

## R bootcamp exercise - Reef Life Survey data

Daniel Viana and Jessica Zamborain Mason

October, 2022

We are going to work with data from underwater surveys conducted in Mozambique. This data is available from the Reef Life Survey website ([reeflifesurvey.com](http://reeflifesurvey.com)) and contains fish counts per species and size classes. Therefore, for each survey site and species, divers counted the number of fish observed for each size class. For each survey, one transect of 5m x 50m was conducted (total area of 250m<sup>2</sup>). We will use this data to calculate total biomass of fish in each site.

- 1) Read explore the data ("RLS\_MOZ.csv"). How many variables does the dataset have? What information it contains?
- 2) This data is in the wide format, where each size class is a different column. Transform this data into the long format, where size class is only one column and each row represent the observed number of fish (tip: use the `reshape2::melt()` function). Delete all rows with missing data or NA's (tip: use the `drop_na()` function). Rename columns.
- 3) Read the data ("RLS\_MOZspp\_a\_b.csv"). This data contains the "a" and "b" parameters needed to convert fish length to weight (All these values were taken from [fishbase.se](http://fishbase.se)). Now, join the RLS survey data with the length/weight parameters (tip: use `left_join()` function).
- 4) Calculate the estimated biomass for each row of data using the following equation:  
$$biomass = abundance * a * (length^b)$$
- 5) Plot the total biomass of fish per site.
- 6) Plot total biomass of fish with size greater than 20cm.
- 7) What other interesting analysis can you do with the data? What other interesting questions can you ask? try to do it!

##Answers 1) Read explore the data ("RLS\_MOZ.csv"). How many variables does the dataset have? What information it contains?

```
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse
1.3.0 —
```

```
## ✓ ggplot2 3.3.2    ✓ purrr   0.3.3
## ✓ tibble  2.1.3    ✓ dplyr   0.8.4
## ✓ tidyr   1.0.2    ✓ stringr 1.4.0
## ✓ readr   1.3.1    ✓ forcats 0.5.0
```

```
## — Conflicts —————
tidyverse_conflicts() —
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

dta <- read_csv("~/Rbootcamp_Madagascar/data/RLS_MOZ.csv")

## Parsed with column specification:
## cols(
##   .default = col_logical(),
##   country = col_character(),
##   site_code = col_character(),
##   species_name = col_character(),
##   `2.5` = col_double(),
##   `5` = col_double(),
##   `7.5` = col_double(),
##   `10` = col_double(),
##   `12.5` = col_double(),
##   `15` = col_double(),
##   `20` = col_double(),
##   `25` = col_double(),
##   `30` = col_double(),
##   `35` = col_double(),
##   `40` = col_double()
## )

## See spec(...) for full column specifications.

head(dta)

## # A tibble: 6 x 31
##   country site_code species_name `2.5` `5` `7.5` `10` `12.5` `15`
##   <chr>    <chr>    <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Mozamb... MOZ1      Abudefduf v...   NA    NA     2    NA    NA    NA
## 2 Mozamb... MOZ1      Acanthurus ...   NA     5    15   40    25    NA
## 3 Mozamb... MOZ1      Centropyge ...   NA     1    NA    NA    NA    NA
## 4 Mozamb... MOZ1      Chaetodon d...   NA     1    NA    NA    NA    NA
## 5 Mozamb... MOZ1      Chaetodon l...   NA    NA     1    NA    NA    NA
## 6 Mozamb... MOZ1      Cheilinus t...   NA    NA    NA    NA    NA     1
## # ... with 21 more variables: `25` <dbl>, `30` <dbl>, `35` <dbl>, `40`
## #   `50` <lgl>, `62.5` <lgl>, `75` <lgl>, `87.5` <lgl>, `100` <lgl>,
## #   `112.5` <lgl>, `125` <lgl>, `137.5` <lgl>, `150` <lgl>, `162.5` <lgl>,
## #   `175` <lgl>, `187.5` <lgl>, `200` <lgl>, `250` <lgl>, `300` <lgl>,
## #   `350` <lgl>, `400` <lgl>
```

```
names(dta)
```

```
## [1] "country"      "site_code"    "species_name" "2.5"          "5"
## [6] "7.5"          "10"           "12.5"         "15"           "20"
## [11] "25"           "30"           "35"           "40"           "50"
## [16] "62.5"         "75"           "87.5"         "100"          "112.5"
## [21] "125"          "137.5"        "150"          "162.5"        "175"
## [26] "187.5"        "200"          "250"          "300"          "350"
## [31] "400"
```

```
str(dta)
```

