Exploring data of Native Species and Cultivated Native Species in Santa Catarina

Júlia Kuse Taboada

August 18th 2022

Abstract

In this exercise, I tried to understand differences in taxonomic diversity and trait diversity between native cultivated species and the whole range of native species in the state of Santa Catarina. For that, it was used the package BIEN to access data of occurrence and traits. The diversity indices used were Shannon and inverse Simpson, and also Rényi diversity profiles. The trait that was explored was the whole plant growth form.

Introduction

It is known the importance of biodiversity for human life on planet, since it provides and regulates ecosystem services. We've been facing a biodiversity crisis, specially due to human activities that enhance biodiversity loss and climate change. Thus, it is necessary to assess information about biodiversity, the factors that influence it and the services it provides (Nock, Vogt, & Beisner (2016)) . There are several ways to measure plant diversity, being the main ones taxonomic, functional and phylogenetic. The main goal of this exercise was to explore data of species occurrence and their traits, and try to understand if there are differences in diversity when comparing the categories of native species with native species that are cultivated in the state of Santa Catarina. The main hypothesis is that the subset of "All species" should present higher richness than the subset of "Cultivated species", since we are analyzing native species and it is likely that the "Cultivated" category is a subset of the "All Species" category. It is also expected that "Cultivated" species present a higher concentration in a few trait values, simply because if they are cultivated they are being chosen by humans, and this could be related with some trait of interest.

Methodology

It was used data from BIEN database (Maitner (2022)), and it was assessed with functions from the BIEN package in R. It was generated two data frames with the cultivated and non cultivated native species' occurrence. `library(BIEN) BIEN_occurrence_state(country = "Brazil", state = "Santa Catarina")` and `BIEN_occurrence_state(country = "Brazil", state = "Santa Catarina", cultivated = TRUE)`. For the cultivated species, it was removed the rows with NA value in the column "Is_cultivated_observation". After that, the occurrence data frame was transformed in an abundance matrix for each category. So it was possible to sort the species and find the most abundant species in each category. It was created a table with the 50 most abundant species for "All species" and "Cultivated". `library (dplyr)   table_abund_sp <- sp_sc %>% select(scrubbed_species_binomial) %>% table ()` The same code was run for the "Cultivated". It was also created a data frame with the species and abundance of both categories, so it was possible to see species that were in one subset but not in the other. The code to create the data frame was: `total_abund_50 <- dplyr::full_join(sorted_50_all, sorted_50_cult, by = ".")`. Then, it was created two vectors, one with the abundance of "All species" and other with the abundance of "Cultivated". They were used to calculate Shannon, inverted Simpson and Rènyi diversity. They were calculated using function from the package vegan (Oksanen et al. (2020)).

The Rényi values from both categories was joined in one single data frame, so it was possible to plot then together. The Rényi was also calculated for the whole set of species, and it was plotted with the following code.

```
matplot(renyi_bind_sp, type = "l", axes = F, main = "Renyi Diversity")
box()
axis(side = 2)
axis(side = 1)
legend("topright",
    legend = c("All species", "Cultivated species"),
    lty = c(1,2),
    col = c(1,2))
```

With the abundance tables, it was created a vector containing the names of the most abundant species. It was used another BIEN function, to request information of the functional traits of those species: `BIEN_trait_species()`. Inside the parenthesis of the function it was passed the vector with the names of the 50 most abundant species. The data of traits was cleaned to have only information from the name of species, trait name and trait value. The trait that was chosen to be analyzed was "Whole plant growth form", that was filtered using the code: `gf_all <- dplyr::filter(trait_all, trait_name %in% "whole plant growth form")`. This trait data was cleaned, because there were a few values written in different ways but meaning the same thing, like "tree" and "Tree". They were fixed using: `gf_all$trait_value[gf_all$trait_value =="tree"] <- "Tree"`. A bar plot showing the most common traits for each category was plotted with the following code:

```
ggplot(data = gf_all,
    mapping = aes(x = trait_value)) +
ggtitle("Growth form for 50 most abundant species in general") +
geom_bar(col="black",
        fill = "darkgreen") +
xlab("Trait Value") +
ylab("Specie") +
theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

Results from expolatory data analysis and discussion

In Shannon diversity index, the more species you have in a given subset and the more even is their abundance, higher the result (Gotelli & Ellison (2004)). The results obtained when comparing "All species" and "Cultivated" were not that different, although "All species" had a slight higher result. The Simpson diversity index represent the chance of, in a given subset, two randomly picked individuals belong to the same specie. In Simpson index, the lower the result, higher the diversity (Gotelli & Ellison (2004)), and that's why it is usually represented as it's inverse. The results obtained were not that different as well, with "All species" presenting a slightly higher value. This could be explained by the fact that we are analyzing cultivated native species and native species in general in the whole sate of Santa Catarina, without considering different plots in the state, so it is very likely that the "Cultivated" category is representing a subset of the "All species" and not showing much difference. The Rényi diversity profiles were also calculated, and it is better to visualize the results with the plots:

