Navigation in the Ecological Space

1 – Assessing Diversity and Species Composition

Miguel Alvarez & Ildikó Orbán

GreenGaDe

About this Workshop

Timeslots

26th – 27th January 2024

- 09:00 – 12:00 Morning Session
- 13:30 – 16:30 Afternoon Session

10 – 15 min break

Trainers

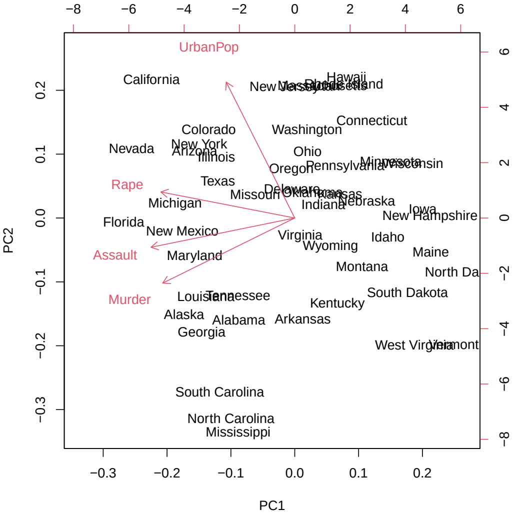
Miguel Alvarez

Ildikó Orbán

About this Workshop

Content

- Data Structure
- Diversity
- Unconstrained Ordination
  - Linear
  - Unimodal
  - Distance-Based
- Dissimilarity/Distance



GreenGaDe

About this Workshop

Disclaimer

- Licenced by Creative Commons 4.0 CC BY-SA
- It may contain a pinch of AI

About this Workshop

https://cran.r-project.org/

Why R?

- Freeware
- Scripting Language
    - Flexibility
    - Reproducibility

About this Workshop

Navigation in the Ecological Space

Welcome to our Workshop!

Plant species composition as a response to environmental factors and anthropogenic disturbance is a central principle in vegetation science and biogeography. Assessing vegetation species composition as a response to environmental factors is not trivial, as most common modelling approaches are restricted to a single response variable and not multiple responses. Multivariate statistics, including indirect (unconstrained) and direct (constrained) ordination analysis, are tools to address this problem.

https://kamapu.gitlab.io/multivar/

GreenGaDe

Introduce yourself

GreenGaDe

Expectations

Write in the Padlet Board

Data Structure

Back from the field

- Digitizing
- Data Cleaning
- Harmonizing Taxonomy
- Structuring Data
    - Relational Databases

Data Structure

Matrix served to the readers

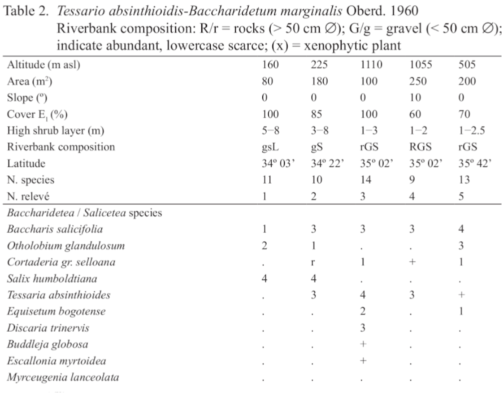


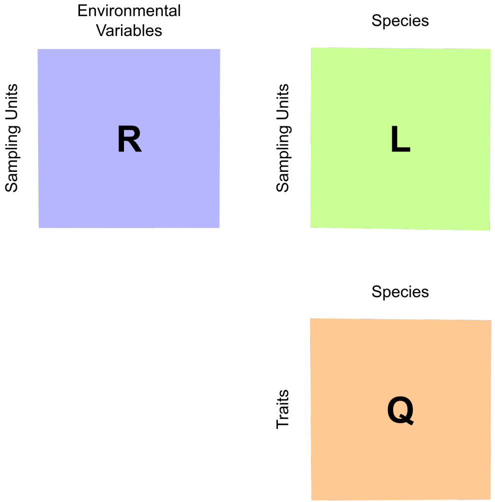
Table 2. *Tessario absinthioidis-Baccharidetum marginalis* Oberd. 1960
Riverbank composition: R/r = rocks (> 50 cm ∅); G/g = gravel (< 50 cm ∅);
indicate abundant, lowercase scarce; (x) = xenophytic plant

