

LTO Benthic Data Analysis SOP

KEY:

Red – used for R functions

Blue – used for changeable information (ie. The names of objects/data frames)

Black – used for non-negotiable parts of the code (must be in there to allow the code to run correctly)

Steps to Follow	Notes
1. Set working directory and/or load in data <pre> setwd('Working Directory Folder 1/Folder2/Folder3') data<-read.csv('data.csv', header = TRUE) </pre>	<p>When using the setwd function, you must create a pathway of folders leading to the files of interest, this can require any number of folders not just 3 as shown in the example.</p> <p>If multiple files are to be used simultaneously, the final and most specific folder name should be the one that holds all of the files of interest.</p>
2. Create a matrix of species and species cover <pre> Library(tidyverse) Library(reshape) matrix<-data %>% group_by(transect_no, site_no, site_name, date, distance) %>% pivot_longer(cols=c(7,9,11,13),values_to='species',names_to='S_number')%>% pivot_longer(cols=c(7,8,9,10),values_to='percentage',names_to='P_number') </pre>	<p>Creates a dataframe which replicates each species by each percentage cover value for every 'group_by' value.</p> <p>IE. For each distance on the transect we have four possible species and their associated four possible percentage cover values. Therefore 16 possible combinations for each grouped value.</p>

Commented [AF1]: Maybe add to the additional notes section, how to coalate multiple files

LTO Benthic Data Analysis SOP

3. Extract identifiers from columns to pair species names and species cover	
<pre>matrix\$species_num <- as.numeric(gsub("[^0-9]", "", matrix\$S_number)) matrix\$cover_num <- as.numeric(gsub("[^0-9]", "", matrix\$P_number))</pre>	<p>Creates two additional columns which are a copy of S-number and P-number columns from previous step. Uses a gsub function to only include numbers 0-9, removing characters seen in these columns.</p> <p>This range of numbers can be changed dependent on how many species and percentage cover values that you have for each set of group_by values.</p>
4. Filter data and remove unnecessary rows/columns	
<pre>summarised_data <- matrix %>% filter(species_num == cover_num) %>% select(-S_number, -P_number, -species_num, -cover_num)</pre>	<p>Produces a new dataframe which holds only rows where the species_num and cover_num match using the filter function.</p> <p>Select is used to remove any unnecessary columns in the data frame after rows where numbers match have been filtered for.</p>
OPTIONAL: Clean data to remove NA's	
<pre>library(dplyr) clean <- summarised_data %>% group_by(transect_no, site_no, year, date, site_name, distance, species) %>% drop_na(species) %>% dplyr::summarise(total = sum(percentage_cover))</pre>	<p>Will drop all rows with NA's</p> <p>Can also remove distance if wanting to look across entire transect.</p>

LTO Benthic Data Analysis SOP

Example Outputs & Explanations

1. Set working directory and/or load in data

transect_no	site_no	site_name	year	date	distance	species1	cover1	species2	cover2	species3	cover3	species4	cover4
1	LT08	CAVES	2017	06/01/2017	0	N/A	0.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	2.5	Favia	55.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	5	Porites	11.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	7.5	Porites	22.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	10	N/A	0.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	12.5	Porites	44.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	15	N/A	0.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	17.5	Porites	60.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	20	N/A	0.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	22.5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	25	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	0	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	2.5	Porites	11.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	5	Porites	33.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	7.5	Lobohyllia	11.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	10	Porites	77.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	12.5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	15	Porites	100.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	17.5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	20	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	22.5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	25	N/A	0.0	N/A	0.0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	0	N/A	0.0	N/A	0.0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	2.5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	5	N/A	0.0	N/A	0.0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	7.5	Favia	11.0	N/A	0.0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	10	N/A	0.0	N/A	0.0	N/A	0	N/A	0

The output of the read.csv function should display data in the same format as the excel/csv file being read in.

