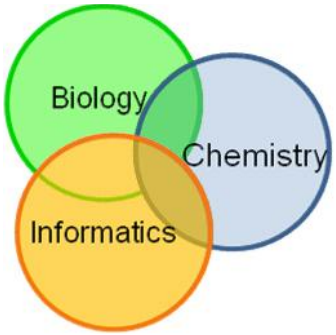


Evaluation of metabolomic sample processing methods using hierarchical cluster analysis

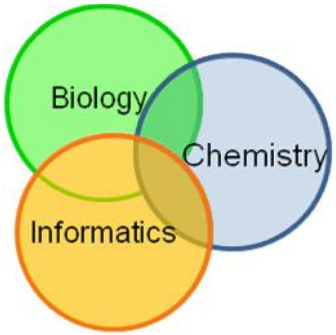
Cluster Analysis

Goal:

Use hierarchical cluster analysis (HCA) to evaluate data variance structure

Topics:

1. Evaluate sample and variable similarities
2. Identify the effect of data transformation, distance and linkage methods on data similarities



Clustering data

Cluster Analysis

Use DATA: **Pumpkin data 1.csv**

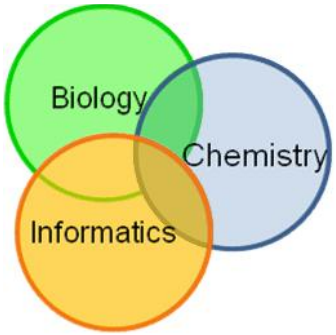
Goal: Use HCA to cluster samples

Visualize:

1. Sample (row) raw similarities as a heat map
2. Annotate heatmap with extraction and treatment type
3. Select cluster distance and linkage method to cluster the samples
4. Determine the effect of data transformations on the cluster structure (view as a dendrogram)

Exercises:

1. What factor, extraction or treatment, has the greatest contribution to the data variance structure?
2. Describe the effect of clustering raw data or sample correlations