Altitude (m asl)	160	225	1110	1055	505
Area (m^2)	80	180	100	250	200
Slope (°)	0	0	0	10	0
Cover E_1 (%)	100	85	100	60	70
High shrub layer (m)	5–8	3–8	1–3	1–2	1–2.5
Riverbank composition	gsL	gS	rGS	RGS	rGS
Latitude	34° 03'	34° 22'	35° 02'	35° 02'	35° 42'
N. species	11	10	14	9	13
N. relevé	1	2	3	4	5
Baccharidetea / Salicetea species					
Baccharis salicifolia	1	3	3	3	4
Otholobium glandulosum	2	1	.	.	3
Cortaderia gr. selloana	.	r	1	+	1
Salix humboldtiana	4	4	.	.	.
Tessaria absinthioides	.	3	4	3	+
Equisetum bogotense	.	.	2	.	1
Discaria trinervis	.	.	3	.	.
Buddleja globosa	.	.	+	.	.
Escallonia myrtoidea	.	.	+	.	.
Myrceugenia lanceolata	.	.	.	.	.

… served to the statistical software

```
    Heliindi Panisuba Phraaust Ricicomm Pycrpoly
213      0.5       50       80      0.5      0.0
214      0.0       30       90      0.0      0.0
215      0.0       20       90      0.1     10.0
219      0.0        0        0      0.0      0.0
220      0.0        0        0      0.0      0.0
221      0.0        0        0      0.0      0.0
222     50.0        0        0      0.0      0.5
223     25.0        0        0      0.0     15.0
234      0.0        0       60      0.0      0.0
235      0.0        0       70      0.0      0.0
236      0.0        0       70      0.0      0.0
237      0.0        0       80      0.0      0.0
238      0.0        0       80      0.0      0.0
245      0.0        0        0      0.0      0.0
246      0.0        0        0      0.0      0.0
247      0.0        0        0      0.0      0.0
249      0.0        0        0      0.0      0.0
250      0.0        0        0      0.0      0.0
251      0.0        0        0      0.0      0.0
255      5.0        0        0      0.0      0.5
```

Data Structure



- **R** environmental table (header)
- **Q** species-trait table (vegetation matrix)
- **L** species composition table (species attributes)

Dolédec et al. (1996) *Environmental and Ecological Statistics*

Data Structure

Vegetation matrix

cross table

wide table

```
    Heliindi Panisuba Phraaust Ricicomm Pycrpoly
213      0.5       50       80      0.5      0.0
214      0.0       30       90      0.0      0.0
215      0.0       20       90      0.1     10.0
219      0.0        0        0      0.0      0.0
220      0.0        0        0      0.0      0.0
221      0.0        0        0      0.0      0.0
222     50.0        0        0      0.0      0.5
223     25.0        0        0      0.0     15.0
234      0.0        0       60      0.0      0.0
235      0.0        0       70      0.0      0.0
236      0.0        0       70      0.0      0.0
237      0.0        0       80      0.0      0.0
238      0.0        0       80      0.0      0.0
245      0.0        0        0      0.0      0.0
246      0.0        0        0      0.0      0.0
247      0.0        0        0      0.0      0.0
249      0.0        0        0      0.0      0.0
250      0.0        0        0      0.0      0.0
```

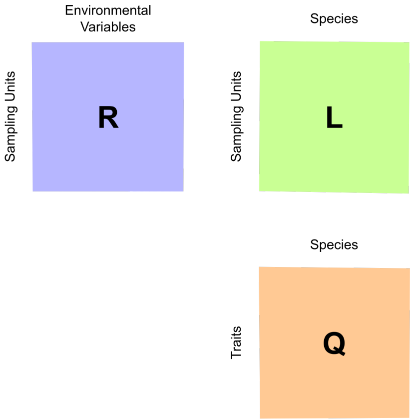
Database list

column-oriented table

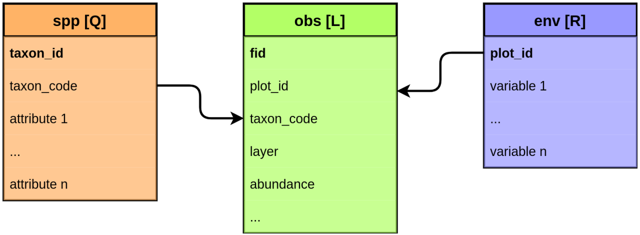
long table

```
     taxon_code taxon_id plot_id cover_perc cover_class
53     Ricicomm      473     213        0.5           1
112    Heliindi      305     213        0.5           1
689    Phraaust    51996     213       80.0           5
2780   Panisuba    50146     213       50.0           3
690    Phraaust    51996     214       90.0           5
2781   Panisuba    50146     214       30.0           3
54     Ricicomm      473     215        0.1           1
249    Pycrpoly    50067     215       10.0           2
691    Phraaust    51996     215       90.0           5
2644   Sphabull      525     215        0.5           1
2782   Panisuba    50146     215       20.0           2
1755   Cypepapy      206     219       60.0           4
2676   Cyclinte    50074     219       40.0           3
2703   Lepiowar    50075     219       50.0           3
398    Leerhexa      356     220        0.5           1
822    Typhdomi    50105     220       10.0           2
1756   Cypepapy      206     220       10.0           2
2260   Melascan      379     220        0.5           1
```

