# 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 |

# Variables and Summary Statistics

**Independent variables:** Method of predation (categorical, 3 levels)  
**Response variables, either proportions or counts of:** Diptera, Acarina, Collembola, Hymenoptera (except Formicidae)\*, Thysanoptera, Homoptera, Coleoptera, Araneae, Lepidoptera, Hemiptera, Formicidae, Orthoptera. (Numeric)

\*Formicidae have a highly complex association with many carnivorous plants and thus are separated out in their own category for the time being. We may eventually lump them in with the rest of the order, but we wanted to keep them distinct to start.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 (except Formicidae) | 24.06 | 6.62 | **Hymenoptera (except 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 |

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

Our current concern is merging the proportion data and the individual count data in order to have the maximum number of replicates. When this is accomplished, we will update our assumptions of normalcy…

-need a read on homogeneity of variance

-and normalcy 🡪 we would expect it to have a Poisson distribution due to the fact that it is count data (we can test for this with the following lines of code:

Library(vcd)

Gf = goodfit (dataset$column, type = “poisson”, method= “ML”)

Summary(Gf)

If the P value is greater than or equal to 0.05 we cannot reject Poisson)

🡪 we should also mess with some histograms and qqlines

**Previous Publications:**

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.

**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).