

Reorganisation following disturbance: multi trait-based methods in R

Part 1: Getting to know the dataset and the ecosystem

We will start by getting to know the Lizard Island fish data presented in the introductory practical lecture. Whenever using data provided by others it is important that you fully understand how the data were collected, what was measured and the units used. Before you get going though, get ready to work in R using your Rstudio account.

Go to RStudio, and open the `Functional_diversity_practical` project you created in preparation for the practical. Open the R script you saved previously if it is not already open.

Preparing your R workspace

In preparation for the practical you installed the libraries we will be using for the practical. The functions available within these libraries are not available until you load them to your R workspace – you need to load the libraries you want every time you open a new R session.

You load R packages using the `library()` function. In your `Day_1_code` script type:

```
library(reshape)

## Warning: package 'reshape' was built under R version 4.3.2

options(max.print = 999999999) ## changes output view settings
```

Over time you will build up instructions in the script – and this forms the record of your analysis. This way you save your instructions should you need to repeat it later. But make sure you save your script as you go!

Getting to know the datafile structure

The first requirement for calculating a trait-based functional diversity index is a dataset on species abundance. Underwater visual surveys on coral reef fish assemblages at different sites have been conducted by SCUBA divers, and provided to you. In this exercise, you will get to know these data.

Make sure the datafile is loaded into the R environment:

```
dat <- read.csv("LI_fish_abundance_pre_post_bleaching.csv")
```

Look at the size of your dataset using `dim(dat)`. To understand the two numbers that R returned – you can use the help function on the `dim` function (!) by typing `help(dim)`.

Note that you can apply the help function to any other function in R to learn more about it.

```
dim(dat)
help(dim)
```

Q1. What are the dimensions of the dataset in rows and columns?

__ rows

__ columns

Familiarize yourself with the dataset by looking at the names of the columns and what they contain. You can see column names by typing `names(dat)` and view the first few rows of data by typing `head(dat)`.

```
names(dat)
head(dat)
```

The default setting for the head function is to return the first six rows of the dataframe, but if you wish to see more or less, just include the number after the name of the R object you are passing to the function, like this:

```
head(dat, 12)
```

The first four columns of the dataset contain information on the fish survey: Habitat, Site, Replicate and Bleaching. The remaining columns are species abundance estimates i.e. how many of each species were seen on each survey.

Another thing that is useful to learn very early on when handling a new dataset is the type of data you are dealing with. Datasets typically contain data of different data modes. You can find out what type of data is in a dataset by using the `str()` function. Here `str` stands for structure.

```
str(dat)
```

Scroll further up in the R console until you can see the command you just passed to R – this puts you at the start of the R output and it tells you that you have a dataframe of ____ observations of ____ variables – in other words, ____ rows and ____ columns.

We want the first 4 columns of the dataset (Habitat, Site, Replicate, and Bleaching) to be recognised as factors, so that R knows that each different value should be treated as a level of a grouping variable, rather than simply text (e.g. character 'chr') or a number (e.g. integer 'int'). The Replicate column will likely show as an integer type 'int' which is numeric data; we want this to be treated as a factor. The Habitat, Site, and Bleaching columns may show as type character 'chr' or as type factor 'fct', depending on the default settings in your R setup.

The Replicate column needs converting to a factor, and if your Habitat, Site and Bleaching columns are showing as character 'chr' types then you also need to convert these to factors using the below code:

```
dat$Habitat<-as.factor(dat$Habitat)
dat$Site<-as.factor(dat$Site)
dat$Replicate<-as.factor(dat$Replicate)
dat$Bleaching<-as.factor(dat$Bleaching)
```

For columns that contain factors (i.e. the first 4), you can find out the factor levels of a column by using the levels function. Type: `levels(dat$Bleaching)`. This tells you that you have two factor levels within the column named Bleaching – Pre and Post bleaching – i.e. fish surveys were done before and after the bleaching event.

The expression `dat$Bleaching` refers to the Bleaching component in the R object named `dat` (in this case a column called Bleaching in a dataframe called `dat`). You can use this dollar sign to access any column in a dataframe.

Q2 What are the different factor levels in the habitat column?

Use the `levels()` function shown above to find out what habitat types are contained in the Habitat column. Then, using the details presented in Richardson et al. (2018), describe each habitat factor level in the data frame and using the results in the paper report how coral cover changed pre (2015) and post (2016) the thermal stress. Richardson et al. (2018) can be accessed free of charge and without subscription here: <https://doi.org/10.25903/5b57c26b0beb7> (see Chapter 4).

(HINT: read the benthic composition section in the results of Richardson et al. 2018. Increased temperature can lead to coral bleaching. Which of these habitats might be more and less impacted by bleaching? How then do you expect these habitats to differ pre and post bleaching?)

Getting to know the dataset sampling design

To report on these data you need to understand the sampling design as well as the sampling method used to survey fish.

Q3a What was the sampling method used to collect these data?

Using the details presented in Richardson et al. 2018, describe in your own words how the fish survey transects were carried out. Consider drawing a diagram to visually display this - You can use this text and diagram later for your poster content.