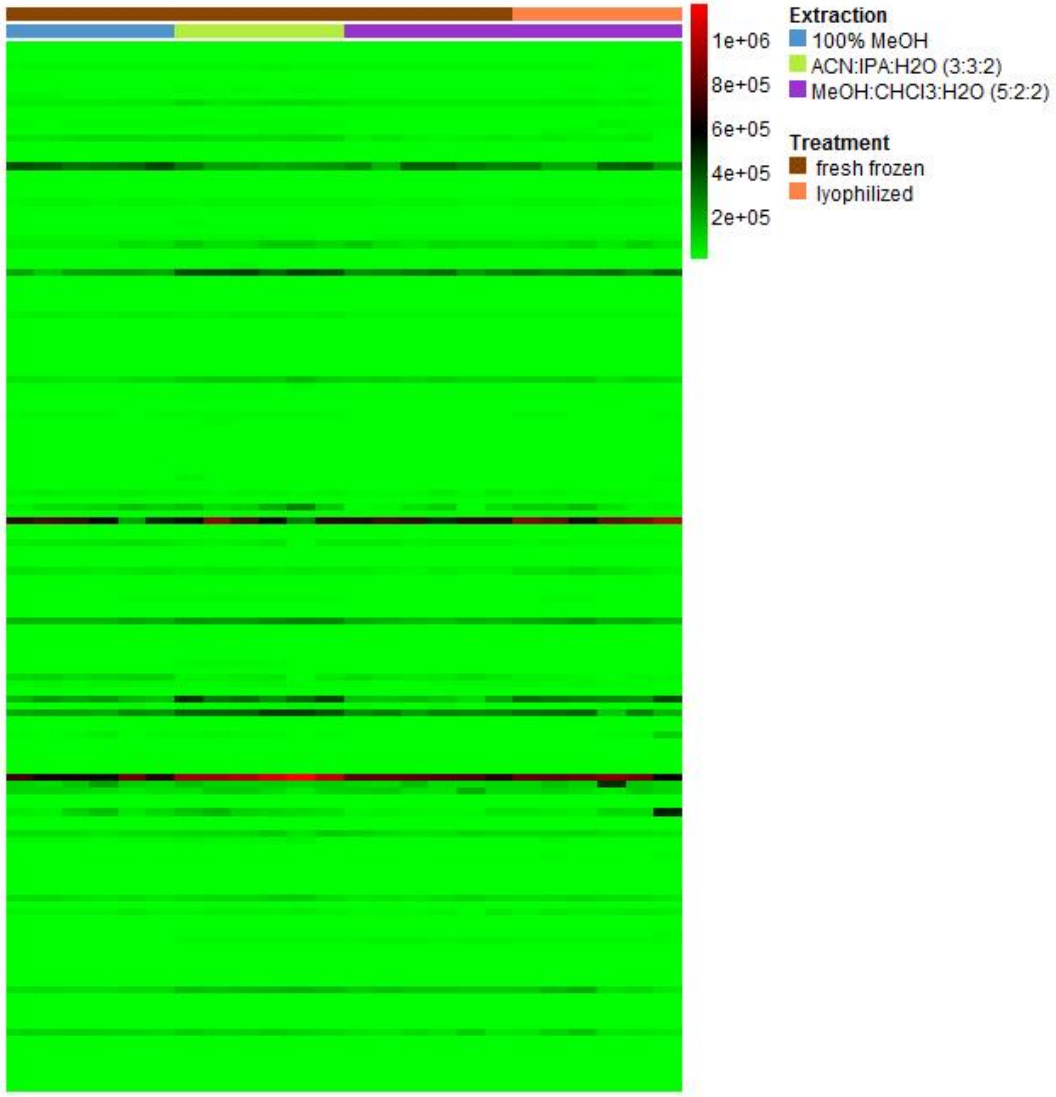


Raw data matrix visualized as a heatmap

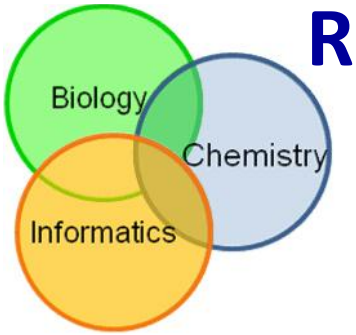
Cluster Analysis

samples

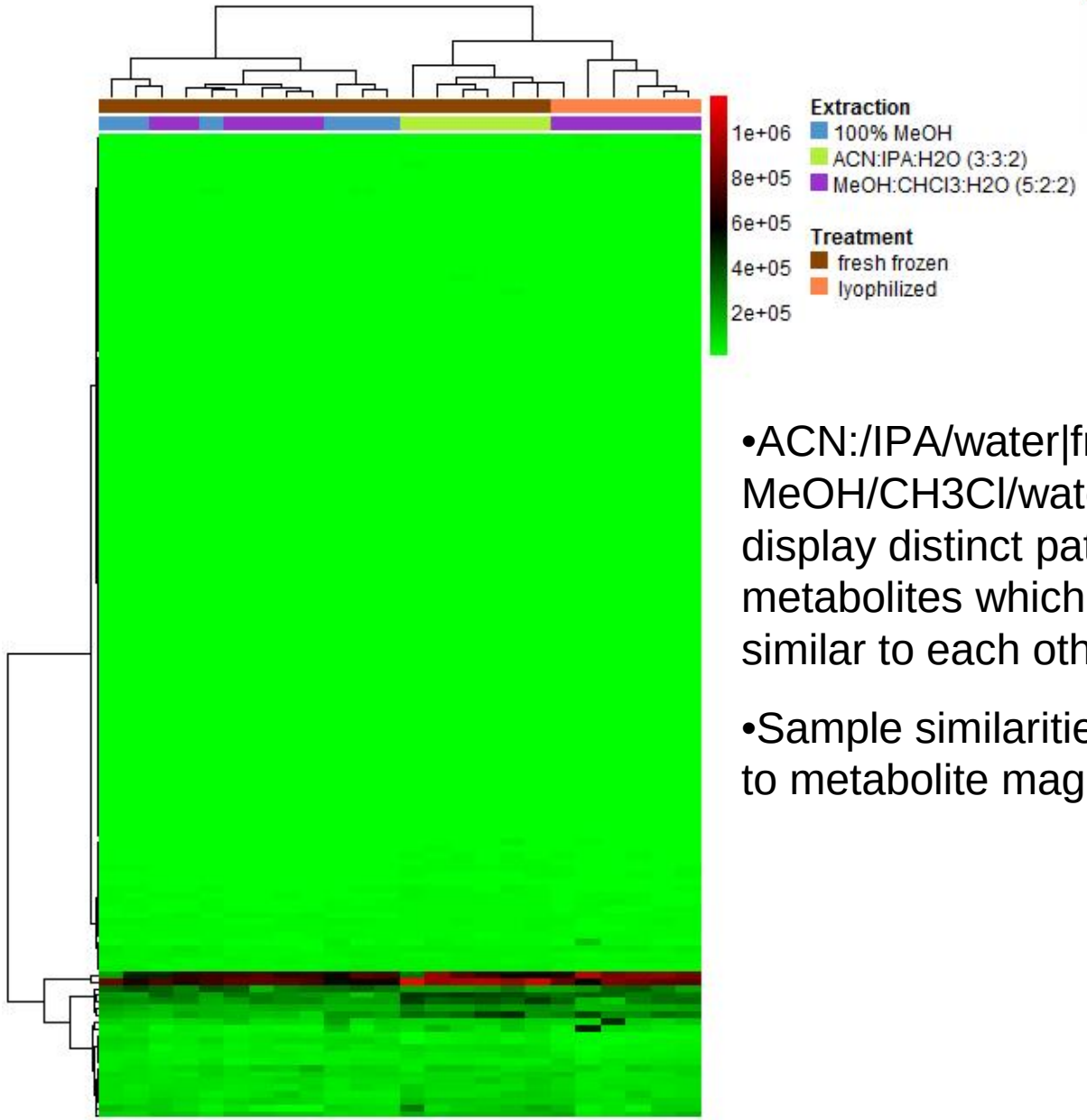
variables



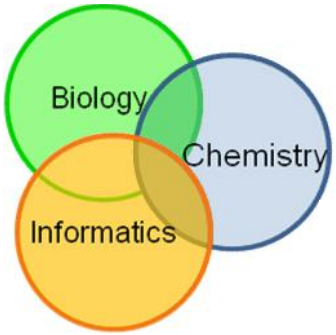
Raw data matrix organized by HCA



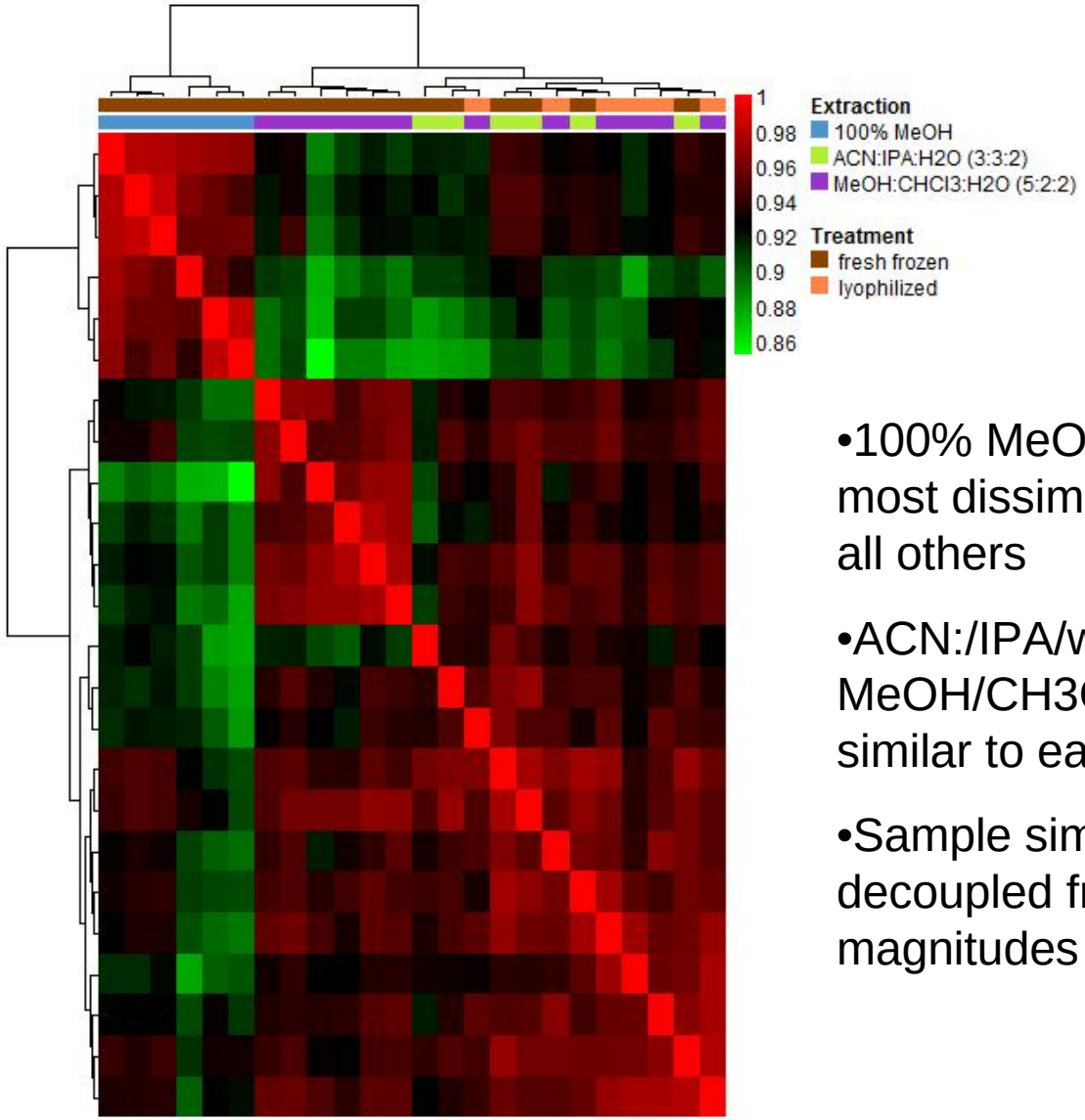
Cluster Analysis