```
knitr::include_graphics(c("renyi_div_50_abund.png", "renyi_div.png"))
```

In the profile, a horizontal shape of the line indicates that all species have the same evenness. Also, the starting point in the left side of the profile is an indicator of richness, so, profiles that start in a higher level in the y axis have a higher richness. Another important thing is that, if the profile of one site is totally
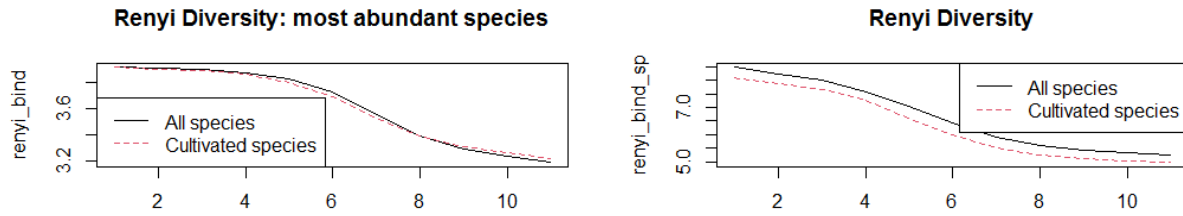
Figure 1: Rényi diversity for the 50 most abundant species(left) and for the whole set of species, considering both categories (all species and cultivated) (right).

above the profile of another site, that means that the site with the highest profile is more diverse when compared with the other. If the profiles intersect at some point, it's not possible to rank the sites from lower to higher diversity, because one can have higher richness but lower evenness and so on (Kindt & Coe (2005)) . What we can see in the profiles is that, when comparing the subset of the 50 most abundant species of each category, it is not possible to tell which is the most diverse. But when comparing the whole range of species from both categories, the "All species" category shows higher diversity.

Exploring trait data, I tried to understand which were the growth forms most expressed in species in the state of Santa Catarina. I thought that could have some differences in the "All species" and "Cultivated", because cultivated species are usually chosen in order to some value that they have, and that could be related to some specific growth form. When plotting a bar plot for the 50 most abundant species (Fig. 2), it is possible to see that "Cultivated" show slightly less diversity in growth form values, but species in general present a similar pattern in the expressed growth form, when comparing "Cultivated" and "All species".

```r
knitr::include_graphics(c("growth_form_50_all_ggplot.png", "growth_form_50_cult_ggplot.png"))
```
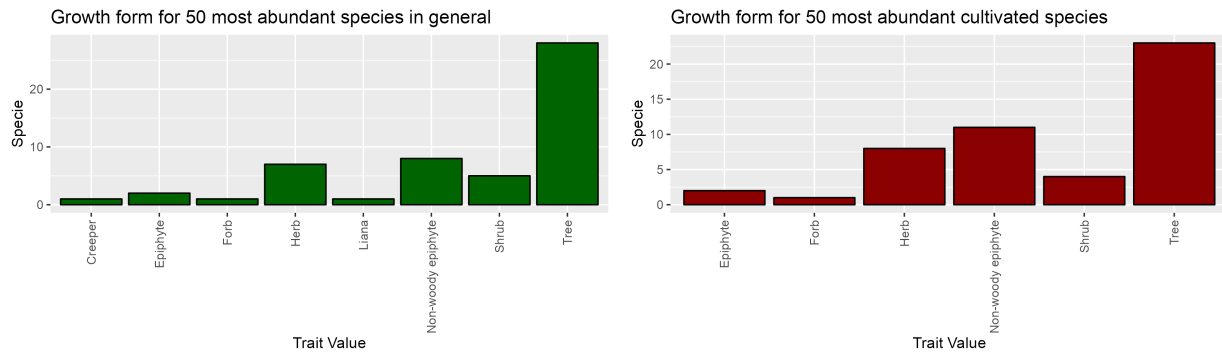

Figure 2: Bar plots showing the distribution of Growth Forms trait among species, for the 50 most abundant species. The barplot in the left is from the All Species category, and the barplot in the right is from the Cultivated category.

I also wanted to see if this similar pattern was maintained when comparing the whole set of species of both categories. These plots can be seen in Figure 3. There are a few small differences, for example, "All species" presents a bit more diversity in growth form values, but, in general, both profiles look almost the same. And that can be related with what was presented before in this report, that the cultivated species may be only a subset of the "All species", since we only worked on native species data, and it was used data from the whole state.