You can now type `table(dat$Site, dat$Habitat, dat$Bleaching)` to count the number of replicate transects there are for each site (A, B, C) and habitat pre and post bleaching.

The `table()` function is a useful one for getting to know a dataset, as it will build a contingency table, counting each combination of factor levels. In this case, the number of sites surveyed per habitat and bleaching combination.

Q3b How many replicate fish surveys were done at each site and how many sites are surveyed per habitat type?

This means you are dealing with replicate samples per habitat type pre and post bleaching. You will later use these samples to calculate mean and standard error estimates per habitat type and you can compare these estimates pre and post bleaching.

Exploring the fish data

By now you should be getting a good sense of how many fish surveys were done, but what about the actual fish counts? Type `head(dat)` in your script again and run this. All of the columns beyond column 4, Bleaching, are the counts of fish, with each column header being the species name. To understand the range of counts use type:

```
summary(dat)
```

This is a handy function, as when it is applied to integers, it returns summary metrics for those numbers, i.e. the minimum, maximum, median and mean. For example – the mean estimate of the first fish column, a species named *Abudefduf bengalensis* is 0.76 and the maximum count observed is 17.

Q4 Which species has the highest maximum count and what is it?

Eyeballing data like this can be error prone, as you can easily miss details, but it is also important that you spend time looking at the data, to get to know it before you formally analyse it. Scanning over the summary output is a good way to explore a new dataset.

For much of this practical, you will only focus on one habitat type. Your instructor will assign you to a habitat.

Based on your habitat assignment, note their mean counts for the following species (A-B). N.B. You only need to inspect the species that relate to your habitat group:

Branching porites A. *Pomacentrus moluccensis* B. *Chrysiptera rollandi*

Mixed coral A. *Pomacentrus moluccensis* B. *Cheilodipterus quinquelineatus*

Soft coral A. *Pomacentrus chrysurus* B. *Halichoeres nebulosus*

Low coral cover A. *Archamia zosterophora* B. *Pomacentrus amboinensis*

We will next use the online database FishBase to learn a little bit more about the ecology and function of these individual species.

1. Open a new web browser and go to [FishBase](#)
2. In the Genus + species bar type the species name e.g. *Pomacentrus moluccensis*
3. Scroll down and read the Biology tab, then scroll to the bottom of the page and click the Ecology tab
4. Read the available information about the distribution, habitat and diet Repeat for all the species (A-B) above and fill out this table

Species name Habitat Diet Notes on biology and ecology

A

Species name Habitat Diet Notes on biology and ecology

B

Q5 Given what you have learned about the ecology of species A-B write down how you expect each to be impacted by a coral bleaching event and briefly discuss why.

Based on what you have learnt about the habitat, dietary requirements and body size of each species, make a prediction as to whether these fish would increase, stay the same, or decrease in abundance after coral bleaching associated habitat change.

These predictions are hypotheses of how you think these specific species might be impacted by coral bleaching. You can confront the null hypotheses later using a statistical test.

Summarizing the fish count data

You know from your preliminary exploration of the datafile that there were replicate surveys (samples) taken per habitat type before and after bleaching. You will use these samples to calculate the mean and standard error estimate of species abundance per habitat, to see whether overall there was a change pre and post bleaching.

Because there are so many species, we will continue to focus on the species A-B above, so the first step is to subset the dataframe with these species of interest.

A quick way to subset a dataframe in R is to extract only the data columns you want, using square brackets. To explain how we will do this, consider the dataframe `dat`.

The code below returns row 1, column 1 of the dataframe called 'dat':

```
dat[1,1]
```

This next code returns rows 1-4 and column 1:

```
dat[1:4, 1]
```

This next code returns rows 1-4 and columns 1-4 of the dataframe:

```
dat[1:4, 1:4]
```

So the format is dataframe[rows, columns], if you leave either element before or after the comma blank, then all rows or all columns will be returned e.g. this would return all rows but just columns 1-2: dat[, 1:2]

You can also include columns by name when subsetting. This is the code for if you wanted to take all rows and only the columns named Habitat and Bleaching, and store it in a new dataframe called datsub. The c() function means 'concatenate' or 'join together' so it is essentially saying take both the Habitat and Bleaching column and put them in the subset.

```
datsub <- dat[, c("Habitat", "Bleaching")]
```

Edit the below code to create a subset that includes the columns Habitat, Site, Bleaching and the two columns containing your two species of interest (species A and B) above, stored in a new dataframe called datsub.

```
datsub <- dat[, c("Habitat", "Site", "Bleaching", "Zebrasoma.velifer", "Chrysiptera.rollandi")]
```

So far we have only created subsets of the dataframe that take whole columns or rows of data. Another way to subset a dataframe is by using the subset function, where you select only data that meets a certain criteria (e.g., only taking the data from a particular habitat type).

Subsetting the data to your habitat of focus

This is the point at which you will start working with your own group allotted subset of data.

For this rest of the practical, you will only focus on your assigned habitat type. Based on this habitat assignment, you should subset the dataset using the code below based on your habitat group.

