README File

File name: Fish\_summarizer\_function.Rmd

Created by Anela K. Akiona, [aakiona@ucsd.edu](mailto:aakiona@ucsd.edu)

Last Edited: 10 January 2025

**Description**

Fish\_summarizer\_function.Rmd is an R markdown function that calculates density and biomass of fish transect count data, using user-provided length-weight parameters, and produces summary files and plots. The motivation for developing this function was to provide a one-stop-shop for colleagues and collaborators to utilize the length-weight parameters published in \*Master Blaster citation here\*. This function was conceived and created by members of the Sandin Lab at Scripps Institution of Oceanography at UC San Diego, and as such the function is tailored for their belt transect methods, though it is meant to be customizable to other styles of belt transects.

**Arguments**

fishdata

The CSV file containing the field observation data of fish counts and lengths.

LWfile

The CSV file containing the species-specific length-weight parameters.

filename

Unique identifier to be included in output file names.

sp\_data

Name of the column in fishdata which contains the species names. The values in this column must match the values in sp\_lw.

sp\_lw

Name of the column in LWfile which contains the species names. The values in this column must match the values in sp\_data.

integrate

If TRUE, uses the integral method to calculate biomass. If FALSE, biomass is calculated using the given length.

size\_correct

Number to be subtracted from the total length of each fish prior to biomass calculation. Cannot be used if integrate = TRUE.

bin

Bin size used for length measurements. This can either be a single number or vector of numbers.

max\_size

If bin is a single number, this is the maximum size to which biomass should be calculated.

size\_threshold

Fish length at which belt transect width changes (cm). If belt width is the same for the entire transect, enter 0.

width\_below

Width of transect for fish smaller than size\_threshold (m).

width\_above

Width of transect for fish larger than size\_threshold (m). If belt width is the same for the entire transect, enter it here.

transect\_length

Length of transect (m).

**Outputs**

*Data files:*

* fish\_summary\_station\_'filename'.csv
  + Biomass and density for each station.
* fish\_summary\_group\_'filename'.csv
  + Biomass and density by trophic group at each station.
* fish\_summary\_species\_'filename'.csv
  + Biomass and density by species at each station.

*Plots*:

* fish\_summary\_density\_'filename'.png
  + Stacked bar chart of density (#/m2) by trophic group at each station.
* fish\_summary\_biomass\_'filename'.png
  + Stacked bar chart of biomass (g/m2) by trophic group at each station.

**Details**

*Data file requirements*

This code calls specific column names to conduct the calculations, and as such the data files passed to the function must contain those column names. Below are the required column names for the respective data files. Please note that unless indicated otherwise, the names must be *identical*, including capitalization.

Required columns in the fish survey data file (with order of columns not constrained):

(1) Data\_Type -- For all count data, record as "QUAN" (differentiates from "PRES" used to document presence in larger swath for remote island surveys). Only data with Data\_Type = QUAN will be included in the calculations,

(2) Date -- Part of metadata for record,

(3) Transect -- Metadata (continued), noting from which replicate transect the data were collected,

(4) Diver -- Metadata (continued), noting unique identifier of diver collecting data,

(5) Station -- Name of the station (needs to be unique within the data file, but can be repeated across dates, though the data will be lumped by station),

(6) sp\_data -- Unique species codes or taxon names which must match sp\_lw (see below). This column can have any name, but it must be passed to the function as an argument,

(7) Number -- Records the number of fish of the particular size observed in the particular transect,

(8) Size -- Records the size of fish in cm.

Required columns in length-weight file (with order of columns not constrained):

(1) sp\_lw -- Unique species codes or taxon names. The list in this column must include all species identifiers used in sp\_data. This column can have any name, but it must be passed to the function as an argument,

(2) a\_cm -- Best estimate of alpha from length-weight relationship,

(3) b\_cm -- Best estimate of beta from length-weight relationship,

(4) LTLRat -- Length conversion factor linking field-estimated length (in Total Length) to length type for length-weight relationship (can be Standard, Fork, or Total Length),

(5) Trophic -- Trophic level classification for each species. Any types of unique categorizations will work.

*Examples*

# Using the integral method  
fish\_summary(fishdata = "Pohnpei\_2018-08\_Fish\_Data.csv",   
 LWfile = "Pacific\_LW\_params.csv",  
 filename = "Pohnpei 2018",  
 sp\_data = "Species",   
 sp\_lw = "NewName",  
 integrate = TRUE,   
 bin = c(2, 5, 10, 15, 25, 50, 75, 100, 150, 200),   
 size\_threshold = 20,  
 width\_below = 2,  
 width\_above = 4,  
 transect\_length = 25)

# Using size bins with size correct  
fish\_summary(fishdata = "Pohnpei\_2016-08\_Fish\_Data.csv",   
 LWfile = "Pacific\_LW\_params.csv",  
 filename = "Pohnpei 2016",  
 sp\_data = "Species",   
 sp\_lw = "Taxon",  
 integrate = FALSE,  
 size\_correct = 2,  
 bin = 5,  
 max\_size = 300,  
 size\_threshold = 0,   
 width\_above = 2,  
 transect\_length = 30)

*Notes on size correcting and the integral method of biomass calculation*

When estimating fish lengths in the field, it is often prudent to classify individual fish into size bins, rather than attempting to estimate length to the nearest 1 or 0.1 cm. While this is a commonplace and practical solution to the challenge of estimating the length of often rapidly moving organisms, it can introduce bias when calculating biomass – using the length estimated in the field assumes that every individual is at the maximum of the size bin, and thus biomass calculations can become artificially inflated. One such way to correct this bias is the use of a ‘size correct,’ which subtracts a user-defined amount from each fish size (e.g. 2 cm off every length for fish in 5 cm bins). This, however, can be quite arbitrary and may not be applicable when bin size changes with fish length.

Here, we standardize the correction using the mean value theorem for integrals, which states that if *f(x)* is a continuous function on the closed interval [*a, b*], then there exists a number *c* in the closed interval such that

Where is the average value of the function from *a* to *b*. For the length-weight relationship, this can be written as,

Where and are the lower and upper limits of the size bin, respectively, and is the mean weight of the size bin. This can be rearranged to give,

This method corrects for potential bias introduced in the belt survey methods and will work regardless of the width of the size bin. When integrate=TRUE, the function will calculate biomass using the mean value theorem for integrals. When integrate=FALSE, the function will calculate biomass using the given length, i.e. the maximum value for that size bin.