Storage and Analysis of Vegetation Data:
a crash course

Miguel Alvarez*[1] and Kai Behn†[1]

[1]Vegetation Ecology, INRES, University of Bonn, Germany

18th–19th November 2015

Contents

1 Introduction to the Workshop **2**

2 The common structure of vegetation data **2**
2.1 The phytosociological relevé . . . 2
2.2 Cross tables and column oriented tables . . . 2
2.3 Exercises . . . 2

3 Relational databases **2**
3.1 Basics on relational databases . . . 2
3.2 Importing Tables to Base and Building Relationships . . . 3
3.3 Query building . . . 3
3.4 Exercises . . . 3

4 Introduction to Juice **4**
4.1 Starting in Juice . . . 4
4.2 Sorting Plots and Species . . . 6
4.3 Inserting Separators . . . 6
4.4 Building Synoptic Tables . . . 6
4.5 Exercises . . . 6

*malvarez@uni-bonn.de
†kaibehn@uni-bonn.de

1 Introduction to the Workshop

This workshop attempts to combine theoretical introductions with interactive sessions. The afternoon sessions are reserved for exercises and coaching and is the opportunity for participants to solve personal challenges.

Every participant receives a flash disk including example data, documents and software required for the sessions. Distributed software can be executed from the disk, thus no installations will be required.

2 The common structure of vegetation data

2.1 The phytosociological relevé

The structure of data always depends on the information contained and the sampling strategy used to collect it. One of the most common ways to collect information regarding the floristic composition and other structural patters of vegetation is the **phytosociological relevé**.

Basically, a phytosociological relevé is the record of abundance of species in plots. Additional information on the plots are usually collected during a survey (e.g. slope, aspect) or results from further sample analyses (e.g. soil chemical and physical properties). Moreover, surveys done in woody stands also split records of species abundance by layers (e.g. herb, shrub and tree layers).

The degree of complexity of such records makes almost impossible to store all information in just a single table, thus multiple table information models are more convenient. Here we will give a short introduction on relational databases applied to the storage of vegetation data.

2.2 Cross tables and column oriented tables

The cross table is the most common way used to display information in publications and other documents. In fact, many software dealing with the analysis of vegetation data use cross tables as input format. Therefore we have the tendency to store rough data in cross table format, but is this really the best way to store vegetation data?

Usually it is more convenient to select **column oriented tables** (also called **database lists**) for the storage of data. The database lists are usually more economical than cross tables for the storage of vegetation data, since the later usually contain a lot of empty cells. Database lists are also the most common input format for statistical packages and it is easier to produce cross tables from data base lists than the opposite way.

For creating a cross table in **Calc**, you have to use the function **pivot table** (the same as in Microsoft Excel):

1. Select the whole table, for it press `control + shift + end`.
2. Go to menu `Data -> Pivot Table -> Create...`.
3. Use the default `Current selection` and click `OK`.
4. Drag and drop the fields you like to display in the pivot table.
5. In the button `Options...` you can set the function used to calculate the `Data Fields`.
6. In the button `More`, you can expand the options. Here select in `Results to`, `- new sheet -`.

2.3 Exercises

- Among the data distributed in the course there is a file called. **Germination.ods**. Produce a cross table with the average germination for every treatment.
- Test this function with own data.

3 Relational databases

3.1 Basics on relational databases

As previously discussed, it is difficult to organize vegetation data in a single table. Such information can be more efficiently stored in multiple tables, each of them as **column oriented tables**. While each column in a table represent one **variable**, each row is a **record** in the same table. One of those column may be a **key**, which represents the identity of the record and therefore does not allow the occurrence of duplicates.

In one database the different tables may share at least one column with another table in the same data set. This link is called **relationship**.

The interaction with a relational database is usually done by **SQL** (Structured Query Language).