transect_no	site_no	site_name	year	date	distance	species1	cover1	species2	cover2	species3	cover3	species4	cover4
1	LT08	CAVES	2017	06/01/2017	0	N/A	0	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	2.5	Favia	55	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	5	Porites	11	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	7.5	Porites	22	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	10	N/A	0	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	12.5	Porites	44	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	15	N/A	0	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	17.5	Porites	60	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	20	N/A	0	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	22.5	N/A	0	N/A	0	N/A	0	N/A	0
1	LT08	CAVES	2017	06/01/2017	25	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	0	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	2.5	Porites	11	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	5	Porites	33	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	7.5	Lobohyllia	11	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	10	Porites	77	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	12.5	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	15	Porites	100	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	17.5	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	20	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	22.5	N/A	0	N/A	0	N/A	0	N/A	0
2	LT08	CAVES	2017	6/13/2017	25	N/A	0	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	0	N/A	0	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	2.5	N/A	0	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	5	N/A	0	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	7.5	Favia	11	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	10	N/A	0	N/A	0	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	12.5	Favia	11	Porites	11	N/A	0	N/A	0
3	LTO13	PAO	2017	6/13/2017	15	Porites	11	N/A	0	N/A	0	N/A	0

LTO Benthic Data Analysis SOP

2. Create a matrix of species and species cover

Group_by columns which remain the same

transect_no	site_no	site_name	year	date	distance	S_number	species	P_number	percentage_cover
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover1	0
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover3	0
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover4	0
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover1	0
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover3	0
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover4	0
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover1	0
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover3	0
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover4	0
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover1	0
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover3	0
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover4	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover1	55
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover2	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover3	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover4	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover1	55
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover3	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover4	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover1	55
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover2	0
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover3	0

Shows where each of the four species for each set of group_by values is paired with each of the four possible percentage cover values for the same set of group_by values.

LTO Benthic Data Analysis SOP

3. Extract identifiers from columns to pair species names and species cover

transect_no	site_no	site_name	year	date	distance	S_number	species	P_number	percentage_cover	species_num	cover_num
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover1	0	1	1
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover2	0	1	2
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover3	0	1	3
1	LTO8	CAVES	2017	06/01/2017	0	species1	NA	cover4	0	1	4
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover1	0	2	1
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover2	0	2	2
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover3	0	2	3
1	LTO8	CAVES	2017	06/01/2017	0	species2	NA	cover4	0	2	4
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover1	0	3	1
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover2	0	3	2
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover3	0	3	3
1	LTO8	CAVES	2017	06/01/2017	0	species3	NA	cover4	0	3	4
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover1	0	4	1
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover2	0	4	2
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover3	0	4	3
1	LTO8	CAVES	2017	06/01/2017	0	species4	NA	cover4	0	4	4
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover1	55	1	1
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover2	0	1	2
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover3	0	1	3
1	LTO8	CAVES	2017	06/01/2017	2.5	species1	Favia	cover4	0	1	4
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover1	55	2	1
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover2	0	2	2
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover3	0	2	3
1	LTO8	CAVES	2017	06/01/2017	2.5	species2	NA	cover4	0	2	4
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover1	55	3	1
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover2	0	3	2
1	LTO8	CAVES	2017	06/01/2017	2.5	species3	NA	cover3	0	3	3

The numbers in the **S_number** column should match those in the new **species_num** column and the same with **P_number** & **cover_num**. Ensure that this is correct as a sanity check.

The only difference between columns should be the removes of characters from the new columns completed by the **gsub** function.

