

The difference between predator-size to prey-size

Jacob Passfield

1 Predators of Atlantic mackerel and shrimp

The aim of this document is to download a dataset and make some graphs based on it. After searching on <https://www.cefas.co.uk/data-and-publications/fish-stomach-records/> I came across two different species of prey with enough data to make some graphs. The two I chose were the Atlantic mackerel (*Scomber scombrus*) and caridean shrimps (*Caridea*) and so I downloaded the two `csv` files containing any record of these two prey being found in the stomach contents of their predators.

1.1 Preparing the data

First of all I load the following packages.

```
library(tidyverse)
library(ggplot2)
```

Then I read the data into R using the `read_csv()` function.

```
rawamack <- read_csv("_data/DapstomExportAMack.csv")
rawshrimps <- read_csv("_data/DapstomExportShrimps.csv")
```

After inspecting the data I decided I may be able to create something from the length of the predators and prey recorded. Because of this decision I used the `select()` function to create two new and refined tibbles including the variables of interest. Using the `rename()` function to change the variable names leads to easier coding.

```
#creating a new tibble for Atlantic mackerel
amack <- rawamack %>%
  rename("Predator" = "Predator.common.name", "PredatorMeanLength" =
    "Mean.length.of.predator", "Prey" = "Prey.common.name",
    "PreyLength" = "Prey.Length") %>%
  select(Predator, PredatorMeanLength, Prey, PreyLength)

#doing the same for caridean shrimp
shrimps <- rawshrimps %>%
  rename("Predator" = "Predator common name", "PredatorMeanLength" =
    "Mean length of predator", "Prey" = "Prey common name",
    "PreyLength" = "Prey Length") %>%
  select(Predator, PredatorMeanLength, Prey, PreyLength)
```

After experimentation I identified that the data types between the two tibbles did not match and so I ensured they did.

```
#checking whether data types match
sapply(shrimps, class)
```

```
##      Predator PredatorMeanLength      Prey      PreyLength
##      "character"      "numeric"      "character"      "character"
```

```
sapply(amack, class)
```

```
##      Predator PredatorMeanLength      Prey      PreyLength
##      "factor"      "numeric"      "factor"      "numeric"
```

```
#changing the data types
```

```
amack$Predator <- as.character(as.factor(amack$Predator))
amack$Prey <- as.character(as.factor(amack$Prey))
shrimps$PreyLength <- as.numeric(as.character(shrimps$PreyLength, na.rm=TRUE))
#and now they match
sapply(shrimps, class)
```

```
##      Predator PredatorMeanLength      Prey      PreyLength
##      "character"      "numeric"      "character"      "numeric"
```

```
sapply(amack, class)
```

```
##      Predator PredatorMeanLength      Prey      PreyLength
##      "character"      "numeric"      "character"      "numeric"
```

After this I checked the data for any errors or inconsistencies. First for Atlantic mackerel, then for

```
#changing the capital letters in the predators and prey columns to lowercase letters
amack <- amack %>%
```

```
  mutate(Predator = tolower(Predator), Prey = tolower(Prey))
#looking at the range and distribution of the numeric variables
summary(amack$PredatorMeanLength)
```

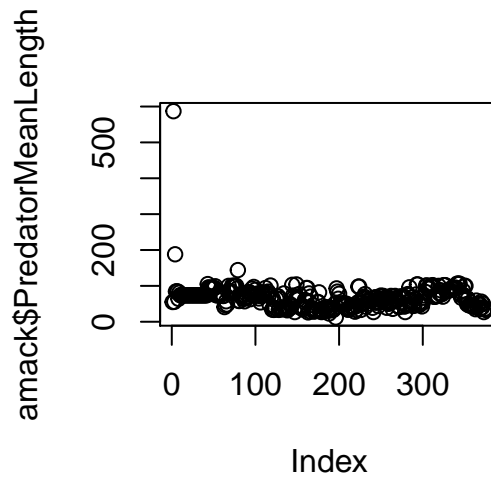
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      12.00  46.00  68.00  66.57  78.00  587.00         2
```

```
summary(amack$PreyLength)
```

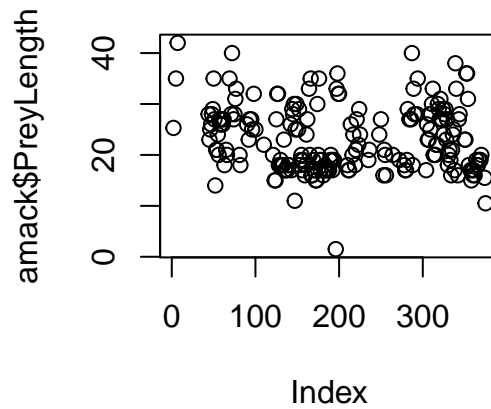
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      1.50  18.00  21.00  23.07  27.00  42.00      166
```

