**Paper Analysis SOP**

Required & Suggested:

Programs: Excel (and the variable template found in GitHub), Able2Extract, GitHub Desktop (and account), Google Drive (<https://drive.google.com/open?id=11DqXtYlkBuakdUGe7XPtY2kE32Sj71O3>)

Useful sites: <https://www.mathsisfun.com/measure/months.html>

1. **Select papers to extract in Google Drive:** All papers were contained in a google drive folder titled “papers\_to\_extract” along with folders named after each person working on the project. Individuals took papers from “papers\_to\_extract” and moved them into their named folder (e.g. “Justin Extract”) and downloaded them onto their computer.
2. **Extract raw data from tables in Able2Extract or Acrobat Pro and save in GitHub:** Papers were then opened with Able2Extract and scanned for tables containing relevant data (e.g. % volume, % abundance, total abundance, etc). Frequency of occurrence and other derived measures (e.g. indices) were not retained. If no data was found, the papers were moved into a folder titled “pdfNoData”. These tables were then extracted as .csv files and individually saved as “<name>\_<table number>” in the folder named “dataUneditedRawFiles” on GitHub. If the extraction process didn’t work, or manual entry appeared faster, papers were moved respectively into the “pdfNotExtractable” or “pdfNotExtractable” folders on GitHub to be added at a future date.
3. **Use template to collect information about samples and transpose the data:** The raw Able2Extract files were checked for extraction issues and cleared of any non-relevant data, usually leaving only the data points and prey items. Often this was done by editing the raw file, though not saving the progress and pasting the content into the following document. An excel template containing a row of all the database’s variables was then, or at a previous time, opened. Data from Able2Extract was then pasted and transposed into the aforementioned excel template, directly next to the row containing the variables. Information for each variable was then extracted from the paper manually and placed next the relevant row of data and below the relevant variable.
4. **Save as .csv to GitHub:** At the end of all this, the data in each individual excel document was then individually saved as files named “<author>\_<year>\_<table number>” into the “dataEditedRaw” folder on GitHub.
5. **Compile in R and analyze:** The rest of the data tidying was done in R, excluding any manual error fixing.

**Variable Explanations**

Site\_name: Name of the site listed in the paper. Will be used to automatically fill out lat and long towards the end of the project.

Lat: The latitude of the site surveyed. Will be automated with google later based off site name.

Long: The longitude of the site surveyed. Will be automated with google later based off site name.

Predator\_min\_length: The minimum length given for the range of the fish species (only the numerical value).

Predator\_max\_length: The maximum length given for the range of the fish species (only the numerical value).

Predator\_average\_length: The average length given for the range of the fish species (only the numerical value).

Length\_units: The units used to measure the predator lengths.

Length\_measure: The type of fish measurement used (total, fork, or caudal).

Type\_of\_fish: The name given for the species of fish in the paper.

Sample\_size: The sample size of the predator given for a specific dataset within a paper’s table.

Start\_date: The date predator collection started.

End\_date: The date predator collection ended.

Sampling\_interval: The frequency of predator collection (monthly, weekly, daily, etc).

Habitat: The habitat the predators were collected in.

Measurement\_type: The type of measurement used to numerically represent the predator’s diet (volume, mass, ect).

Measurement\_units: The units associated with the predator diet data.

Author: The author(s) of the paper.

Year: The year the paper was published.

Journal: The journal the paper was published in.

Citation: The APA citation of the paper. Fastest method is to search on google scholar.

Table\_figure: The table the data came from.

Notes: Any notes about the data set not covered by the variables above.

**GitHub Use**

Before doing anything, click “fetch origin” to get the updated GitHub data. Make all changes and additions to the database locally on your own computer, and then click “commit to master” after giving your additions a title and a summary of what you did. Immediately after, click “push origin” to update the Git server’s master branch. Code development should be done in your own development branch. Consult the group with any new ideas before adding anything into the master branch.