Dissimilarity, cluster analysis, diversity

Gavin Simpson

November 21, 2022

Welcome

Logistics

Today's topics

- Multivariate data
- Diversity
- Dissimilarity
- Cluster analysis

Multivariate data

Multivariate data

Multivariate != multiple regression

Multivariate means we have two or more *response* variables

We are interested in learning about the common patterns or modes of variation among those multiple response variables

Multivariate data require special statistical methods

Multivariate data in biology

Community composition — many species as the responses

In modern biology we have OTUs and ASVs

In chemistry we have metabolites or spectra or masses or ...

All of these constitute multivariate data

Species composition

Species composition is the set of species in a site or a sample

Typically comes with a measure of abundance

"abundance" could also be whether each species is present or absent in the sample

Relative abundance expresses the abundance of each species as its proportion out of the total abundance in each sample

Dissimilarity

Measuring association — binary data

		Object j	
		+	-
Object i	+	a	c
	-	b	d

Dissimilarity based on the number of species present only in i (b), or j (c), or in present in both (a), or absent in both (d).

Jaccard similarity

$$s_{ij} = \frac{a}{a+b+c}$$

Simple matching coefficient

$$s_{ij} = \frac{a+d}{a+b+c+d}$$

Jaccard dissimilarity

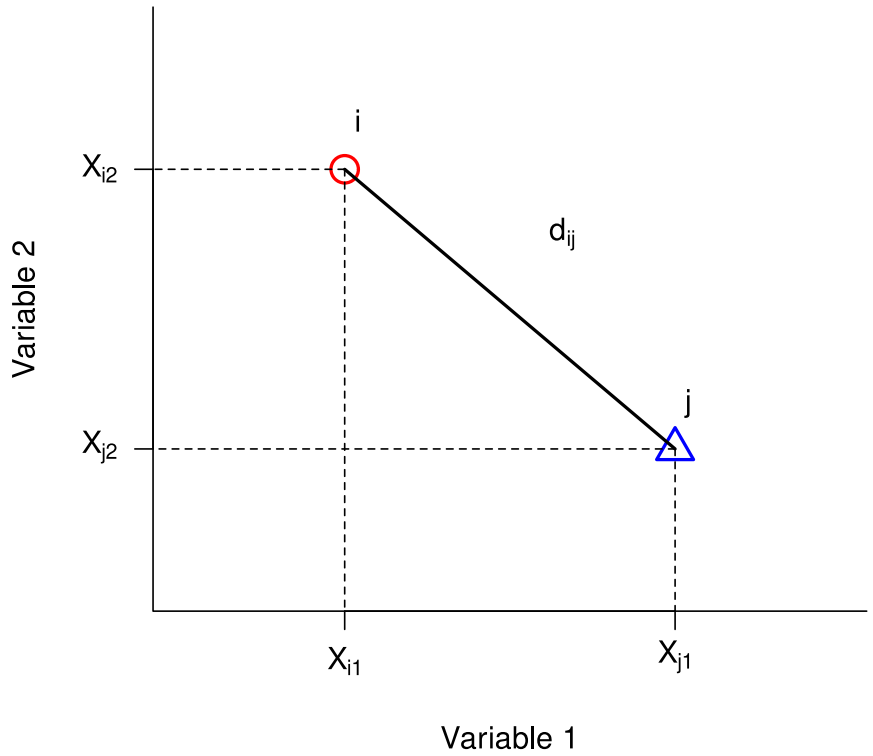
$$d_{ij} = \frac{b+c}{a+b+c}$$

Simple matching coefficient

$$d_{ij} = \frac{b+c}{a+b+c+d}$$

Dissimilarity

$$d_{ij} = \sqrt{(X_{j1} - X_{i1})^2 + (X_{j2} - X_{i2})^2}$$



$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}$$

$$d_{ij} = \sum_{k=1}^{m} |x_{ik} - x_{jk}|$$

$$d_{ij} = \frac{\sum_{k=1}^{m} |x_{ik} - x_{jk}|}{\sum_{k=1}^{m} (x_{ik} + x_{jk})}$$

Measuring association — quantitative data

- Euclidean distance dominated by large values.
- Manhattan distance less affected by large values.
- Bray-Curtis sensitive to extreme values.
- Similarity ratio (Steinhaus-Marczewski $\equiv$ Jaccard) less dominated by extremes.
- Chord distance, used for proportional data; *signal-to-noise* measure.

Similarity ratio

Chord distance

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (\sqrt{p_{ik}} - \sqrt{p_{jk}})^2}$$

Measuring association — mixed data

- s_{ijk} is similarity between sites i and j for the kth variable.
- Weights w_{ijk} are typically 0 or 1 depending on whether the comparison is valid for variable k. Can also use variable weighting with w_{ijk} between 0 and 1.
- w_{ijk} is zero if the kth variable is missing for one or both of i or j.
- For binary variables s_{ijk} is the Jaccard coefficient.
- For categorical data s_{ijk} is 1 of i and k have same category, 0 otherwise.
- For quantitative data $s_{ijk} = (1 - |x_{ik} - x_{jk}|)/R_k$

Gower's coefficient

$$s_{ij} = \frac{\sum_{i=1}^{m} w_{ijk} s_{ijk}}{\sum_{i=1}^{m} w_{ijk}}$$

Transformations

- Can transform the variables (e.g. species) or the samples to improve the gradient separation of the dissimilarity coefficient.
- No transformation of variables or samples leads to a situation of quantity domination — big values dominate d_{ij}.
- Normalise samples —gives all samples equal weight.
- Normalise variables;
  - gives all variables equal weight,
  - inflates the influence of rare species.
- Double (*Wisconsin*) transformation; normalise variables then samples.
- Noy-Meir *et al*. (1975) *J. Ecology* **63**; 779--800.
- Faith *et al*. (1987) *Vegetatio* **69**; 57--68.

Dissimilarity

Two key functions

1. `vegdist()`
2. `decostand()`

```
data(varespec)

euc_dij ← vegdist(varespec, method = "euclidean")

bc_dij ← vegdist(varespec)

hell_dij ← vegdist(decostand(varespec, method = "hellinger"),
                   method = "euclidean")
```

Cluster analysis

Basic aim of cluster analysis

- Partition a set of data (objects, samples) into groups known as clusters.
- Partitions formed that minimise a stated mathematical criterion, e.g. sum of squares (SS)
  - Minimise within groups SS $\rightarrow$ maximising between group SS
- Cluster analysis is a compromise however:
  - With 50 objects there are 10^{80} possible ways of partitioning the objects
- Compromise is made in selecting a clustering scheme that reduces the number of partitions to a reasonable value.
- Commonest approaches impose a hierarchy and then either fuse (*agglomerative*) or split (*divisive*) samples and clusters.

A taxonomy of clusterings

- Clustering techniques can be characterised in many ways:

Formal	Informal
Hierarchical	Non-hierarchical
Quantitative	Qualitative
Agglomerative	Divisive
Polythetic	Monothetic
Sharp	Fuzzy
Supervised	Unsupervised
Useful	Not useful

A cautionary tale

- Cluster analysis *per se* is *unsupervised*; we want to find groups in our data.
- We will define *classification* as a *supervised* procedure where we know *a priori* what the groups are.