PreyLength looks okay but the value 587.00 for PredatorMeanLength looks a bit out of place.

```
#now visualising the data to identify errors
plot(amack$PredatorMeanLength)
```

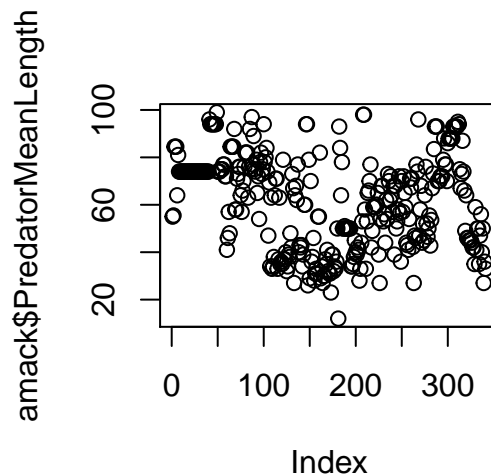


```
plot(amack$PreyLength)
```



For the sake of this task, I decided to remove this value.

```
#removing the error
amack <- amack %>%
  filter(PredatorMeanLength < 100)
plot(amack$PredatorMeanLength)
```



```
#doing the same with the caridean shrimps
```

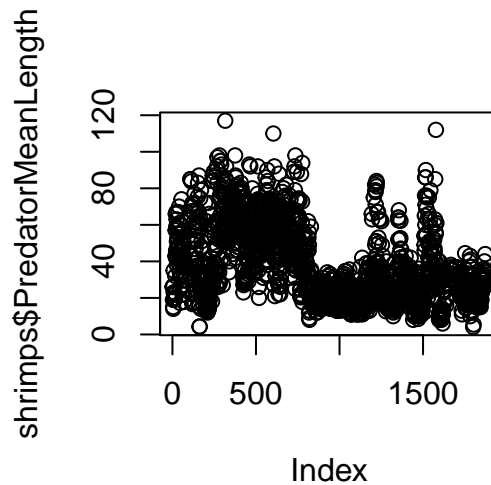
```
shrimps <- shrimps %>%  
  mutate(Predator = tolower(Predator), Prey = tolower(Prey))  
summary(shrimps$PredatorMeanLength)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      4.10  20.00   31.00   37.12  52.00  117.00     111
```

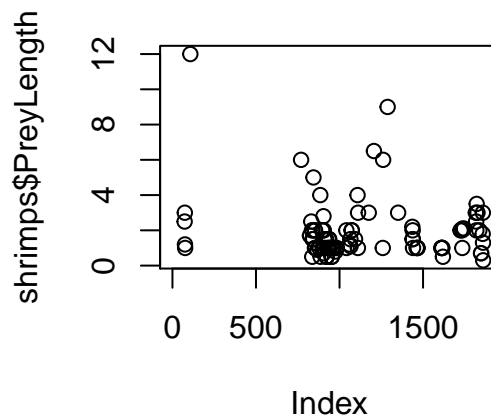
```
summary(shrimps$PreyLength)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      0.300  1.000   1.400   1.859  2.000  12.000     1772
```

```
plot(shrimps$PredatorMeanLength)
```

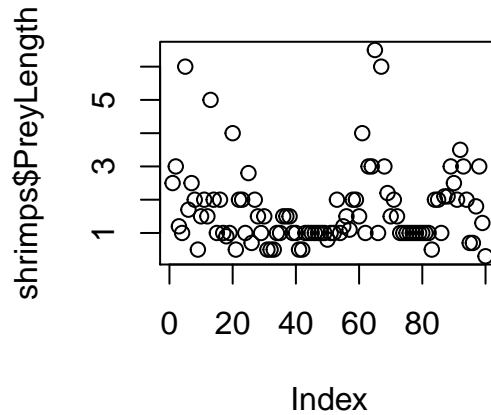


```
plot(shrimps$PreyLength)
```



```
#removing the abnormally high values
```

```
shrimps <- shrimps %>%  
  filter(PreyLength < 8)  
plot(shrimps$PreyLength)
```



I then decided to combine the two data frames into one tibble.

```
prey <- bind_rows(amack, shrimps)
```