3.2 Importing Tables to Base and Building Relationships

For the import of tables in `Base`, you can copy and paste from `Calc`:

- Open `Base` and create a new, empty database.
- Open the vegetation data set in `Calc` and select the sheet to be imported.
- Select all the data in the sheet (put the cursor at the top left corner and press `control + shift + end`).
- Paste the table in `Base` (you can press `control + v`). Then a wizard will be opened.
- Name the table and select the columns.
- In the next step `Type formatting` you can define the fields you import. It is important to choose or create a **primary key**.
- Further define the types of every field. Data fields must be set as a numeric type (e.g. Decimal, Integer, Small Integer, etc.) in order to to allow mathematical calculations with them.

After importing all required tables (see the distributed template), you have to build the relationships between them.

- Go to the menu `Tools -> Relationships`
- Add the tables in the view, one by one.
- To build a new relation between two tables, click `Insert` and then `New Relation`.
- Select under `Tables involved` two tables, between which you want to build the relation and under `Fields involved` the two fields.
- Important to consider is that the definition of two columns connected by a relationship have to be identical.

3.3 Query building

Both, manage and querying databases can be done by using **SQL** statements (Structured Query Language). Fortunately, **Base** has some user friendly applications supporting the building of queries. In this case, we will focus on the use of the `Design View`.

- In the database area (left side), select `Queries`.
- Under `Tasks` click on `Create Query in Design View...`
- For a better overview, add to the design the tables containing the information you like to include in the query. By this way you get an overview of the relationships.
- By drag and drop you can select the respective fields in the lower table.
- Complete the lower table with additional options for the query.
- Push on button `Run Query`.

3.4 Exercises

- In how many plots does *Andropogon abyssinicus* occur?
- Find all plots containing Chamaephytes. Which is their maximum cover in these plots?
- Which annuals (therophytes) occur in the communities of *Andropogon pratensis* and the communities of *Pennisetum schimperi-Commelina africana*?
- Which plant communities are recorded in the locality **Zegamel**?
- Make a table of all families and their sum of the cover values (in percent) for the communities of *Andropogon pratensis*. Transform the cover into percentage cover.
- Make a table of lifeforms and their sum of cover values (life form spectrum) for the communities of *Fimbristylis complanata*.
- Build a database with own data and apply similar exercises.

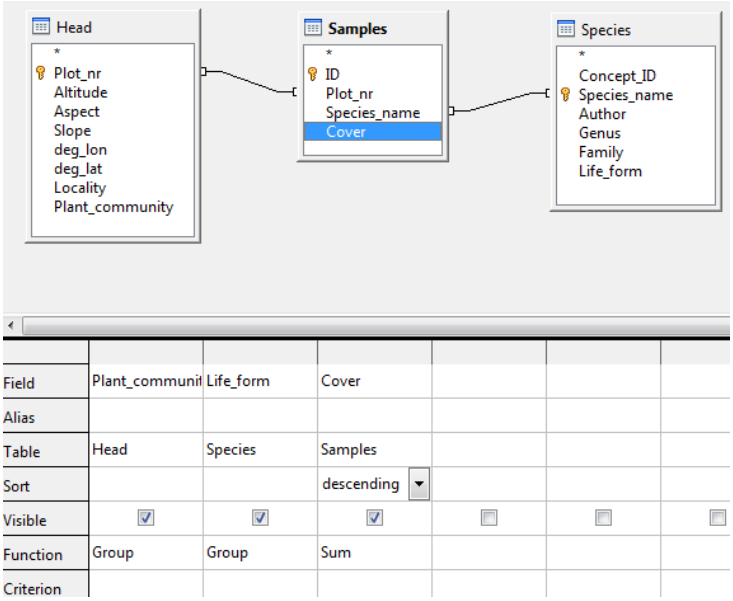

Figure 1: Query building in Base using the Design View.

```
SELECT "Head"."Plant_community", "Species"."Life_form",
SUM( "Samples"."Cover" ) FROM "Samples", "Head", "Species" WHERE
"Samples"."Plot_nr" = "Head"."Plot_nr" AND "Samples"."Species_name"
= "Species"."Species_name" GROUP BY "Head"."Plant_community",
"Species"."Life_form" ORDER BY SUM( "Samples"."Cover" ) DESC
```

Figure 2: The same query as SQL statement.