Data Structure



Relational database schema



spp [Q]
- **taxon_id**
- taxon_code
- attribute 1
- ...
- attribute n

obs [L]
- **fid**
- plot_id
- taxon_code
- layer
- abundance
- ...

env [R]
- **plot_id**
- variable 1
- ...
- variable n

- Efficient storage
  - Efficient data assessment?
- Little redundancy
- Consistency

GreenGaDe

Data Structure

Before Assessing

- Merge taxa to species
- Filter unknown species
- Reclassify abundance
    - In databases, harmonize abundance

Experience with reshaping

Write in the Padlet Board

Diversity Metrics

- Species Richness
- Shannon Index
- Evenness

Species Attributes

- Taxonomy
- Functional Traits
- Indicator values

Use of taxonomic attributes

Write in the Padlet Board

Ordination Analysis

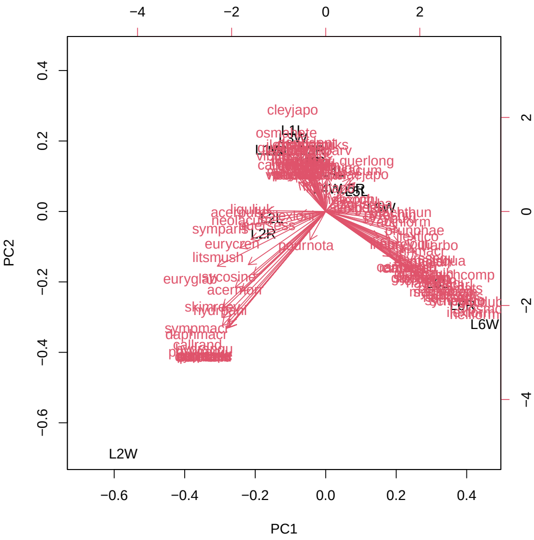
- Reduce dimensions (detect redundancy)
- Explorative assessment

Ordination Analysis

	Indirect / Unconstrained Ordination	**Direct / Constrained / Canonical Ordination**
Linear Response	Principal Component Analysis (PCA)	Redundance Analysis (RDA)
Unimodal Response	Correspondence Analysis (CA) Detrended Correspondence Analysis (DCA)	Canonical Correspondence Analysis (CCA)
Distance-Based	Principal Coordinates Analysis (PCOA) Non-Metric Multidimensional Scaling (NMDS)	Canonical Analysis on Principal Coordinates (CAP) Distance-Based Redundance Analysis (DBRDA)

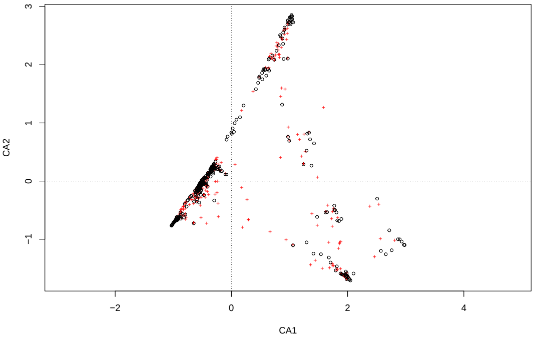
Principal Component Analysis (PCA)

```
pca_ord <- prcomp(lalashan$cross, scale. = TRUE)
biplot(pca_ord)
```



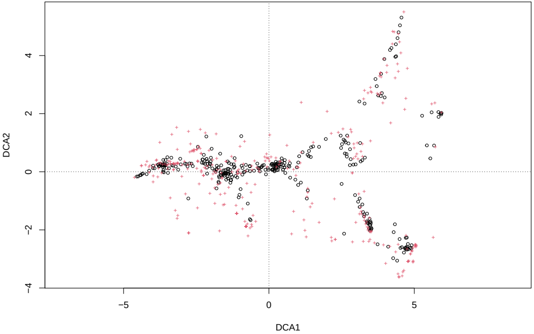
Correspondence Analysis (CA)

```
ca_ord <- cca(wetlands$cross)
ordiplot(ca_ord)
```