A cautionary tale

> The availability of computer packages of classification techniques has led to the waste of more valuable scientific time than any other "statistical" innovation (with the possible exception of multiple regression techniques). R.M. Cormack (1970) *J. Roy. Stat. Soc. A* **134(3)**; 321--367.

Hierarchical cluster analysis

Agglomerative hierarchical clustering

- Agglomerative methods start with all observations in separate clusters and fuse the two most similar observations.
- Fusion continues, with the two most similar observations and/or clusters being fused at each step.
- Five main stages in this analysis:
  - calculate matrix of *(dis)similarities* d_{ij} between all pairs of m objects,
  - fuse objects into groups using chosen *clustering strategy*,
  - display the results graphically via a *dendrogram* or superimposed on to an ordination,
  - check for distortion,
  - validate the results

Clustering strategies

- Different strategies can be used to determine the dissimilarity (*distance*) between a sample and a cluster or two clusters.
- Single link or nearest neighbour;
  - the distance between the closest members of two clusters,
  - finds the minimum spanning tree, the shortest tree connecting all points,
  - can produce *chaining*, producing groups of unequal size.
- Complete link or furthest neighbour;
  - the distance between the farthest members of two clusters,
  - produces compact clusters of roughly equal size,
  - may make compact groups even when none exist

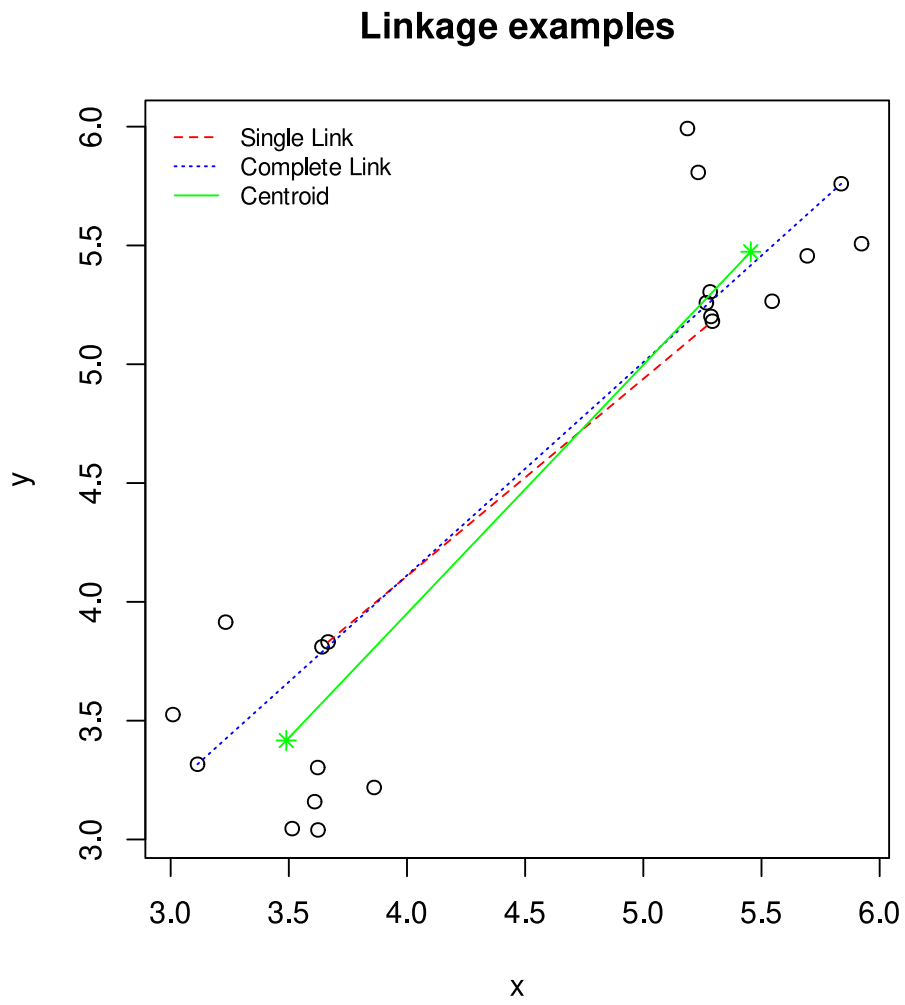
Clustering strategies

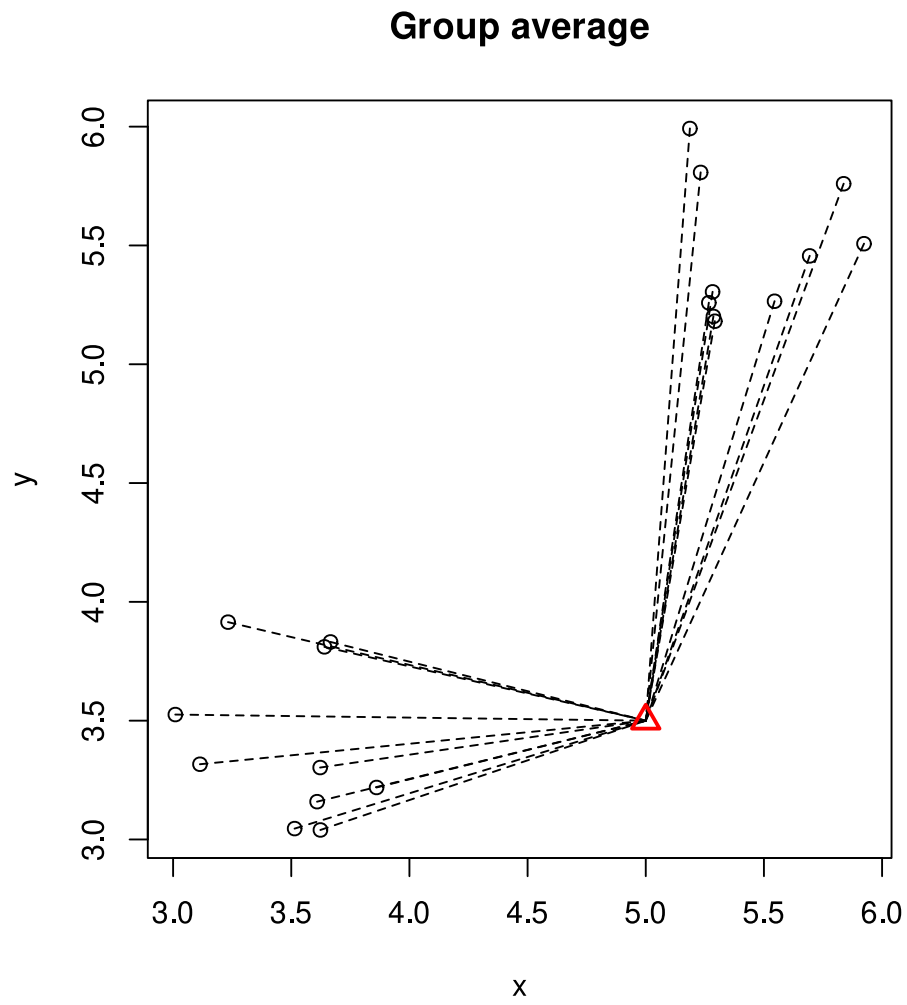
- Centroid;
  - the distance between the centroids (the centre of gravity) of two clusters,
  - centroid is the point that is the average of the coordinates (variables) of all objects in the cluster
  - can produce reversals
- Unweighted group average
  - the average of the distances between the samples in one cluster and the samples of another,
  - intermediate between single and complete link methods,
  - maximises the \alert{cophenetic correlation},
  - may make compact clusters where none exist,

