**Preliminary data analysis**

**GitHub link:** https://github.com/sydwil/ENT6707-G4

**Source:** The data was obtained from the Environmental Data Initiative database

**Any data wrangling undertaken to prepare the data for analysis**

We first added a “genus” column to group plants by genus. Because each row was an observation of one plant species from a given study, we wanted to subset the rows. There were a few typos in the genera of the species cells, so we fixed errors there first. Then, we added a column of “predation\_method” based on the table below. While all studies measured prey capture, there were several different types of units for prey capture, so we subset the data into two data sets of the most common units: 1 data set that measured proportions (n = 30) and one data set measuring actual counts of trapped arthropods (n = 52). We have reported summary statistics below of these two datasets, but we have found several of the studies that include proportions and have found total counts, so we may soon be able to combine most of these datasets together for analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Method of Predation** | **Genera** | **Number of Replicates** | **Number of replicates (w/o % value data)** |
| Passive Trapping | *Pinguicula* | 16 | 10 |
| *Drosera* | 23 | 16 |
| Pitchers | *Nepenthes* | 15 | 1 |
| *Brocchinia* | 1 | 0 |
| *Sarracenia* | 25 | 9 |
| *Triphyophyllum* | 1 | 1 |
| Active Trapping | *Utricularia* | 5 | 5 |
| *Dionaea* | 2 | 0 |

* **Summary statistics for relevant variables, variable types (e.g., categorical vs. continuous**

*Response variables, either proportions or counts of*:   
diptera, acarina, collembola, hymenoptera.not.formicidae,

thysanoptera, homoptera, coleoptera, araneae, lepidoptera, hemiptera,

formicidae, orthoptera. (numeric)

*Independent variables*: Method of predation (categorical, 3 levels).

* **Potential challenges in analysis (e.g., violations of normality; pseudoreplication), and pertinent graphical depictions**
* **If using a published database or data from a publication, please provide a complete citation and explain how your analyses build upon any existing analyses of those data.**

We found through additional research that there was a paper that utilized this data (Ellison & Gotelli 2009) but was not cited on the original Environmental Data Initiative page. Fortunately, this paper compares the prey capture of plants on a genus level, so we may still be able to achieve unique analyses with our subsetting according to their prey capture method. With this prey capture variable, we can compare catch within each group and between the three groups, hopefully demonstrating that the same trap type are more similar to each other. We intend to fit a generalized linear mixed effect model to analyze the data, given that the majority of the data is count data and should theoretically follow a Poisson distribution. For the data using proportional values, we would assume a normal distribution.

The first of three graphs using this data is a “star” plot, where each genus has been broken down into proportions of prey. The second graph addresses PIE, or probability of interspecific encounter, which is meant as a means to demonstrate how specialized to certain orders the genus of plants is. The final graph illustrates a similarity index between a few selected samples prey capture. While all analyses are very interesting and may provide useful for further work, we do not intend to use these exact statistical or graphical representations to guide our project. (don’t think I’ll include this in the final draft)

**Citations:**

Ellison, A. M., & Gotelli, N. J. (2009). Energetics and the evolution of carnivorous plants—Darwin's ‘most wonderful plants in the world’. *Journal of Experimental Botany*, *60*(1), 19-42.

Ellison, A. and N. Gotelli. (2023). Prey Capture by Carnivorous Plants Worldwide 1923-2007 ver 18. Environmental Data Initiative. https://doi.org/10.6073/pasta/38b5bf8887b039b3c0b81b4a317f8cbc (Accessed 2024-09-03).