Detrented Correspondence Analysis (DCA)

```r
dca_ord <- decorana(wetlands$cross)
ordiplot(dca_ord)
```

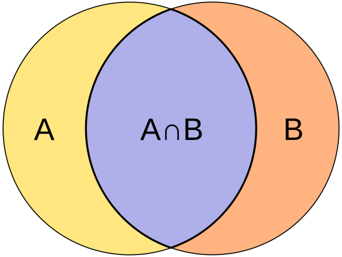


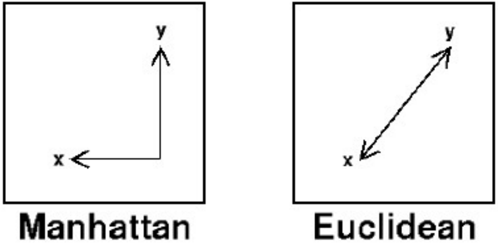
Data Pre-Processing

- Filter rare species
- Reclassify cover values
- Transformations
    - Standardize
    - Log-Transforms etc.

(Dis-)similarity and Distance Metrics

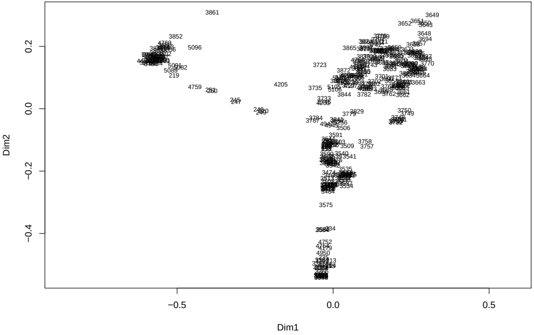
- Jaccard/Soerensen
- Bray-Curtis
- Euclidean Distance
- Manhattan Distance





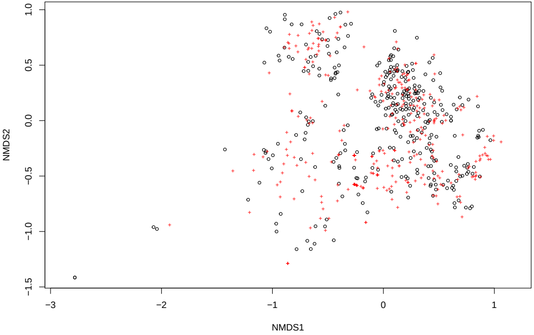
Principal Coordinates Analysis (PCoA)

```r
pcoa_ord <- wcmdscale(vegdist(wetlands$cross,
    method = "bray"), eig = TRUE)
plot(pcoa_ord)
```



Non-Metric Multidimensional Scaling (NMDS)

```
nmds_ord <- metaMDS(comm = wetlands$cross,
  distance = "bray")
plot(nmds_ord)
```



:::: {.columns} ::: {.column width=“60%”}

Objects and Attributes

Write in the Padlet Board

::: ::: {.column width=“40%”}

taxlist and vegtable

taxlist

Introduction

`taxlist` is a package designed to handle and assess taxonomic lists in R, providing an object class and functions in `S4` language. The homonymous object class `taxlist` was originally designed as a module for taxa recorded in vegetation-plot observations (see `vegtable`), but became as an independent object with the ability of contain not only lists of species but also synonymy, hierarchical taxonomy, and functional traits (attributes of taxa).

The main aim of this package is to keep consistence in taxonomic lists (a set of rules are checked by the function `validObject()`), to enable the re-arrangement of such data, and to statistically assess functional traits and other attributes, for instance taxonomy itself (function `tax2traits()` set taxonomic information as trait).

While this package only includes a function for the import of taxonomic lists from Turboveg, almost any data source can be structured as `taxlist` object, so far the information is imported into data frames in an R session and the consistency rules are respected (validity).

The use of `taxlist` is recommended for people cleaning raw data before importing it to relational databases, either in the context of taxonomic work or biodiversity assessments. The other way around, people having relational databases or clean and structured taxonomic lists may use `taxlist` as recipient of this information in R sessions in order to carry out further statistical assessments. Finally, the function `print_name()` makes `taxlist` suitable for its implementation in interactive documents using `rmarkdonw` and `knitr` (e.g. reports, manuscripts and check-lists).