4 Introduction to Juice

4.1 Starting in Juice

The software Juice has been developed at the Marsaryk University (Czech Republic) and has as main tasks the arrangement of vegetation tables, especially those containing a big amount of plots and species. Additionally, this package implements some statistical analyses and can be also connected to other software (e.g. Canoco, PC-Ord and R) working as interface.

Besides its popularity among vegetation ecologists and the implementation of new statistical approaches in this software, Juice is also a freeware (download site http://www.sci.muni.cz/botany/juice).

For importing data to **Juice**, we will first save the required files in **CSV** format (Comma Separated Values). The tables required are in the sheets **Vegetation** and **Head**:

1. In Calc, open the sheet you like to convert to CSV. Remember, a CSV file can only contain one sheet of a book.
2. In menu follow `File -> Save as...`
3. In the checkbox `Save as type:` select `Text CSV`.
4. You will some warning messages. To the first click on `Save current format` and to the others just `OK`.
5. Then close your book.

Before importing the tables in Juice, some slight modifications are required:

1. In your Windows System, open the **Notepad**.
2. In the menu `File -> Open...` select the CSV file containing the vegetation table. Remember to select in the bottom right corner `All Files (*.*)`.
3. Insert two empty rows at the begining of the table.
4. In the first row, write a name for the data set.
5. Delete the first column name in the table.
6. Save and close.

For the head table you have also to open it in **Notepad** and then replace the name of the first column by **Releve number**.

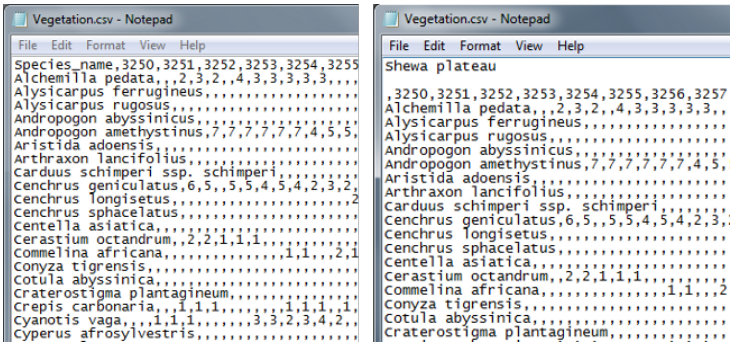

Figure 3: Vegetation table exported from Calc (left) and modified for its import in Juice (right).

Now you may open **Juice** and follow those steps:

1. In menu, `File -> Import -> Table -> from Spreadsheet File` and select the CSV file containing the vegetation table
2. A wizard will support you in the import process step by step
3. Set `Character delimiting columns` as **Comma**
4. In `Layer information` unmark the checkbox
5. In `Cover values` you may set it as `User Defined`. Double click in the respective entries and set the values according to the sheet cover in the Calc book.
6. Click `Finish`