1.2 Basic statistics

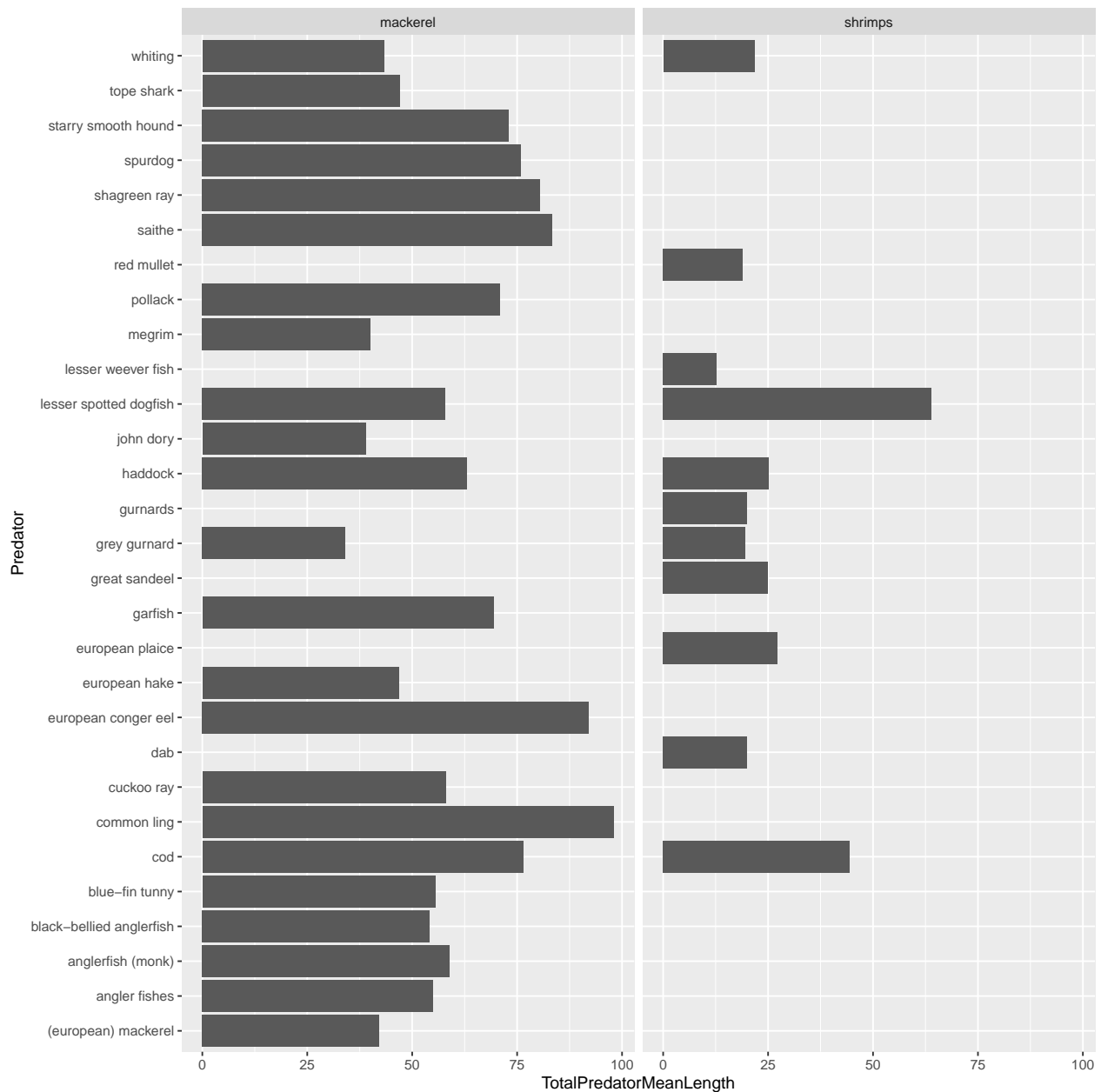
I thought of making a bar chart to show the difference between predators for the particular prey chosen.

To do this I created a new tibble which grouped all the data by predator whilst taking the mean length of all combined entries.

```
GroupedPredatorMeanLength <- prey %>%
  #grouping data by predator and prey
  group_by(Predator, Prey) %>%
  #summarising the total mean length of predator across all data
  summarise(TotalPredatorMeanLength = mean(PredatorMeanLength, na.rm=TRUE))
```

I then used this to create my box plot.

```
barchart1 <- ggplot(data = GroupedPredatorMeanLength,
  aes(x = Predator, y = TotalPredatorMeanLength)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ Prey) +
  coord_flip()
barchart1
```



