# Managing Vegetation-Plot Data in R

IAVS-LAC

# Contents

# Introduction to the Course

How the course is organized.

Focus in own packages (mention that there are other options).

# Chapter 1

# Managing Vegetation Data in R

While in this book we assume a background knowledge on statistical programming in **R**, we will refresh some basics on data and objects. We will also discuss about the proper arrangement of tabulated data and organizing tables as in relational databases. Besides the `data.frame`, which is the queen of data sets in R, there are also other ways to arrange data, such as `list`, `tribble`, and ad-hoc classes designed in **S4** (see Genolini, 2008).

## 1.1 Data in R

The basic structure in **R** to represent data is the vector, which corresponds to one or many connected values. A property of the vector is its length, namely how many elements are connected in the vector and can be retrieved by the function `length()`. The other important attribute of a vector is its class, which corresponds to the nature of the values of the vector, for instance, `numeric`, `integer`, `character`, etc.

***Here a table with basic classes***

The access to parts of the vector can be done using squared brackets and including a value or a vector inside of it, for instance an `integer` or a `logical` vector are suitable for it.

```
class(letters)
```

```
## [1] "character"
```

```
# Fifth letter in alphabet
letters[5]
```

```
## [1] "e"
```

```
# Using logical vector
letters[letters %in% c("x", "b", "f")]
```

```
## [1] "b" "f" "x"
```

Special attention is needed for the class `factor`, which is useful to represent categorical variables in R and may not be confused with `character`.

***Here a mention about coercion***

***Discussing time and Date***

***Table with time and Date classes***

***Mention on symbols,expressions and formulas***

## 1.2 Objects in R

Vectors can be elements of more complex objects in R, for instance arranging vectors in columns and rows result in a `matrix`. For the access to parts of the matrix you can use squared brackets with two indices, the first for the rows and the second for the columns.

```
M <- matrix(data = 1:20, nrow = 5, byrow = FALSE)
M
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    6   11   16
## [2,]    2    7   12   17
## [3,]    3    8   13   18
## [4,]    4    9   14   19
## [5,]    5   10   15   20
```

```
# Second row
M[2,]
```

```
## [1]  2  7 12 17
```

```
# Third column
M[,3]
```

```
## [1] 11 12 13 14 15
```

```
# Crossing both
M[2,3]
```

```
## [1] 12
```

Different objects can be grouped into a `list` of objects. It is even possible to contain lists into lists resembling hierarchical trees. Elements of a list can be accessed by using double squared brackets. If the elements are named, you can also index by using the dollar symbol (`$`).

***data.frame***

***tribbles***

***spatial objects (sf)***

***S3, S4, R6***

## 1.3 Functions and Methods

Functions are, besides the vectors, among the most important object class in R. Functions are defined in the form of `foo <- function() {definition}`. In the definition of functions you will find arguments sometimes with pre-set default values. Arguments are also known as parameters, while their values can be called arguments, as well.

```
add_two <- function(x) {
  x <- x + 2
  return(x)
}
class(add_two)
```

```
## [1] "function"
```

```
add_two(x = 10)
```

```
## [1] 12
```

*Hint?*

In the context of object oriented programming, a function can behave differently according to the class of one or more input objects. Take for instance the function `summary()`, which may do different thinks if **`object`** is `numeric`, `factor`, or `data.frame`.

```
# Applied to numeric vector
summary(iris$Sepal.Length)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.300   5.100   5.800   5.843   6.400   7.900
```

```r
# Applied to factor
summary(iris$Species)
```

```
##     setosa versicolor  virginica 
##         50         50         50
```

```r
# Applied to data.frame
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width   
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100  
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300  
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300  
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199  
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800  
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500  
##        Species  
##  setosa    :50  
##  versicolor:50  
##  virginica :50  
##                 
##                 
## 
```

Usually these alternatives correspond to different function definitions and are called **methods**. In **S3** the method depends on the class of the first argument, while in **S4** the method can be determined by more than one argument.

# Chapter 2

# Taxonomic Lists

Vegetation data and any kind of diversity information will preferably focus on the species rank. Besides having a clean list of species, depending on the aims of the respective research project, additional information regarding higher taxonomic ranks or attributes of recorded taxa (i.e. life forms, chorology, functional traits) may need to be retrieved and properly stored for further statistical assessments.

## 2.1 Taxonomic Names Resolution

One of the most important steps previous to statistical assessments is resolving names. The main causes of systematic errors are:

- Typos
- Differing Taxonomic References

In such cases records of the same species will be accounted as different ones by the computer.

Matching names with different tools

- taxize
- Taxonstand
- WorldFlora

### 2.1.1 Resolving names with taxize

Introduction to the package `taxize` (Chamberlain and Szöcs, 2013).

Example.

### 2.1.2 Resolving names with WorldFlora

Introduction to the package `WorldFlora` (Kindt, 2020).

### 2.1.3 Other Alternative Resolutions

Package `taxadb` (Norman et al., 2020).

Package `Taxonstand` (Cayuela et al., 2012).

Package `taxizedb`.