Clustering strategies

- Minimum variance (Ward's method)
  - fuse two groups if and only if neither group will combine with any other group to produce a lower within group \alert{sum of squares}
  - forms compact clusters of equal size; even where none exist,

Clustering strategies

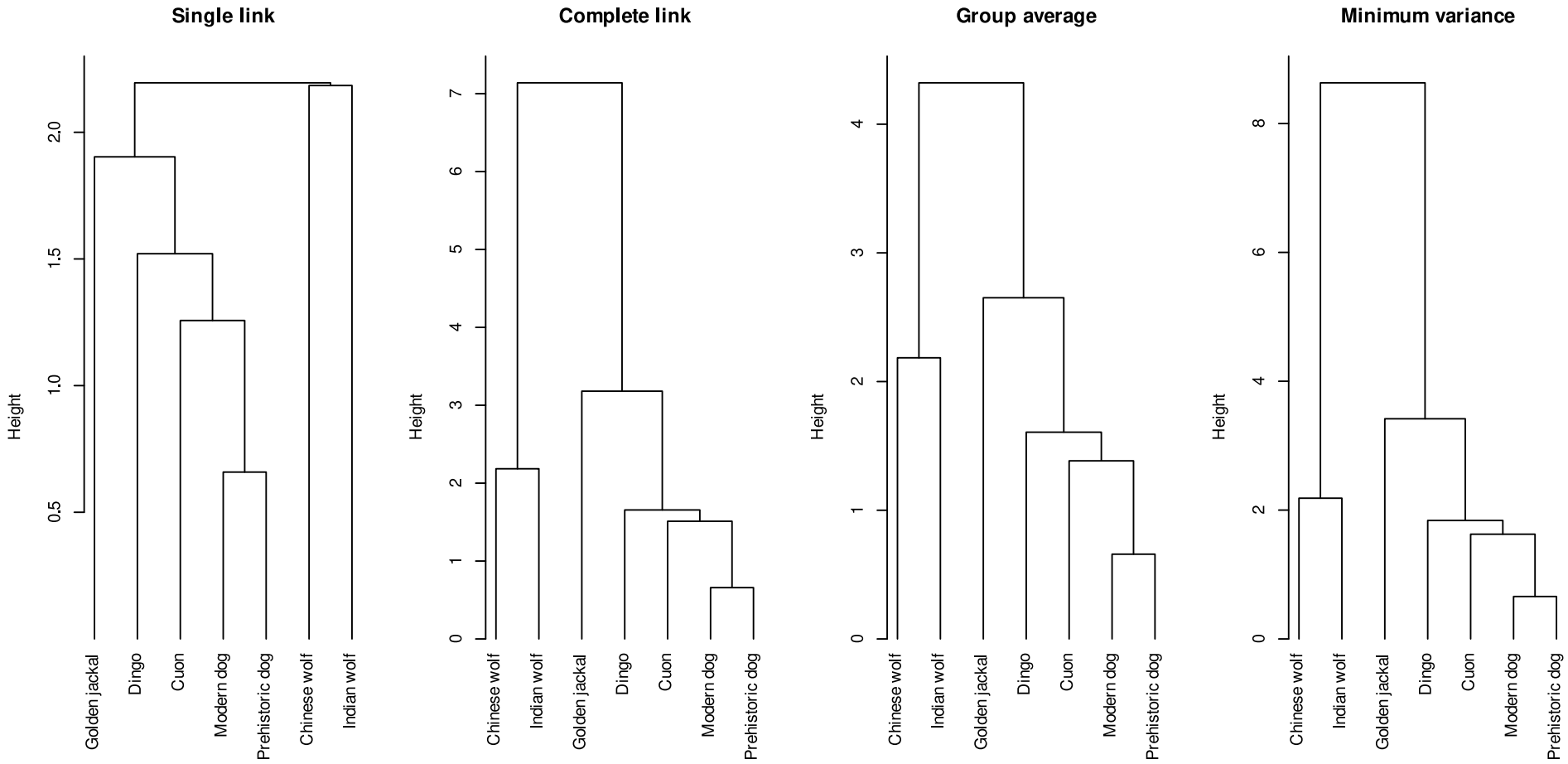




Prehistoric dogs from Thailand

- Archaeological digs at prehistoric sites in Thailand have produced a collection of canine mandibles, covering period 3500BC to today
- Origins of the prehistoric dog uncertain
  - Could possibly descend from the golden jackal or the wolf
  - Wolf not native to Thailand; nearest indigenous wolves in western China or Indian subcontinent
- Data are mean values of each of six measurements of specimens from 7 canine groups

Prehistoric dogs from Thailand

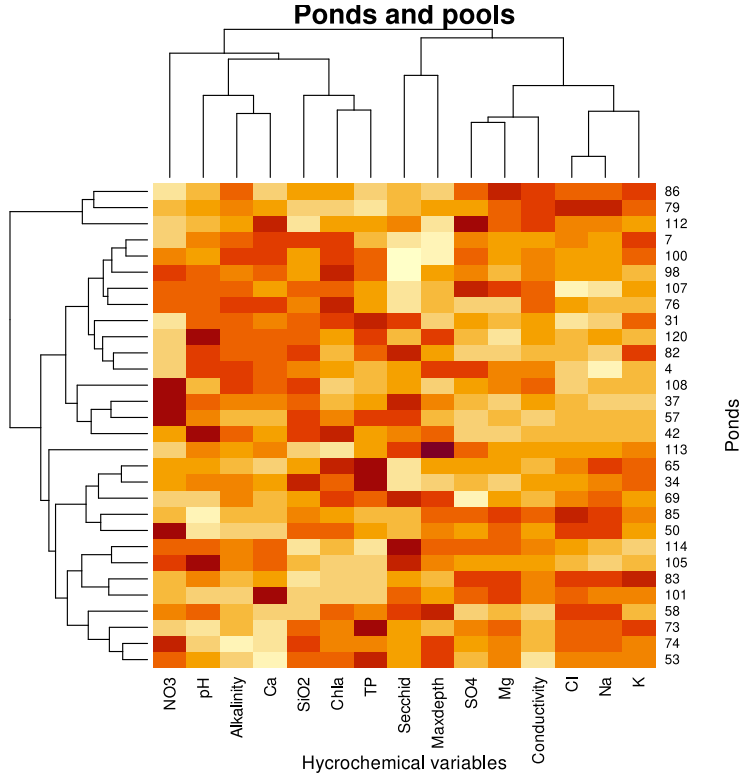


Graphical display; dendrograms

- The height is the dissimilarity at which the fusion was made.
- The length of the *stem* represents the dissimilarities between clusters.
- *Nodes* represent clusters; internal or terminal.
- Configuration of nodes and stems is known as the tree *topology*.
- Can flip two adjacent stems without affecting topology.
- Form a clustering by cutting the dendrogram.