```
knitr::include_graphics(c("growth_form_all_ggplot.png", "growth_form_cult_ggplot.png"))
```
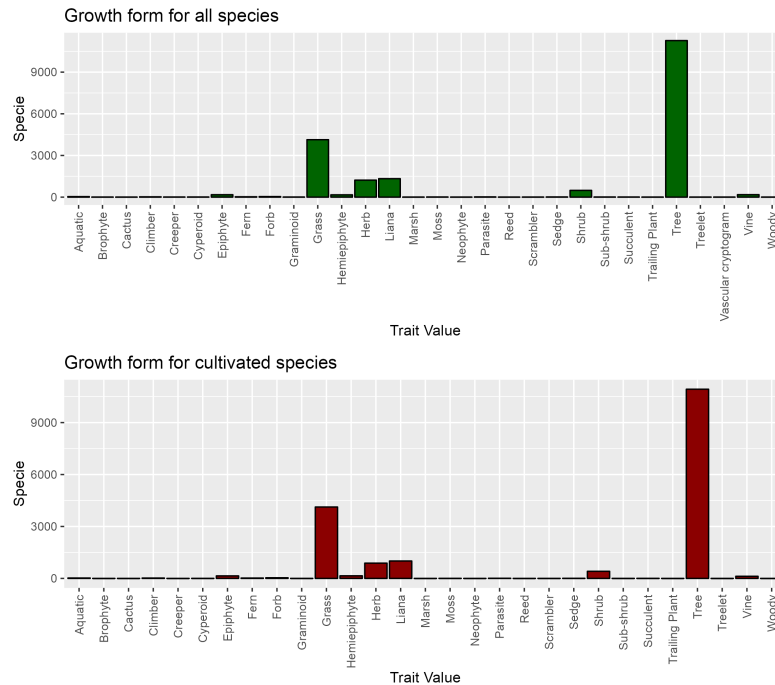

Figure 3: Bar plots showing the distribution of Growth Forms trait among species, for the whole set of species. The barplot in the top is from the All Species category, and the barplot in the bottom is from the Cultivated category.

So, for further analysis, it would be very interesting to use data from exotic species too. Exotic species might have a big influence in the profiles of cultivated species. Also, in further analysis it would be nice to divide the state in plots, and compare plots instead of subsets with occurrence and abundance data from the whole state.

Conclusions

The scripts here presented show a brief exercise on plant diversity analysis. It was done a lot of work requesting and cleaning data, and trying to understand a few patterns of native species in general and cultivated species in the state of Santa Catarina. Exploring data is a very important step before actually starting the analysis, so the method can be designed knowing the data you are working with, and fitting the best models and techniques to analyse it. That said, I do believe that using data from exotic species and dividing the area of the state in plots would be good advances for further analysis in this area.

References

Gotelli, N. J., & Ellison, A. M. (2004). *A Primer of Ecological Statistics.* Sunderland: Sinauer Associates.

Kindt, R., & Coe, R. (2005). *Tree diversity analysis.* Nairobi, Kenya: World Agroforestry Centre.

Maitner, B. (2022). *BIEN: Tools for accessing the botanical information and ecology network database.* Retrieved from https://CRAN.R-project.org/package=BIEN

Nock, C. A., Vogt, R. J., & Beisner, B. E. (2016). Functional Traits. In *eLS* (pp. 1–8). Wiley. doi: 10.1002/9780470015902.a0026282

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2020). *Vegan: Community ecology package.* Retrieved from https://CRAN.R-project.org/package=vegan