- ACN:/IPA/water|fresh and MeOH/CH3Cl/water|dried display distinct patterns in metabolites which are most similar to each other
- Sample similarities are linked to metabolite magnitudes

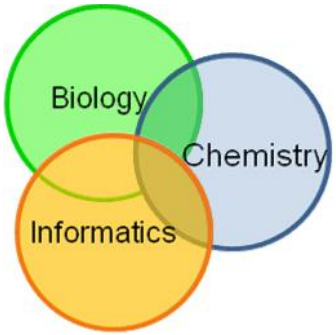


Clustering based on sample correlations (spearman)



Cluster Analysis

- 100% MeOH/fresh is the most dissimilar protocol from all others
- ACN:/IPA/water and MeOH/CH3Cl/water are most similar to each other
- Sample similarities are decoupled from metabolite magnitudes



Clustering metabolites

Cluster Analysis

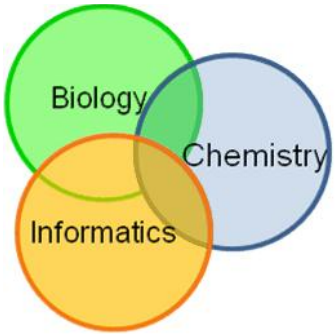
Goal 2: Use HCA to evaluate metabolite similarities

Visualize:

1.Z-scaled and correlation based variable clustering
2.Use a dendrogram to extract variable clusters
3.Select two variables from the same cluster and visualize their correlation