LTO Benthic Data Analysis SOP

4. Filter data and remove unnecessary rows/columns

transect_no	site_no	site_name	year	date	distance	species	percentage_cover
1	LTO8	CAVES	2017	06/01/2017	0	NA	0
1	LTO8	CAVES	2017	06/01/2017	0	NA	0
1	LTO8	CAVES	2017	06/01/2017	0	NA	0
1	LTO8	CAVES	2017	06/01/2017	0	NA	0
1	LTO8	CAVES	2017	06/01/2017	2.5	Favia	55
1	LTO8	CAVES	2017	06/01/2017	2.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	2.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	2.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	5	Porites	11
1	LTO8	CAVES	2017	06/01/2017	5	NA	0
1	LTO8	CAVES	2017	06/01/2017	5	NA	0
1	LTO8	CAVES	2017	06/01/2017	5	NA	0
1	LTO8	CAVES	2017	06/01/2017	7.5	Porites	22
1	LTO8	CAVES	2017	06/01/2017	7.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	7.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	7.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	10	NA	0
1	LTO8	CAVES	2017	06/01/2017	10	NA	0
1	LTO8	CAVES	2017	06/01/2017	10	NA	0
1	LTO8	CAVES	2017	06/01/2017	10	NA	0
1	LTO8	CAVES	2017	06/01/2017	12.5	Porites	44
1	LTO8	CAVES	2017	06/01/2017	12.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	12.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	12.5	NA	0
1	LTO8	CAVES	2017	06/01/2017	15	NA	0
1	LTO8	CAVES	2017	06/01/2017	15	NA	0
1	LTO8	CAVES	2017	06/01/2017	15	NA	0

New dataframe should display correct number of rows for species and percentage cover values for each transect.

LTO Benthic Data Analysis SOP

OPTIONAL STEP

transect_no	site_no	year	date	site_name	distance	species	total
1	LTO8	2017	06/01/2017	CAVES	2.5	Favia	55.0
1	LTO8	2017	06/01/2017	CAVES	5	Porites	11.0
1	LTO8	2017	06/01/2017	CAVES	7.5	Porites	22.0
1	LTO8	2017	06/01/2017	CAVES	12.5	Porites	44.0
1	LTO8	2017	06/01/2017	CAVES	17.5	Porites	60.0
2	LTO8	2017	6/13/2017	CAVES	2.5	Porites	11.0
2	LTO8	2017	6/13/2017	CAVES	5	Porites	33.0
2	LTO8	2017	6/13/2017	CAVES	7.5	Lobohyllia	11.0
2	LTO8	2017	6/13/2017	CAVES	10	Porites	77.0
2	LTO8	2017	6/13/2017	CAVES	15	Porites	100.0
3	LTO13	2017	6/13/2017	PAO	7.5	Favia	11.0
3	LTO13	2017	6/13/2017	PAO	12.5	Favia	11.0
3	LTO13	2017	6/13/2017	PAO	12.5	Porites	11.0
3	LTO13	2017	6/13/2017	PAO	15	Porites	11.0
3	LTO13	2017	6/13/2017	PAO	20	Acropora	33.0
3	LTO13	2017	6/13/2017	PAO	25	Lobophytum	22.0
4	LTO10	2017	6/15/2017	DEVILS	12.5	Lobophytum	11.0
4	LTO10	2017	6/15/2017	DEVILS	17.5	Lobophytum	11.0
4	LTO10	2017	6/15/2017	DEVILS	17.5	Poclipora	22.0
4	LTO10	2017	6/15/2017	DEVILS	20	Lobophytum	11.0
4	LTO10	2017	6/15/2017	DEVILS	22.5	Porites	11.0
4	LTO10	2017	6/15/2017	DEVILS	22.5	Sinularia	44.0
5	LTO10	2017	6/17/2017	DEVILS	17.5	Lobophytum	11.0

LTO Benthic Data Analysis SOP

Additional Useful Functions

- **Str(data)** – outputs a list of data frame columns and the operators by which they are read in R. Most commonly, data will be formatted numerically or as factors, characters and integers.
- **Data\$column_name<-as.numeric(Data\$column_name)** – used to rectify when a column is read incorrectly in R. This code would be used to force R to read the specified column as numeric, however similar code can be used to force data into character and factor operators. This would be done using changes to the function such as:
 - **as.character**
 - **as.factor**
- **install.packages("package_name")** – This SOP requires packages ‘reshape’, ‘tidyverse’ and ‘dplyr’. If when running the **library(package)** function, the output produces an error it may be because the package has not yet been installed. In this instance use the above code.
- **library(openxlsx)**

write.xlsx(data, file = "Name_Data.xlsx") – used to save the reshaped data frame to the device as a new excel file.

Online Code and Data File for Examples can be found using the github:

<https://github.com/amberf02/LTO-benthic-data-reshape.git>