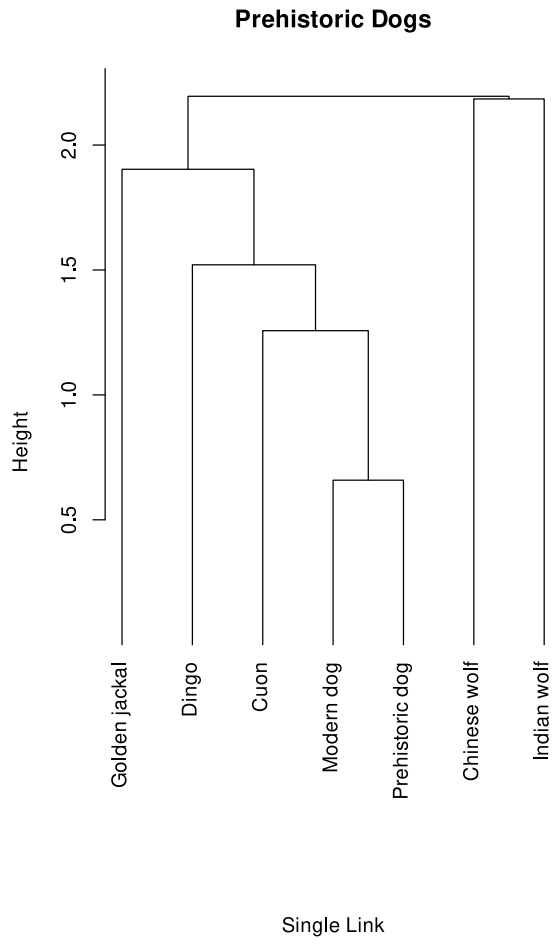
Graphical display; heatmap

- A heatmap is another graphical display of data structure.
- Begin by reordering the rows (sites) and columns (variables) of data matrix.
- Reorder using cluster analysis on rows and columns separately.
- Central panel shows the values of the variables for each site as a shading.



Test for distortion

- The hierarchical cluster analysis represents the underlying original dissimilarities
- Check the dendrogram for distortion
- Calculate the *cophenetic* distances; the heights on the dendrogram where samples are fused
- Calculate the Pearson correlation between the original dissimilarities and cophenetic distances; *cophenetic correlation*
- Cophenetic correlation = 0.712



k-means clustering

k-means clustering

- Hierarchical clustering has a legacy of history: once formed, clusters cannot be changed even if it would be sensible to do so.
- k-means is an iterative procedure that produces a non-hierarchical cluster analysis.
- If algorithm started from a hierarchical cluster analysis, it will be optimised.
- Best suited with *centroid*, *group-average* or *minimum variance* linkage methods.
- Computationally difficult; cannot be sure that an optimal solution is found.

k-means clustering

- Given n objects in an m-dimensional space, find a partition into k groups (clusters) such that the objects within each cluster are more similar to one another than to objects in the other clusters.
- k-means minimises the within group sums of squares (WSS).
- k is chosen by the user; a scree plot of the WSS for $k = 1, \ldots, a$ where a is some small number.
- Even with modest n cannot evaluate all possible partitions to find one with lowest WSS.
- As such, algorithms have been developed that rearrange existing partitions and keep the new one only if it is an improvement.

k-means clustering

`$n$`	`$k$`	Number of possible partitions
15	3	2,375,101
20	4	45,232,115,901
25	8	690,223,721,118,368,580
100	5	`$10^{68}$`