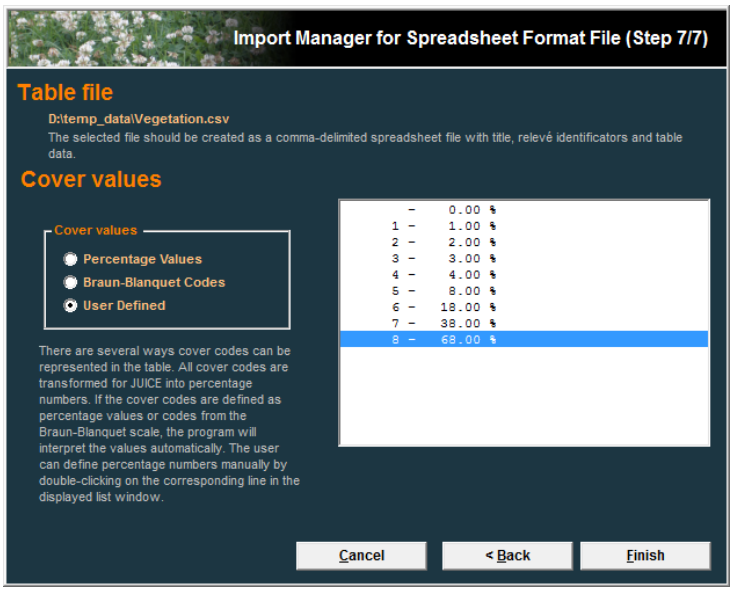

Figure 4: Conversion table for cover values in Juice.

After a successful import, you should directly save the table in menu `File -> Save`. The next step is the import of the head data:

1. Follow the menu `File -> Import -> Header Data -> From Comma Delimited File`, and select the CSV file with the head data
2. Select the CSV file containig the head data and click `Open`

To check you can switch the view from table to extended head.

4.2 Sorting Plots and Species

Sorting species and plots in vegetation tables is a straightforward function in **Juice**. The most easy way is by drag and drop.

You can also select species and plots (relevés) using different colors. For it you select a color in the lower menu and select by using the right mouse click at the margin of the table. You can also select several items by drag with the right mouse key. The arrows in the color menu allow you to group all species or all plots sharing the same color.

4.3 Inserting Separators

Separators are lines inserted to delimit plot groups. You have to go with the cursor to the upper margin of the table and use right mouse click while pressing `shift`.

4.4 Building Synoptic Tables

Synoptic tables are summaries of vegetation tables showing constancy, fidelity or average cover values from species in groups o plots. Those groups are usually plant community types resulting from a classification analysis.

- To build the synoptic tables in Juice, you have first to sort your table and put separators between groups.
- Set in the menu `File -> Options -> Synoptic Tables` the thresholds you like to use for the display in the synoptic table.
- Push `OK`.
- Go to menu `Synoptic Table` and select an alternative of your preference.
- In menu `Synoptic Table -> Threshold Values` you can tune thresholds for fidelity, frequency and cover.

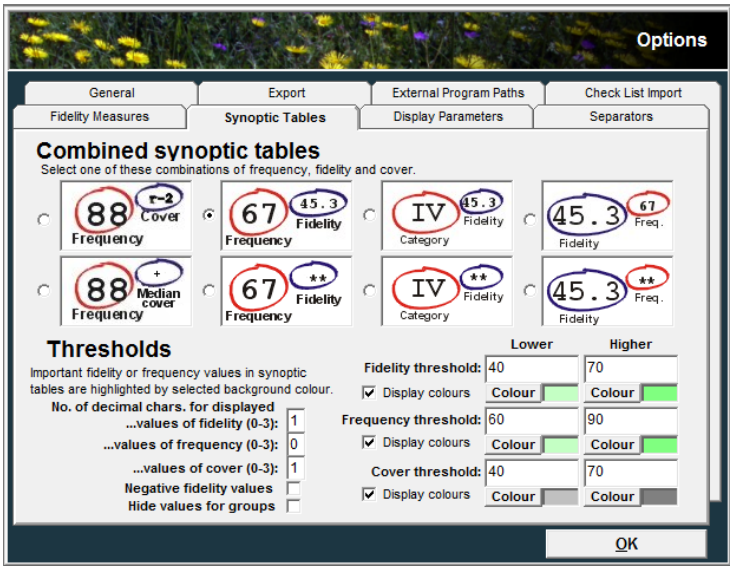

Figure 5: Setting of threshold values for synoptic tables in Juice.

4.5 Exercises

- Identify the species with the highest percentage frequency in every plant community.
- Make separators between localities and create a synoptic table of categorical frequency.