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 152 obs. of 31
variables:
## $ country      : chr  "Mozambique" "Mozambique" "Mozambique" "Mozambique"
...
## $ site_code    : chr  "MOZ1" "MOZ1" "MOZ1" "MOZ1" ...
## $ species_name: chr  "Abudefduf vaigiensis" "Acanthurus triostegus"
"Centropyge bispinosa" "Chaetodon dolosus" ...
## $ 2.5          : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 5            : num  NA 5 1 1 NA NA 3 2 15 NA ...
## $ 7.5          : num  2 15 NA NA 1 NA 2 NA NA 5 ...
## $ 10           : num  NA 40 NA NA NA NA NA NA NA NA ...
## $ 12.5         : num  NA 25 NA NA NA NA NA NA NA NA ...
## $ 15           : num  NA NA NA NA NA 1 NA NA NA NA ...
## $ 20           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 25           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 30           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 35           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 40           : num  NA NA NA NA NA NA NA NA NA NA ...
## $ 50           : logi  NA NA NA NA NA NA ...
## $ 62.5         : logi  NA NA NA NA NA NA ...
## $ 75           : logi  NA NA NA NA NA NA ...
## $ 87.5         : logi  NA NA NA NA NA NA ...
## $ 100          : logi  NA NA NA NA NA NA ...
## $ 112.5        : logi  NA NA NA NA NA NA ...
## $ 125          : logi  NA NA NA NA NA NA ...
## $ 137.5        : logi  NA NA NA NA NA NA ...
## $ 150          : logi  NA NA NA NA NA NA ...
## $ 162.5        : logi  NA NA NA NA NA NA ...
## $ 175          : logi  NA NA NA NA NA NA ...
## $ 187.5        : logi  NA NA NA NA NA NA ...
## $ 200          : logi  NA NA NA NA NA NA ...
## $ 250          : logi  NA NA NA NA NA NA ...
## $ 300          : logi  NA NA NA NA NA NA ...
## $ 350          : logi  NA NA NA NA NA NA ...
## $ 400          : logi  NA NA NA NA NA NA ...
## - attr(*, "spec")=
## .. cols(
## ..   country = col_character(),
```

```
## .. site_code = col_character(),
## .. species_name = col_character(),
## .. `2.5` = col_double(),
## .. `5` = col_double(),
## .. `7.5` = col_double(),
## .. `10` = col_double(),
## .. `12.5` = col_double(),
## .. `15` = col_double(),
## .. `20` = col_double(),
## .. `25` = col_double(),
## .. `30` = col_double(),
## .. `35` = col_double(),
## .. `40` = col_double(),
## .. `50` = col_logical(),
## .. `62.5` = col_logical(),
## .. `75` = col_logical(),
## .. `87.5` = col_logical(),
## .. `100` = col_logical(),
## .. `112.5` = col_logical(),
## .. `125` = col_logical(),
## .. `137.5` = col_logical(),
## .. `150` = col_logical(),
## .. `162.5` = col_logical(),
## .. `175` = col_logical(),
## .. `187.5` = col_logical(),
## .. `200` = col_logical(),
## .. `250` = col_logical(),
## .. `300` = col_logical(),
## .. `350` = col_logical(),
## .. `400` = col_logical()
## .. )
```

## 2) Transform data to the wide format

```
dta_wide = dta %>%
  reshape2::melt(id.vars = c("country", "site_code", "species_name")) %>%
  drop_na(value) %>%
  rename(length = variable,
         abundance = value)
```

## 3) Add length/weight parameters

```
RLS_MOZspp_a_b <- read_csv("~/Rbootcamp_Madagascar/data/RLS_MOZspp_a_b.csv")
```

```
## Parsed with column specification:
## cols(
##   species_name = col_character(),
##   a = col_double(),
##   b = col_double()
## )
```

```
dta_wide = dta_wide %>%
  left_join(RLS_MOZspp_a_b)

