**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. A significant amount of time has passed since this dataset was assembled in 2008, which has led to some modern studies on the topic not being included. In order to address this, and to expand the breadth of our dataset, we are performing a brief literature search adhering to the parameters set by the original author. Any publications found that would have been included if this dataset was assembled more recently will be added to our database.

|  |  |  |  |
| --- | --- | --- | --- |
| **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).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Counts studies data (n = 52) |  |  | Proportions studies data (n = 30) | |  |  |
| Prey group | **mean** | **SE** | **Prey group** | **mean** | | **SE** |
| diptera | 256.4 | 69.76 | diptera | 18.38 | | 3.49 |
| acarina | 52.1 | 21.23 | acarina | 1.09 | | 0.45 |
| hymenoptera.not.formicidae | 24.06 | 6.62 | hymenoptera.not.formicidae | 4.12 | | 0.88 |
| thysanoptera | 5.58 | 1.6 | thysanoptera | 1.28 | | 0.57 |
| homoptera | 38.14 | 13.31 | homoptera | 2.6 | | 0.68 |
| coleoptera | 75.33 | 33.83 | coleoptera | 5.55 | | 1.33 |
| araneae | 13.32 | 4.79 | araneae | 3.17 | | 1.48 |
| lepidoptera | 11.42 | 5.85 | lepidoptera | 2.08 | | 0.56 |
| hemiptera | 11.95 | 3.99 | hemiptera | 0.96 | | 0.29 |
| plecoptera | 0.1 | 0.06 | plecoptera | 0 | | 0 |
| formicidae | 150.72 | 60.08 | formicidae | 50.73 | | 6.15 |
| orthoptera | 2.1 | 0.77 | orthoptera | 0.89 | | 0.56 |

**Potential challenges in analysis (e.g., violations of normality; pseudoreplication), and pertinent graphical depictions**

A graph of a number of numbers

Description automatically generatedA graph with a line

Description automatically generated with medium confidenceA challenge of our analysis is the structure of our dataset, with a large number of observation and variable being present, as this may make it challenging to fit a model to it. Preliminary analysis on the means of individual counts/proportions via a qqplot also shows that this data is not normal. However, the fact that these plots are examining means may be throwing them off a little bit. The raw count/proportion data may be more “normal”.

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. This paper compares the prey capture of plants on a genus level, so our concept of subsetting the data according to prey capture method will still provide a unique analysis. 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 most 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. Application of the Kruskal-Wallis test may also shed light on the relationships between predation type and prey collection, and will be more cooperative in the case of non-normal data.

**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).