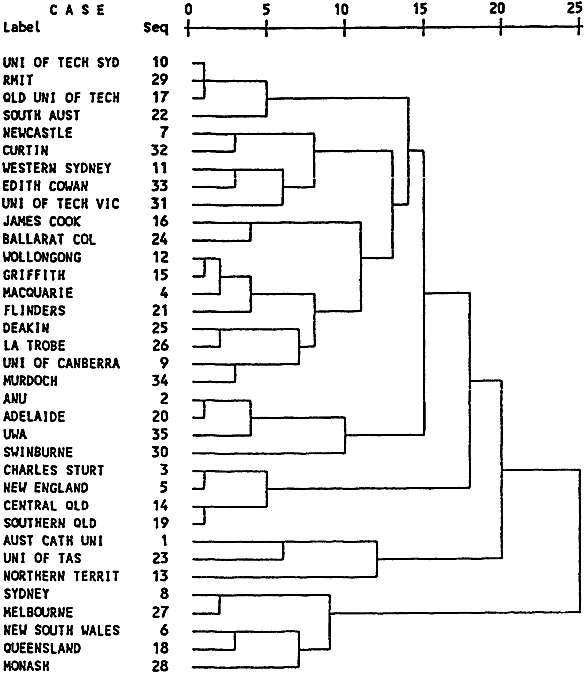
- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities
- Use features such as
  - # of staff in different departments
  - entry scores
  - funding

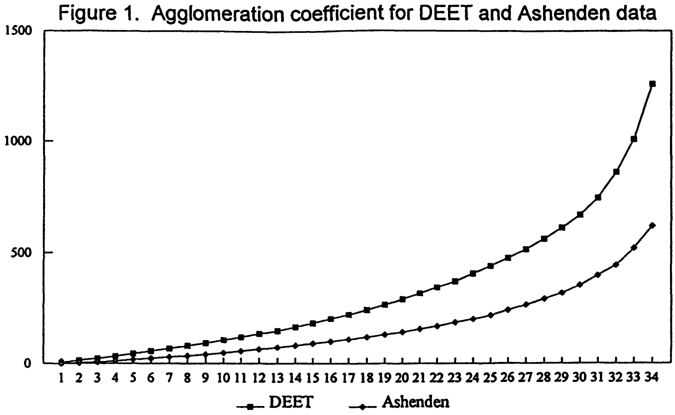
Examples

- "Similarity Grouping of Australian Universities" (Stanley and Reynlds, 1994)
- Use hierarchical clustering on Austrailian universities
- Use features such as
    - # of staff in different departments
    - entry scores
    - funding
    - evaluations



C A S E
Label
Seq
0
5
10
15
20
25
SYDNEY 8
UWA 35
QUEENSLAND 18
ADELAIDE 20
MELBOURNE 27
GRIFFITH 15
FLINDERS 21
LA TROBE 26
RMIT 29
QLD UNI OF TECH 17
SOUTH AUST 22
CENTRAL QLD 14
JAMES COOK 16
BALLARAT COL 24
UNI OF TECH VIC 31
AUST CATH UNI 1
EDITH COWAN 33
NEW ENGLAND 5
UNI OF CANBERRA 9
ANU 2
UNI OF TECH SYD 10
NORTHERN TERRIT 13
DEAKIN 25
CURTIN 32
MACQUARIE 4
MURDOCH 34
NEW SOUTH WALES 6
CHARLES STURT 3
SOUTHERN QLD 19
SWINBURNE 30
WESTERN SYDNEY 11
WOLLONGONG 12
NEWCASTLE 7
UNI OF TAS 23
MONASH 28


CASE
Label
Seq
0
5
10
15
20
25
UNI OF TECH SYD 10
RMIT 29
QLD UNI OF TECH 17
SOUTH AUST 22
NEWCASTLE 7
CURTIN 32
WESTERN SYDNEY 11
EDITH COWAN 33
UNI OF TECH VIC 31
JAMES COOK 16
BALLARAT COL 24
WOLLONGONG 12
GRIFFITH 15
MACQUARIE 4
FLINDERS 21
DEAKIN 25
LA TROBE 26
UNI OF CANBERRA 9
MURDOCH 34
ANU 2
ADELAIDE 20
UWA 35
SWINBURNE 30
CHARLES STURT 3
NEW ENGLAND 5
CENTRAL QLD 14
SOUTHERN QLD 19
AUST CATH UNI 1
UNI OF TAS 23
NORTHERN TERRIT 13
SYDNEY 8
MELBOURNE 27
NEW SOUTH WALES 6
QUEENSLAND 18
MONASH 28



- Split values: They notice that there's no kink and conclude that there is no cluster structure in Austrailian universities.
- **Good**: Cautious interpretation of clustering, analysis of clustering based on multiple subsets of the features.
- **Bad**: Their conclusions—we can't cluster Australian universities—ignores all the algorithmic choices that were made.

Examples

- "Comovement of International Equity Markets: A Taxonomic Approach" (Panton et al., 1976)

Examples

- "Comovement of International Equity Markets: A Taxonomic Approach" (Panton et al., 1976)
- Data: weekly rates of return for stocks in twelve countries

Examples

- "Comovement of International Equity Markets: A Taxonomic Approach" (Panton et al., 1976)
- Data: weekly rates of return for stocks in twelve countries
- Run agglometerative clustering year by year

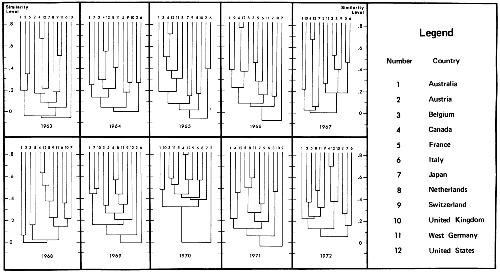
Examples

- "Comovement of International Equity Markets: A Taxonomic Approach" (Panton et al., 1976)
- Data: weekly rates of return for stocks in twelve countries
- Run agglometerative clustering year by year
- Interpret the structure and examine stability over different time periods

Examples

FIGURE II

ONE-YEAR DENDROGRAMS
1963-1972



Good: Cautious. "This study is only descriptive...A logical subsequent research area is to explain observed structural properties and the causes of structural change."