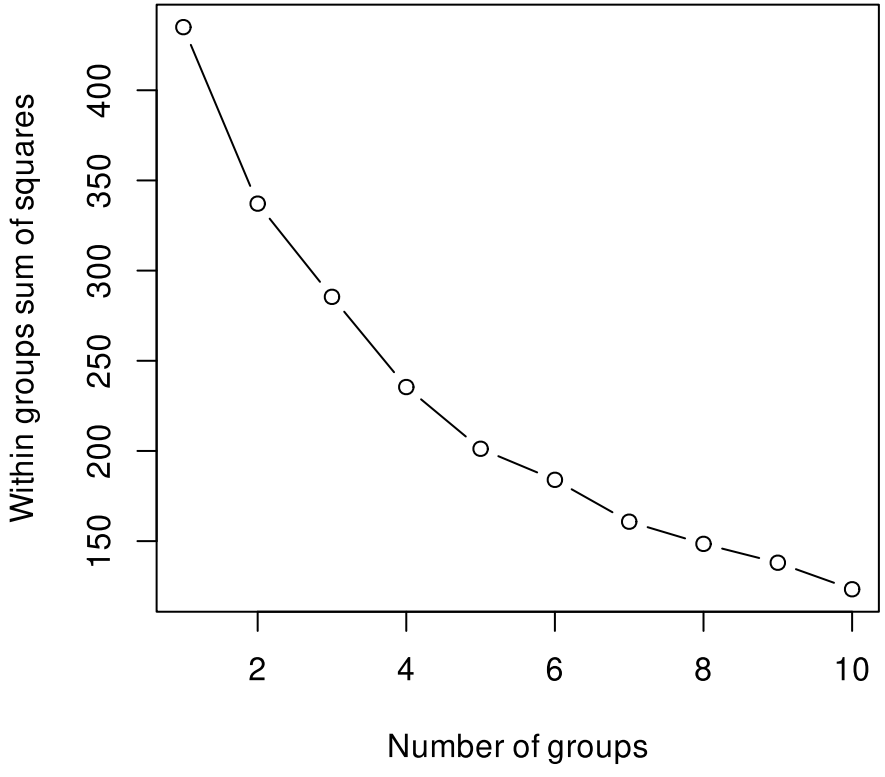
k-means clustering algorithm

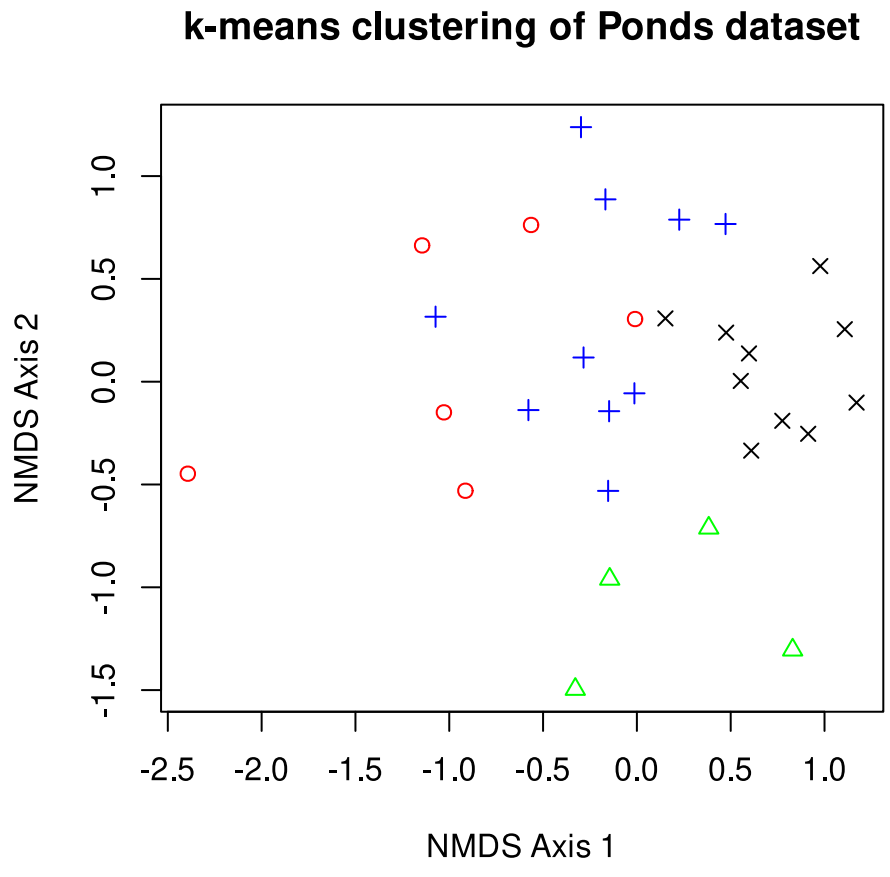
- k-means algorithm proceeds as follows:
  1. Find some initial partition of the individuals into k groups. May be provided by previous hierarchical cluster analysis,
  2. Calculate the change in the clustering criterion (e.g. WSS) by moving each individual from its current cluster to another,
  3. Make the change that leads to the greatest improvement in the value of the clustering criterion,
  4. Repeat steps 2 and 3 until no move of an individual leads to an improvement in the clustering criterion.
- If variables are on different scales, they should be standardised before applying k-means.
- Display results on an ordination; no hierarchy so no dendrogram.

k-means clustering: Ponds

- Cluster 30 shallow ponds and pools from south east UK on basis of water chemistry.
- Run k-means for $k = 1, \ldots, 10$ and collect WSS.
- No clear elbow in scree plot of WSS; change of slope at $k = 4$.
- Display results of k-means with $k = 4$ on an NMDS of the Ponds data set.

k-means clustering: Ponds





Indicator species analysis

Detection of indicator species

- A basic concept and tradition in ecology and biogeography.
- Species characteristic of e.g. particular habitat, geographical region, vegetation type.
- Adds ecological meaning to groups of sites delineated by cluster analysis.
- *Indicator species* are species that are indicative of particular groups of sites.
- *Good* indicator species should be found mostly in a single group *and* be present be present at most sites in that group.
- This is an important duality: *faithfulness* and *constancy*.

INDVAL

- *INDVAL* method of Dufrene and Legendre (1997; *Ecological Monographs* 67, 345--366) is a well respected approach for identifying indicator species.
- INDVAL can derive indicator species for any clustering of objects.
- Indicator species values based only on within-species abundances and occurrence comparisons. Values not affected by abundances of other species.
- Significance of indicator values for each species is determined via a randomisation procedure.

INDVAL

- *Specificity* is $A_{ij} = \mu_{\text{species}_{ij}} / \mu_{\text{species}_i}$.
- *Fidelity* is $B_{ij} = \mu_{\text{sites}_{ij}} / \mu_{\text{sites}_i}$.
- A_{ij} is maximum when species i is present only in group j.
- B_{ij} is maximum when species i is present in all sites within group j.
- $\text{INDVAL}_{ij} = (A_{ij} \times B_{ij} \times 100)\%$.
- Indicator value for species i is the largest value of INDVAL_{ij} observed over all j groups: $\text{INDVAL}_i = \max(\text{INDVAL}_{ij})$.
- INDVAL_i will be 100% when individuals of species i are observed at all sites belonging only to group j.
- A random reallocation of sites among groups is used to test the significance of INDVAL_i.

Interpretting clusters

Silhouette widths

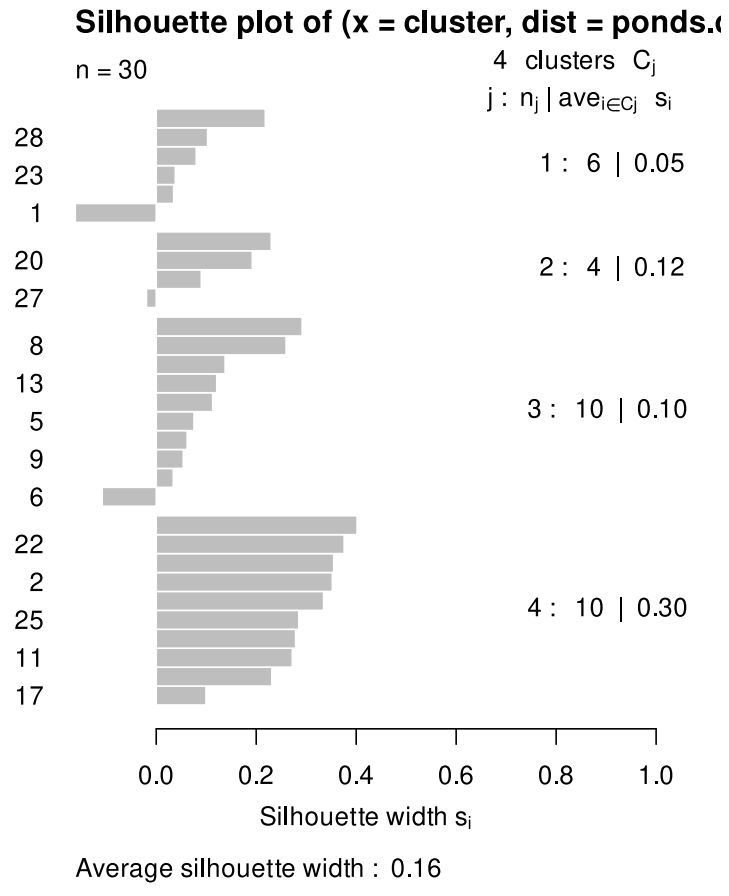
- Deciding how many clusters should be retained is problematic
- Little good theory to guide the choice
- Silhouette plots offer a simple graphical means for assessing quality of a clustering
- Silhouette width s_i of object i is

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

- a_i is average distance of object i to all other objects in same cluster
- b_i is smallest distance of object i to another cluster
- Hence, maximal value of s_i will be found where the intra-cluster distance a is much smaller that the inter-cluster distance b

Silhouette widths: Ponds k-means

- `silhouette()` in the *cluster* package
- Provide a vector of cluster memberships and the dissimilarity matrix used to cluster the data
- Returns object containing the nearest neighbouring cluster and the silhouette width s_i for each observation.
- Low values of s_i indicate observation lies between two clusters
- High values of s_i, close to 1, indicate well-clustered objects