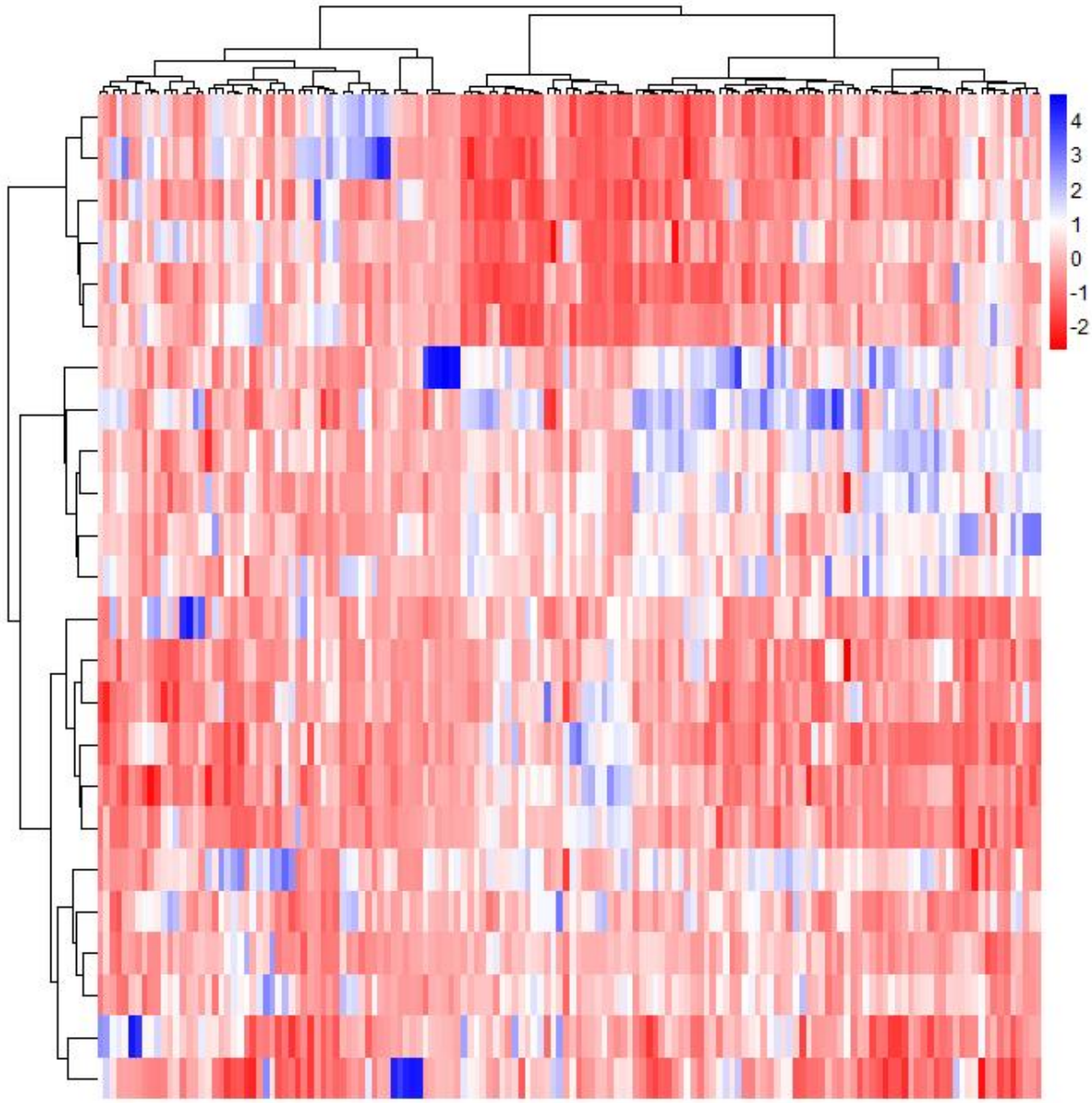
Exercise:

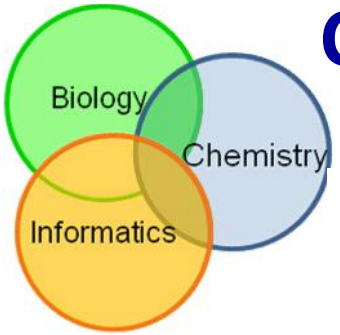
1.Do the clustered variables share biological functions?
2.Which type of correlation is most robust to outliers?
3.Are the correlations for the visualized variable independent of extraction/treatment?