Package `taxize`.

Package `TNRS`.

## 2.2 Structuring Taxonomic Lists with taxlist

Taxlist model (Alvarez and Luebert, 2018).

Matching names with taxlist (intro to vegStore???)

- Step by step
- From data frame
- Comparing with existing species list

## 2.3 Updating Taxonomic Lists

Adding/reseting taxonomic ranks.

Adding synonyms, changing accepted names, merging taxa.

**Potential citations**

(Boyle et al., 2013; Maitner et al., 2017)

# Chapter 3

# Vegetation-plot Data Sets

## 3.1 Data Formats

Again cross table and other data structures

Different data formats (reading files, connecting databases, ...)

Note content of last chapter needs to be moved here.

## 3.2 Data Sets in vegtable

Vegtable concept

Structure of S4 class

## 3.3 Importing and Exporting Vegetation-Plots

Workflow for import

To spreadsheet

To juice

As rda and rds

# Chapter 4

# Using Taxonomic Information

Both, the packages `taxlist` and `vegtable` offer functions to produce information on the basis of taxonomies and attributes, provided that such information is already available in the input data.

## 4.1 Overview of Taxonomic Ranks

Summaries and levels.

Indented lists.

## 4.2 Counting Taxa

Remind the summaries.

Counting all taxa or specific taxonomic ranks.

Counting only present taxa.

## 4.3 Further Taxonomic Attributes

Life forms

Functional traits

Statistics of taxonomic attributes (chorology, life forms, functional traits)

Researching for functional traits in online databases?

# Chapter 5

# Statistics at the Plot Level

Focus on the use of slot **header** to collect information of each plot.

Refreshing statistics in **R**.

Mapping plots

## 5.1 Environmental Information

For instance soil layer information.

Introduction to descriptive statistics

## 5.2 Species Abundance in Plot

Discussion about cover scales.

Objects `coverconvert`.

Cover transformations

Cover reclassification.

## 5.3 Taxonomic Attributes to Plots

- Count taxa
- Trait Proportions
- Trait Statistics

## 5.4 Aggregating Plots

Introduction to the slot **relations**.

Use of the slot **relations** to aggregate plots in groups.

All these summaries applied to clusters (groups of plots).

# Chapter 6

# Data Analysis and Visualization

Those topics in the base of comparisons at the plot level.

## 6.1 Data Exploration

Introduce to a suggested workflow on the basis of `vegtable`.

Some Stats

## 6.2 Hypothesis Tests and Regression Models

Some tests as examples.

conparing groups. regressions among environmental data.

## 6.3 Multivariate Statistics

### 6.3.1 Classification Analysis

### 6.3.2 Ordination Analysis

## 6.4 Maps

- On Graphic devices
- Exporting

# Chapter 7

# Reproducible Assessment with Rmarkdown

Introduction to Rmarkdown

## 7.1 Using Rmarkdown as Notebook

`output: html_notebook` as default format for this course

- Code and console output using chunks

## 7.2 Inline code and scientific names

How to insert inlince code and use `print_name()` from `taxlist`.

## 7.3 Graphics and Tables

- Inserting Graphic and tables
- using captions and cross-links
- referencing chapters

## 7.4 Bibliographic references

- Bibliographic references

## 7.5 Other Output Format

- PDF document
- docx document
- Beamer

## 7.6 Templates

Intro to package `vegTemplates`.

# Bibliography


Alvarez, M. and Luebert, F. (2018). The taxlist package: managing plant taxonomic lists in R. 6:e23635.

Boyle, B., Hopkins, N., Lu, Z., Garay, J. A. R., Mozzherin, D., Rees, T., Matasci, N., Narro, M. L., Piel, W. H., Mckay, S. J., Lowry, S., Freeland, C., Peet, R. K., and Enquist, B. J. (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. 14(1):16.

Cayuela, L., Granzow-de la Cerda, Î., Albuquerque, F. S., and Golicher, D. J. (2012). Taxonstand: an R package for species names standardisation in vegetation databases. 3:1078–1083.

Chamberlain, S. A. and Szöcs, E. (2013). Taxize: taxonomic search and retrieval in r. 2:art 191–.

Genolini, C. (2008). A (not so) short introduction to s4.

Kindt, R. (2020). WorldFlora: An r package for exact and fuzzy matching of plant names against the world flora online taxonomic backbone data. 8(9).

Maitner, B. S., Boyle, B., Casler, N., Condit, R., Donoghue II, J., Durán, S. M., Guaderrama, D., Hinchliff, C. E., Jörgensen, P. M., Kraft, N. J. B., McGill, B., Merow, C., Morueta-Holme, N., Peet, R. K., Sandel, B., Schildhauer, M., Smith, S. A., Svenning, J.-C., Thiers, B., Violle, C., Wiser, S., and Enquist, B. J. (2017). The bien r package: A. tool to access the botanical information and ecology network (bien) database.

Norman, K. E. A., Chamberlain, S., and Boettiger, C. (2020). taxadb: A high-performance local taxonomic database interface. 11(9):1153–1159.