## Joining, by = "species_name"
```

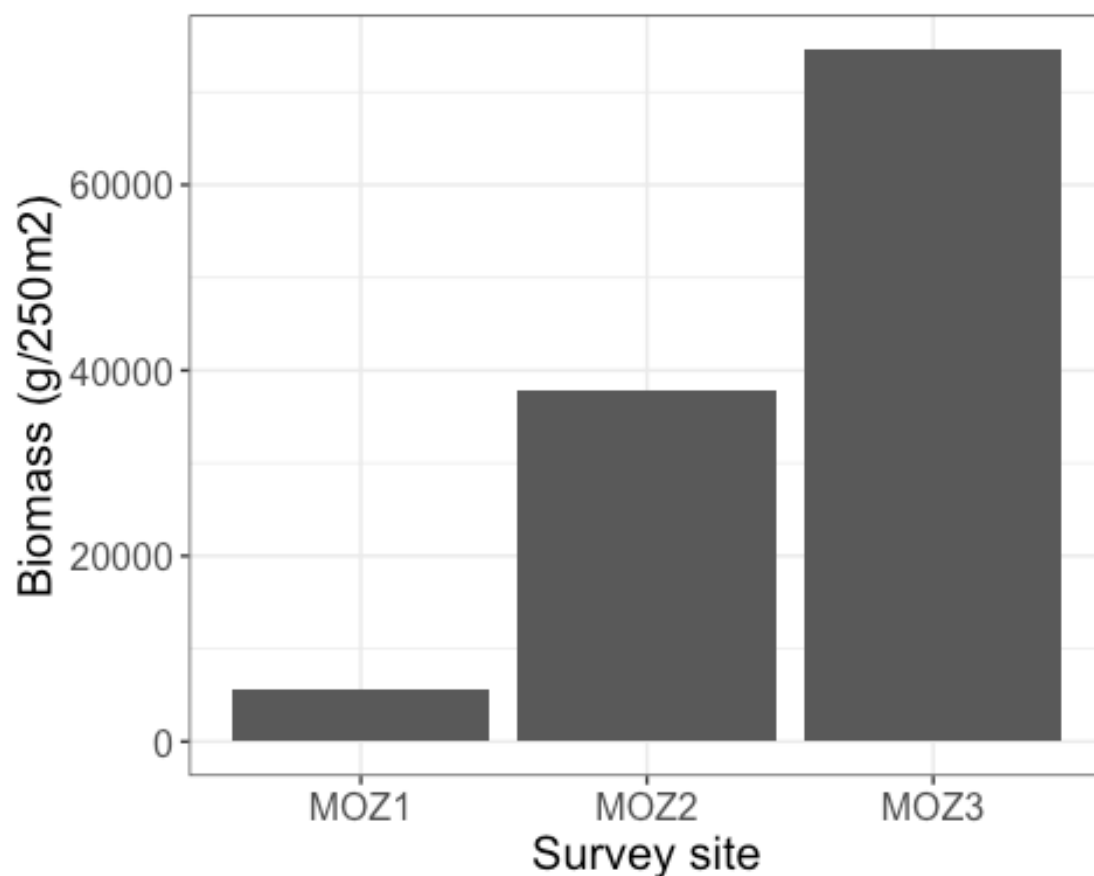
4) Calculate biomass

```
dta_wide = dta_wide %>%
  mutate(length = as.character(length),
         length = as.numeric(length),
         biomass = abundance*a*length^b)
```

5) Plot the total biomass of fish per site

```
bio = dta_wide %>%
  group_by(site_code) %>%
  summarise(biomass = sum(biomass))

ggplot(data = bio) +
  geom_bar(aes(x = site_code, y = biomass), stat = "identity") +
  labs(y = "Biomass (g/250m2)", x = "Survey site") +
  theme_bw() +
  theme(text = element_text(size = 15))
```



6) Plot the total biomass of fish greater than 10cm per site

```

bio = dta_wide %>%
  filter(length>20) %>%
  group_by(site_code) %>%
  summarise(biomass = sum(biomass))

ggplot(data = bio) +
  geom_bar(aes(x = site_code, y = biomass), stat = "identity") +
  labs(y = "Biomass (g/250m2)", x = "Survey site") +
  theme_bw() +
  theme(text = element_text(size = 15))

```