Z-scaled variable clusters

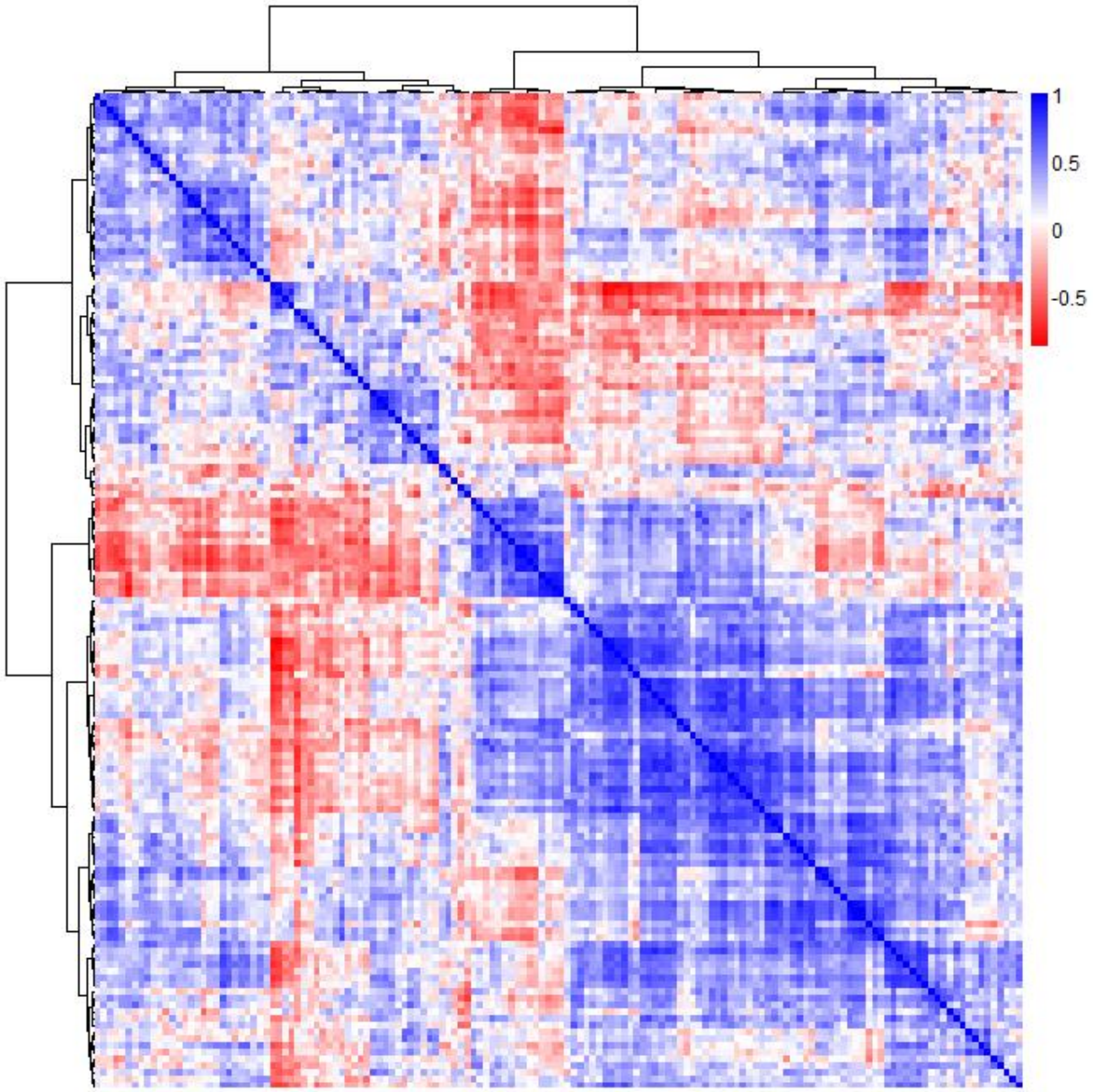
Cluster Analysis

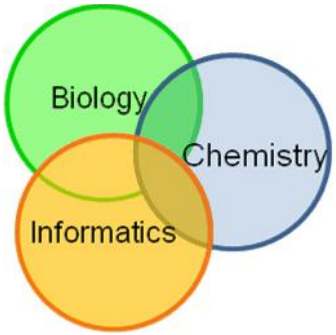




Correlation based variable clusters

Cluster Analysis





Extraction of clusters of correlated variables

cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
cluster 6
cluster 7

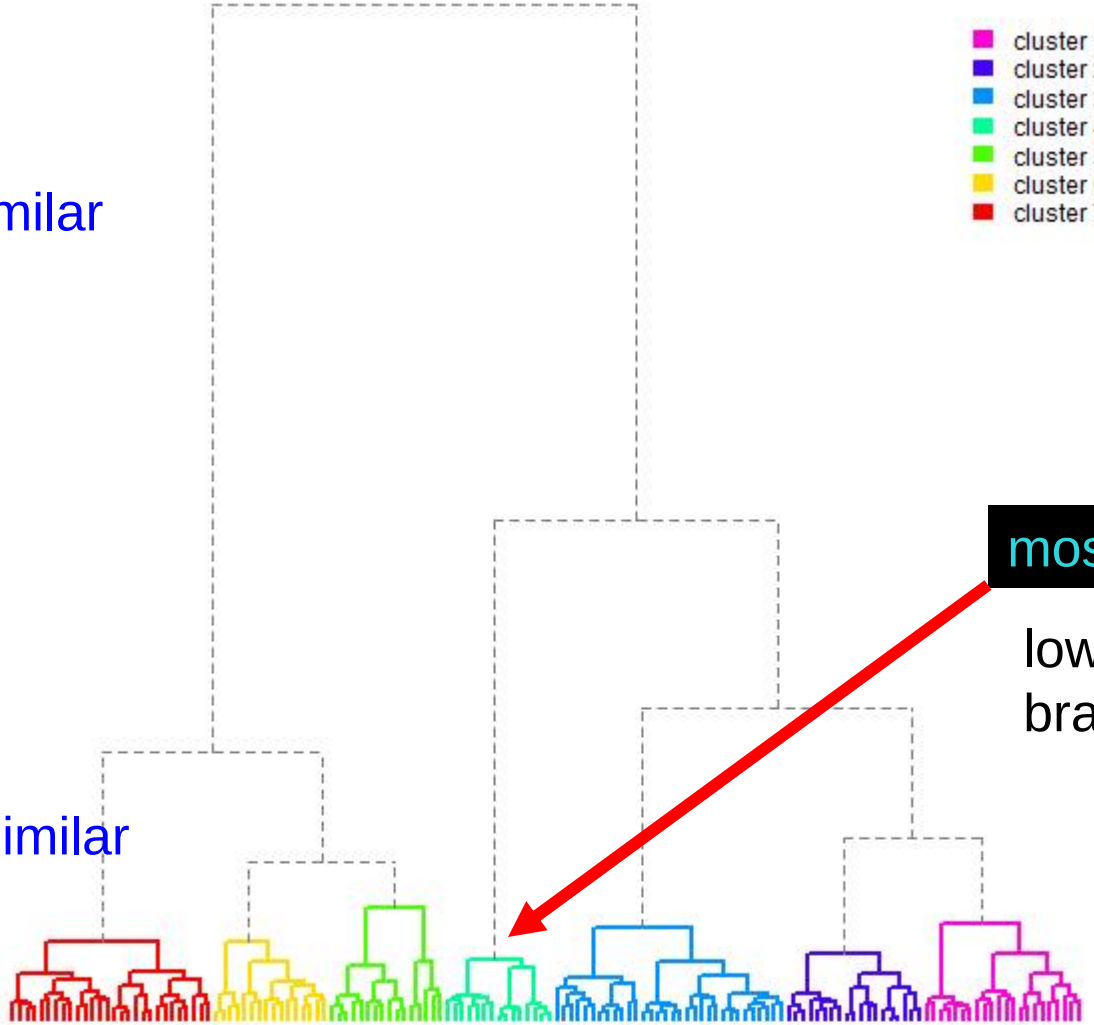
less similar

more similar

Cluster Analysis

most similar cluster

lowest common branch height



Correlation among cluster members (4)

Cluster Analysis