Calinski-Harabasz criterion

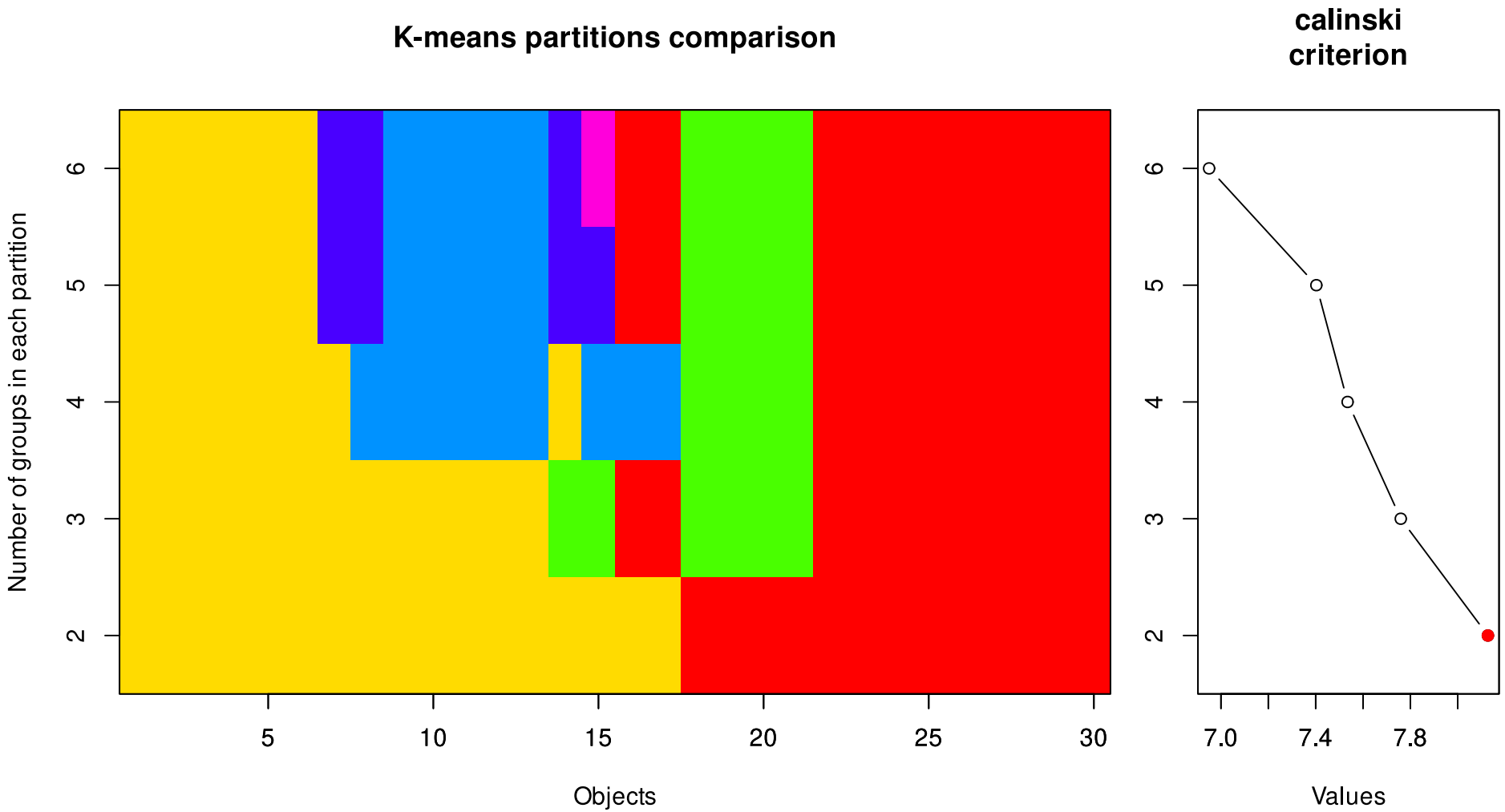
Calinski, T. and Harabasz, J. (1974) A dendrite method for cluster analysis. *Commun. Stat.* **3**: 1--27.

- The Calinski-Harabasz (1974) criterion is suggested as a simple means to determine the number of clusters to retain for analysis
- Function `cascadeKM()` in *vegan* for k-means
- Criterion computed as

$$\frac{\mathrm{SSB}/(k-1)}{\mathrm{SSW}/(n-k)}$$

- SSW \& SSB within cluster and between cluster sums of squares
- k is number of clusters, n is number of observations

Calinski-Harabasz criterion



Relating classifications to external sets of variables

- Cluster analysis is not an end in and of itself. It is a means to an end.
- Interpretation of results of cluster analysis can be aided through the use of external data, e.g. environmental data used to aid interpretation of species-derived clusters.
- Basic EDA graphical approaches, such as boxplots, coded scatterplots.
- Discriminants analysis.
- Indicator species analysis.

Diversity

Diversity

Biodiversity is many things to many people

A measure of the variety of biological life present

Perhaps taking into account the relative abundances

species diversity is a quantitative measure of the variety or number of different species

Diversity

Species	Stream1	Stream2	Stream3	Stream4
Isoperla	20.00	50.00	20.00	0
Ceratopsyche	20.00	75.00	20.00	0
Ephemerella	20.00	75.00	20.00	0
Chironomus	20.00	0.00	140.00	200
Number of species R	4.00	3.00	4.00	1
Shannon-Wiener H	1.39	1.08	0.94	0
Simpson's D	0.75	0.66	0.48	0

How shall we define "diversity"?

Diversity indices attempt to quantify

- the probability of encountering different species at random, or,
- the uncertainty or multiplicity of possible community states (i.e., the entropy of community composition), or,
- the variance in species composition, relative to a multivariate centroid

Measuring variety

Imgaine a species pool with R species

In a sample with only $R = 1$ species the sample can take on only $R = 1$ states

The abundance of that one species is 100% and all others are zero

Measuring variety

Now imagine a sample with two species $R = 2$

The community could take on $R(R - 1)$ different states

The first species could by any of the R species

The second species could be any of the *ther* species

All the other species are zero

Measuring variety

Increasing diversity means increasing the possible states that the community could take,

This increases our uncertainty about community structure

This increasing lack of information about the system is a form of *entropy*

Increasing diversity is increasing entropy

Measuring variety

Three commonly used diversity indices

- Species richness R
- Shannon-Wiener diversity
- Simpson's diversity

Species richness

The count of the number of species in a sample or area

Simple

Most commonly used

Deceptively simple; need to have common sampling effort

Shannon-Wiener diversity

$$H' = -\sum_{i=1}^{R} p_i \log_e(p_i)$$

This is the standard measure of entropy or disorder in a system

Originates from Claude Shannon's work in *information theory*

Simpson's diversity

$$S_D = 1 - \sum_{i=1}^{R} p_i^2$$

- the probability that two individuals drawn at random from a community will be different species
- the initial slope of the *species-individuals curve*
- the expected *variance* of species composition

Measuring variety

The three indices are directly related

They all estimate *entropy*, the amount of disorder or the multiplicity of possible states of a system

R depends most heavily on the *rare* species

S_D depends most heavily on the *common* species

and H' is somewhere in between

Diversity metrics

vegan has many functions for computing diversity metrics

Three popular ones are

1. Shannon-Wiener (Wrongly Shannon-Weaver)
2. Simpson
3. Inverse Simpson

$$H = -\sum_{i=1}^{S} p_i \log_b p_i$$

$$D_1 = 1 - \sum_{i=1}^{S} p_i^2$$

$$D_2 = \frac{1}{\sum_{i=1}^{S} p_i^2}$$

p_i is proportion of species i

b is the base, usually e

S is number of species (richness)