For example, if you are in the branching_Porites group type. The below code filters out only the rows which contain the factor level branching Porites in the Habitat column and creates a new dataframe called datsub2.

```
datsub2 <- subset(datsub, Habitat == "Branching_porites")
```

Check which habitat group you are in and amend the code above to subset it accordingly. It will be one of the following: "Branching_porites", "Low_coral_cover", "Mixed_coral", "Soft_coral".

Now you have only the species and habitat of focus, next we will calculate the mean abundance of these species before and after bleaching. To understand how variable

TIEE

Teaching Issues and Experiments in Ecology - Volume 20, June 2024

this mean estimate of abundance is, we will also calculate the standard error. As a reminder, the standard error is a measure of the uncertainty around the mean estimate.

Before we calculate the mean and standard error estimates, we are going to reformat the data to make it easier to work with. Currently, there are observations of multiple species per row (wide format), R works best with data in long format i.e. one variable per column and only one unique observation per row. Luckily there is a function in reshape package that makes the reformatting data very easy. The function is called melt.

```
dadata<-reshape::melt(datsub2)
```

Now let's give the dataframe sensible column names. You can look at the current names of the object dadata and then overwrite them by typing:

```
names(dadata)
names(dadata)<-c("Habitat", "Site", "Bleaching", "Species", "Count")
```

To calculate the mean and standard error estimates of the abundance of these species in the branching Porites habitat before and after bleaching, we will use the aggregate function.

Aggregate is a useful function to understand, as it allows you to pass a function across multiple columns in a dataframe. In this case the function is mean().

```
means<-aggregate(dadata$Count ~ dadata$Species + dadata$Bleaching + dadata$Habitat, FUN = mean)
means
```

R automatically names the columns in the dataframe generated by the aggregate function. Let's rename them with better names.

```
names(means)<-c("Species", "Bleaching", "Habitat", "Mean")
```

Now we want to calculate the standard error of the mean, to understand how variable the mean estimates are. There isn't a function built into R that does this, but that isn't a problem, we will just write one ourselves. Run the following code:

```
standard.error<-function(x){
  sqrt(var(x)/length(x))
}
```

Now let's apply the newly created standard.error function in the same manner as you did when calculating the mean above, but use it to create an R object called se.

```
se<-aggregate(dadata$Count ~ dadata$Species + dadata$Bleaching + dadata$Habitat, FUN = standard.error)
se
```


We can rename the parts of the `se` object to something sensible.

```
names(se)<-c("Species", "Bleaching", "Habitat", "SE")
```

Let's stick together the dataframe containing the means, with the one column we want from the object `se`, the SE column, using the `cbind` function. Then we will rename them.

```
mean_se<-cbind(means, se[, "SE"])
```

```
names(mean_se)<-c("Species", "Bleaching", "Habitat", "Mean", "SE")
```

Q6 How do the counts of these species differ before and after bleaching. Do they align with your predictions in Q5?

You should inspect the mean and standard error values you have estimated for the species in your dataset to answer this question (type `mean_se` and run this to look at the output).

Write a summary sentence that describes how the species abundance of your two species compares pre and post bleaching. Sketch below a bar plot showing the mean and standard errors you have generated for these species before and after bleaching.

Well done!

You have made it to the end of Practical – Part 1. You've learnt a lot about working with R by getting to know this dataset, and you've learnt a bit about the impact that bleaching had on a few species. There are, however, ~ 200 coral reef fish species in this dataset, and each has its own ecological function on the reef. Next week we will learn how to use functional trait methods to handle all these species at once, to try and understand how the community is affected as a whole. We will also start generating summary plots, to turn your sketch above into a beautiful plot.

For next week you should do 3 things:

- Read the following papers to i) consider the benefits of thinking about diversity in terms of ecosystem function as compared to diversity from a species richness perspective ii) find out what impacts humans can have on the functional diversity

of coral reef fish assemblages and ii) understand the concept of how the trait space is produced.

D'Agata, S. et al. 2014. Human-mediated loss of phylogenetic and functional diversity in coral reef fishes. *Current Biology* 24:555-560. <https://doi.org/10.1016/j.cub.2014.01.049> (open access).

Mouillot, D., et al. 2013. A functional approach reveals community responses to disturbances. *Trends in Ecology & Evolution* 28:167-177. (Author copy here: <http://villeger.sebastien.free.fr/pdf%20publis/Mouillot%20et%20al%20%282013%29%20TREE.pdf>)

- Next week we will handle the outputs of a multivariate analysis, a principal coordinates analysis (PCoA). If you want to understand the underlying approach used to create the trait space – watch these two videos to understand these methods:

https://www.youtube.com/watch?v=HMOI_lkzW08&vl=en

<https://www.youtube.com/watch?v=pGAUHhLYp5Q>

- You will each work with your own subset of the data for the rest of the practical. This means some of you will work with different habitat types. You saw today how you can select species of interest from your dataset and compare the means and standard errors before and after the warming event. For your own poster, we will assess the functional diversity trends, but we would like you to investigate the trends in at least two specific species. We recommend you select a couple based on your reading and exploration of the habitat specific dataset you have been allocated. Before next practical, look into and decide which species you will focus on.