The structure of `taxlist` objects is inspired on the structure of data handled by Turboveg and relational databases.

Figure: Relational model for `taxlist` objects (see Alvarez & Luebert 2018).

Installing taxlist

This package is available from the Comprehensive R Archive Network (**CRAN**) and can be directly installed within an R-session:

Links

View on CRAN

Browse source code

Report a bug

License

GPL (>= 2)

Community

Contributing guide

Code of conduct

Citation

Citing taxlist

Developers

Miguel Alvarez

Author, maintainer

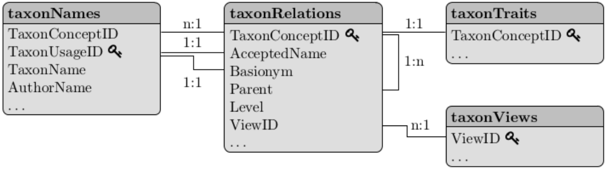
More about authors...

R-universe

r-universe 0.2.4.9002

Software Peer-Revie

Peer Reviewed



Alvarez & Luebert (2018) *Biodiversity Data Journal*

https://docs.ropensci.org/taxlist/

taxlist and vegtable

vegtable

The aim of `vegtable` is to provide a way for handling databases stored in Turboveg. This package incorporates many concepts and some functions included in the package vegdata but defining an homonymous `S4` class containing all elements of a database in just one object. The package `vegtable` also contains several methods for this object class.

Species lists in `vegtable` objects are handled by the package taxlist, thus I will recommend to take a look on it.

This package has been developed as a tool handling data stored in SWEA-Dataveg. Further development is running in the context of the project GlobE-wetlands.

An important source of inspiration for `vegtable` have been the enthusiastic discussions during several versions of the Meetings on Vegetation Databases.

Updating to the last version of vegtable

The very first step is to install the package devtools and dependencies. Then you just need to execute following commands in your R-session:

```
library(devtools)
install_github("kamapu/vegtable")
```

Some examples

The current version of `vegtable` includes an example data, which corresponds to a subset from SWEA-Dataveg. This data set contains plot observations done in Kenya imported from 5 sources.

```
library(vegtable)
#> Loading required package: taxlist
#>
#> Attaching package: 'taxlist'
#> The following objects are masked from 'package:base':
#>
#>     levels, print
#>
#> Attaching package: 'vegtable'
#> The following object is masked from 'package:base':
#>
#>     transform
data(Kenya_veg)
```

Links

View on CRAN

Browse source code

Report a bug

License

GPL (>= 2)

Citation

Citing vegtable

Developers

Miguel Alvarez

Author, maintainer

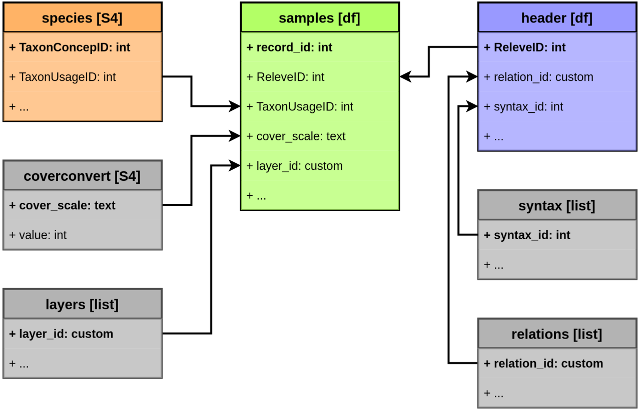
Dev status

- Encapsulation of data processing
- Portability
- Repeatability
- Collaborative data assessment

http://kamapu.github.io/vegtable/

https://kamapu.github.io/posts/vegtable-press-2/

taxlist and vegtable



species [S4]

+ TaxonConcepID: int

+ TaxonUsageID: int

+ ...

coverconvert [S4]

+ cover_scale: text

+ value: int

layers [list]

+ layer_id: custom

+ ...

samples [df]

+ record_id: int

+ ReleveID: int

+ TaxonUsageID: int

+ cover_scale: text

+ layer_id: custom

+ ...

header [df]

+ ReleveID: int

+ relation_id: custom

+ syntax_id: int

+ ...

syntax [list]

+ syntax_id: int

+ ...

relations [list]

+ relation_id: custom

+ ...

GreenGaDe

Retrospective

Thank You!

GreenGaDe