Diversity metrics

```
data(BCI)
H ← diversity(BCI)
head(H)
```

```
##        1        2        3        4        5        6
## 4.018412 3.848471 3.814060 3.976563 3.969940 3.776575
```

```
D1 ← diversity(BCI, index = "simpson")
head(D1)
```

```
##         1         2         3         4         5         6
## 0.9746293 0.9683393 0.9646078 0.9716117 0.9678267 0.9627557
```

```
D2 ← diversity(BCI, index = "invsimpson", base = 2)
head(D2)
```

```
##        1        2        3        4        5        6
## 39.41555 31.58488 28.25478 35.22577 31.08166 26.84973
```

Diversity metrics

Richness

```
head(specnumber(BCI)) # species richness
```

```
##   1   2   3   4   5   6
##  93  84  90  94 101  85
```

```
head(rowSums(BCI > 0)) # simple
```

```
##   1   2   3   4   5   6
##  93  84  90  94 101  85
```

Pielou's Evenness J

```
J ← H / log(specnumber(BCI))
head(J)
```

```
##         1         2         3         4         5         6
## 0.8865579 0.8685692 0.8476046 0.8752597 0.8602030 0.8500724
```

Diversity — Rényi entropy & Hill's numbers

Rényi's *generalized entropy*

$$H_a = \frac{1}{1-a} \log \sum_{i=1}^{S} p_i^a$$

where a is the *order* of the entropy

Corresponding Hill's numbers are

$$N_a = \exp\left(H_a\right)$$

Diversity — Rényi entropy & Hill's numbers

```
R ← renyi(BCI, scales = 2)
head(R)
```

```
##        1        2        3        4        5        6
## 3.674161 3.452678 3.341263 3.561778 3.436618 3.290256
```

```
N2 ← renyi(BCI, scales = 2, hill = TRUE)
head(N2) # inverse simpson
```

```
##        1        2        3        4        5        6
## 39.41555 31.58488 28.25478 35.22577 31.08166 26.84973
```

Diversity — Rényi entropy & Hill's numbers

```
k ← sample(nrow(BCI), 6)
R ← renyi(BCI[k,])
plot(R)
```

Partitioning diversity

Refer to biodiversity at different scales; α, β, and γ

- Alpha diversity, α, is the diversity of a point location or of a single sample
- Beta diversity, β, is the diversity due to multiple localities; β diversity is sometimes thought of as turnover in species composition among sites, or alternatively as the number of species in a region that are not observed in a sample
- Gamma diversity, γ, is the diversity of a region, or at least the diversity of all the species in a set of samples collected over a large area

Partitioning diversity

Diversity across spatial scales can be partitioned in one of two ways

1. *additive*, or,
2. *multiplicative*

partitioning

Additive partitioning

$$\bar{\alpha} + \beta = \gamma$$

$\bar{\alpha}$ is the average diversity of a samples, while γ is the diversity of the pooled samples

β is found by difference

$$\beta = \gamma - \bar{\alpha}$$

We can think of β as the average number of species not found in a sample, but which we know to be in the region

Multiplicative partitioning

$$\bar{\alpha}\beta = \gamma$$

where β is a conversion factor that describes the relative change in species composition among samples

Sometimes this type of β diversity is thought of as the number of different community types in a set of samples

Rarefaction

Species richness increases with sample size (effort)

Rarefaction gives the expected number of species rarefied from N to n individuals

$$\hat{S}_n = \sum_{i=1}^{S}(1 - q_i) \text{ where } q_i = \frac{\binom{N-x_i}{n}}{\binom{N}{n}}$$

x_i is count of species i and $\binom{N}{n}$ is a binomial coefficient — the number of ways to choose n from N

Rarefaction

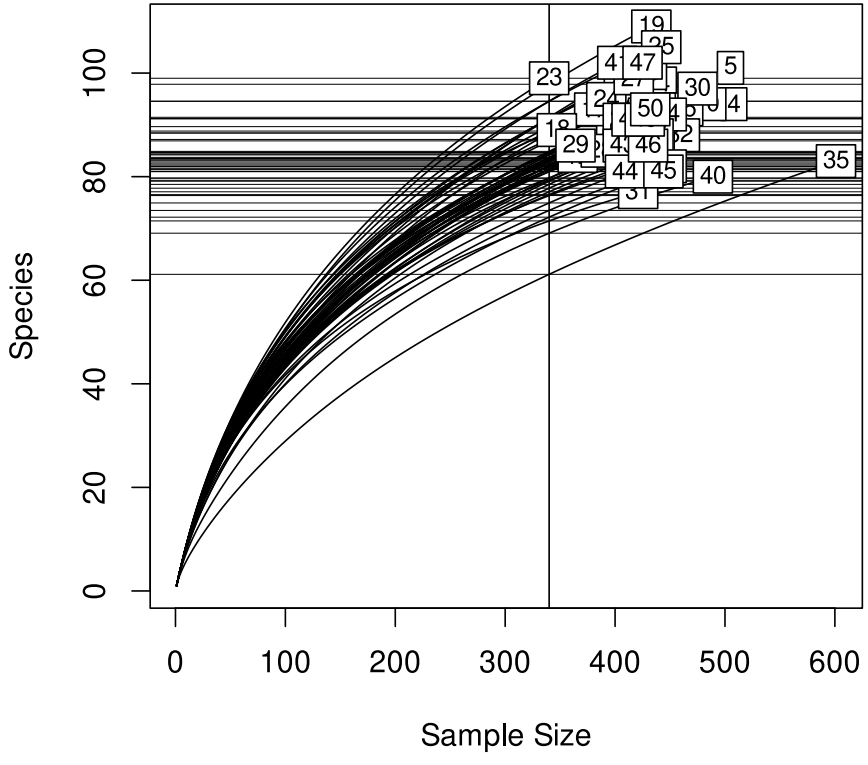
```
rs ← rowSums(BCI)
quantile(rs)
```

```
##    0%   25%   50%   75%  100%
## 340.0 409.0 428.0 443.5 601.0
```

```
Srar ← rarefy(BCI, min(rs))
head(Srar)
```

```
##        1        2        3        4        5
6
## 84.33992 76.53165 79.11504 82.46571 86.90901
78.50953
```

```
rarecurve(BCI, sample = min(rs))
```



Rarefaction

With rarefaction we can be accused of throwing away data

Yet we need to do something about the often large differences in variances of species and samples owing to sampling effort

High-throughput data

Revolutions in biology & biotechnology have lead to exponential increases in our capacity to generate data arising from the counting of biological molecules

- DNA sequencing,
- RNA-Seq — sequence RNA molecules in populations of cells or tissues,
- ChIp-Seq — sequence DNA molecules that are bound to particular proteins,
- ...

Relative cheaply, today we can generate data sets with thousands of variables on tens to hundreds of samples

High-throughput data

Counts of such molecules present a statistical challenge

- the counts typically have a large *dynamic range*, covering many orders of magnitudes
- over this very large range we observe changes in both variance (spread about the mean) and also in distribution of the data — **heterogeneity**
- like other count data, observations are integers and distributions are skewed
- the biotech methodology imparts systematic sampling biases into the data that we need to account for in an analysis — typically called **normalization** in this sub-field

Sequence depth & size factors

The number of reads for each sample is a kind of *effort* variable

All else equal, the more reads we generate the more species (OTUs, ASVs, etc) we would expect to identify

If number of reads differs between samples (libraries) then, all else equal we might assume the counts in different samples are proportional to one another, following some proportionality factor s

A simple estimate for s might be the total reads per sample

But we can do better than this ...