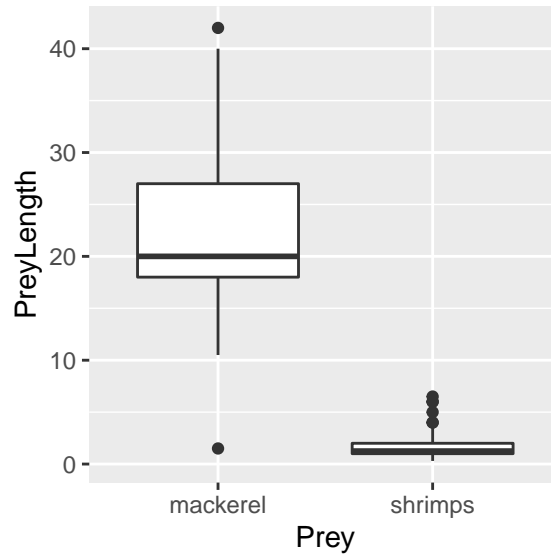
Here I realised this bar chart doesn't really compare predator-size with prey-size. What we can see however is that mackerel has many more predators than shrimp.

I tried to add a line showing the mean length of the prey using `abline()` but it didn't show up.

At this point I remembered that boxplots are a great comparison tool and so I made some.

```
preyboxplot1 <- ggplot(data = prey,
  aes(x=Prey, y=PreyLength)) +
  geom_boxplot()
preyboxplot1
```

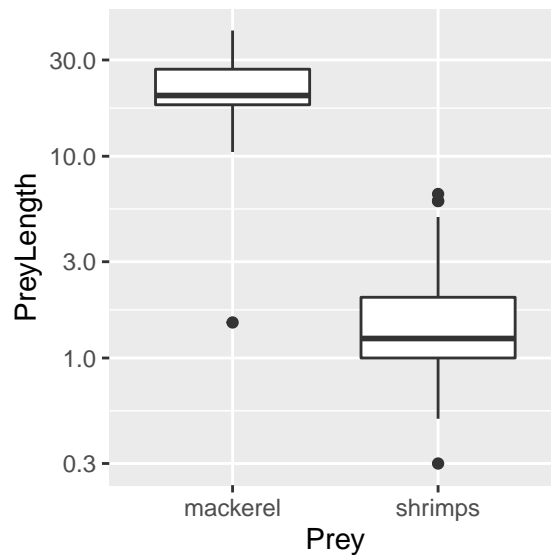
```
## Warning: Removed 157 rows containing non-finite values (stat_boxplot).
```



Seeing the shrimp boxplot too small to distinguish any value, I used `scale_y_log10()`.

```
preyboxplot2 <- ggplot(data = prey,
  aes(x=Prey, y=PreyLength)) +
  geom_boxplot() +
  scale_y_log10()
preyboxplot2
```

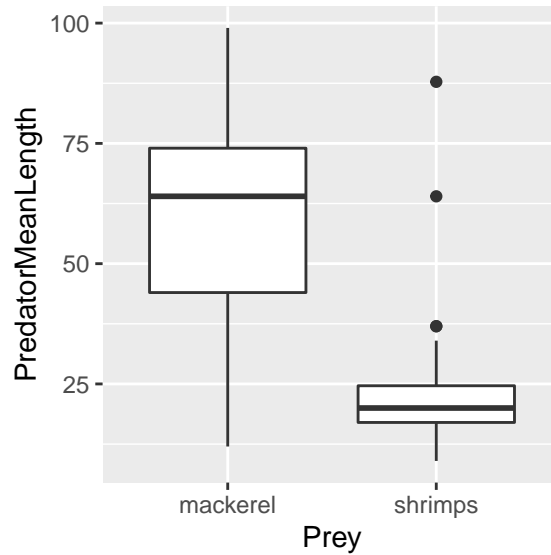
Warning: Removed 157 rows containing non-finite values (stat_boxplot).



It is clear that mackerel are on average bigger than shrimp.

The next boxplot I made was to comparing the predator length with my two prey.

```
predatorboxplot <- ggplot(data = prey,
  aes(x=Prey, y=PredatorMeanLength)) +
  geom_boxplot()
predatorboxplot
```



Here it is evident that mackerel usually have bigger predators compared to shrimp but it is also clear that they can get eaten by the same size predators on occassion.

2 Questions/queries I have picked up along the way

1. Should I be doing the optional assignments?
 - 1a. Will our final report be written in RMarkdown?
2. I had trouble understanding what the data was showing, for example, the units and the variables.
3. How do I convert a `.txt` file to `csv`? In reponse to *Barnes et al.*
4. I had this error: your branch is ahead of 'origin/main' by 4 commits. I resolved the issue but I wonder what could've caused it.

Checking to see if I still remember how to push to GitHub.