Sequence depth & size factors

A small dataset of 5 genes in 2 samples

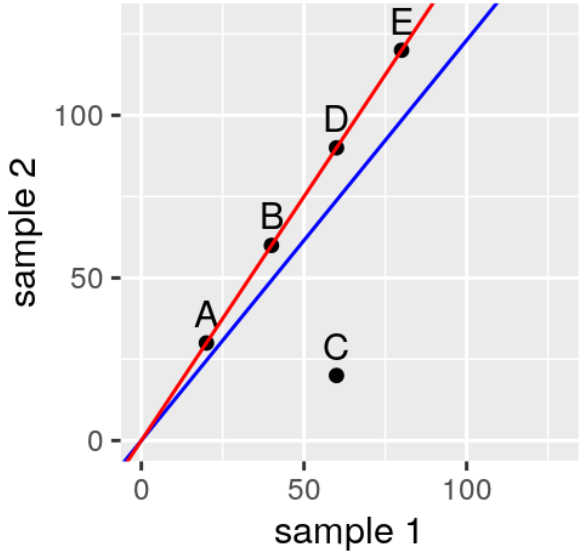
Two views:

1. estimate s_j as $\sum_{i=1}^{m} x_{ij}$, blue line is ratio of s_j
2. instead, estimate s_j such that their ratio is the red line

In 1 we would C is downregulated & A, B, D, & E are upregulated

2 is more parsimonious

`DESeq2::estimateSizeFactorsForMatrix()`



Holmes & Huber (2019) Modern Statistics for Modern Biology

Variance stabilizing transformations

The square-root transformation is known as the *variance stabilizing transformation* for the Poisson distribution

Taking the square root of observations that are distributed Poisson leads to *homoscedasticity*

Can construct variance stabilizing transformations for other distributions

High-throughput count data typically show extra-Poisson variation

`DESeq2::vst()`

Regularized log transform

This transforms counts to a $\log_2$-like scale via a simple model

$$\log_2(q_{ij}) = \sum_k x_{jk}\beta_{ik}$$

where q_{ij} are the transformed data and x_{jk} is a particular design matrix with a dummy variable for the jth sample for ($i = {1, 2, \dots m}$ variables)

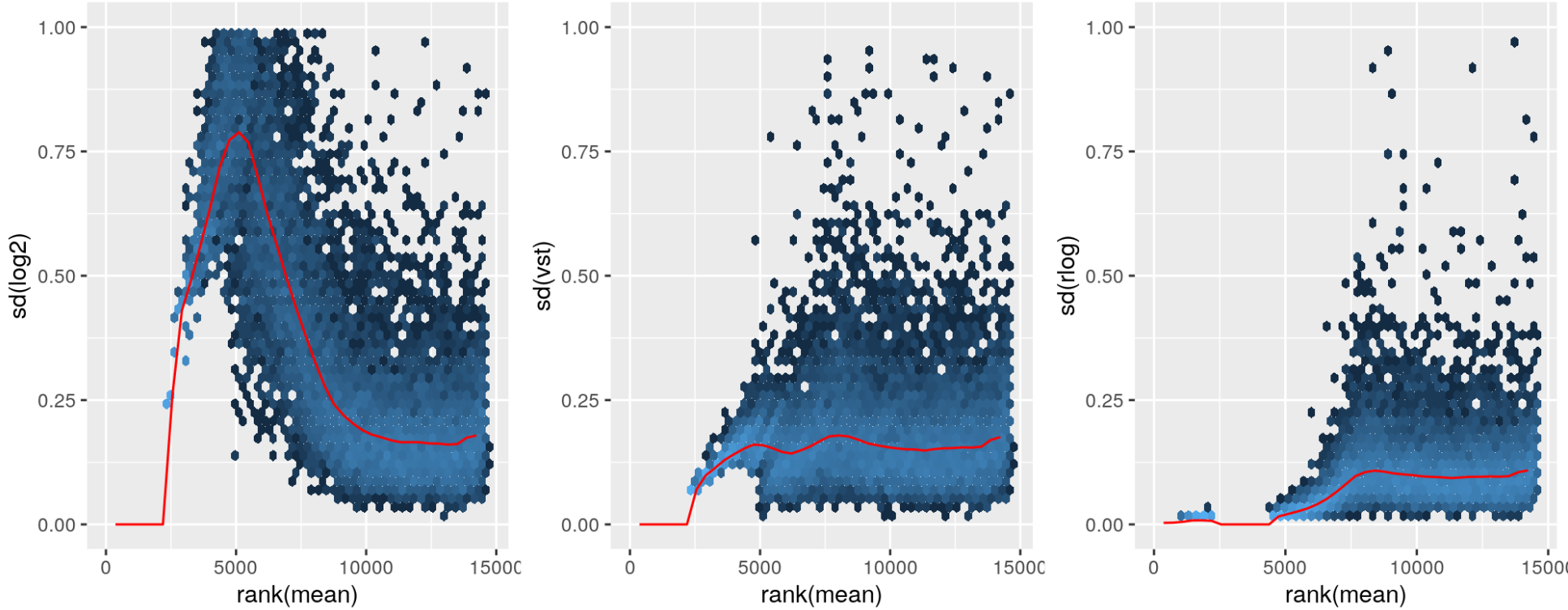
Priors on the β_{ik} make the model identifiable

The *rlog* transformation is ~ variance stabilizing, but handles highly varying size factors across samples

```
DESeq2::rlog()
```

Transformations

Comparison of three transformations



Holmes & Huber (2019) Modern Statistics for Modern Biology

Implications

Size factors can be used in GLMs etc to normalize samples via as `offset()` term in the formula

For ordination we can't easily use the size factors

Need to use *vst* or *rlog* instead

rlog creates data that can't be used in CCA, so if you use it, RDA or db-RDA are the ordination methods to use

Practicalities

Rarefied counts

Can take a random draw n' out of the N individuals in a sample in proportion to species abundances

This yields rarefied counts which some people then go on to analyse

Be careful though as this is a random process — what if you took a different random sample?

Rarefied counts

`avgdist()` in *vegan* tries to get around this by

1. taking a random draw to get rarefied counts
2. computing dissimilarity between smaples on basis of rarefied count

Repeat that many times and then as your dissimilarity d_{ij}, take the average of the many d^*_{ij} values you generated above

See also the help page for other suggestions `?avgdist`

Suggested readings

B. Everitt, S. Landau & M. Leese, 2001, Cluster analysis 4th Edition. Arnold

A.D. Gordon, 1999, Classification. Chapman & Hall

L. Kaufman & P.J. Rousseeuw, 1990, Finding groups in data. An introduction to cluster analysis. Wiley

P. Legendre & L. Legendre, 1998, Numerical ecology. Elsevier (Third English Edition)

M. H. H. Stevens, 2009, A Primer of Ecology with R. Springer