

Firstly, I will try to get the prey selection coefficients from something easy I am going to try anchovy first.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.3.3
```

```
#reading in the data
```

```
load("C:/Users/lucab/Downloads/stomach_dataset.Rdata")
```

```
anchovy <- stom_df%>%filter(pred_species=="Engraulis encrasicolus")
```

```
#okay there arent very many datarows, so I will use another dataset as well
```

```
anc <- read.csv("C:/Users/lucab/Downloads/AnchovyAndSardine.csv")
```

```
anc <- anc%>%filter(Species=="anchovy")
```

```
#I will combine the two datasets, but i need to get the same format
```

```
anchovy <- anchovy%>%select(pre_y_weight_g, pred_weight_g, nprey_perpred, pred_species)%>%  
rename(wprey=pre_y_weight_g, wpredator=pred_weight_g, Nprey=nprey_perpred, Species=pred_species)%>%  
mutate(wprey=wprey/Nprey)
```

```
anchovy <- rbind(anchovy, anc)
```

```
#here is the final dataset to play with,
```

```
#lets add ppmr columns
```

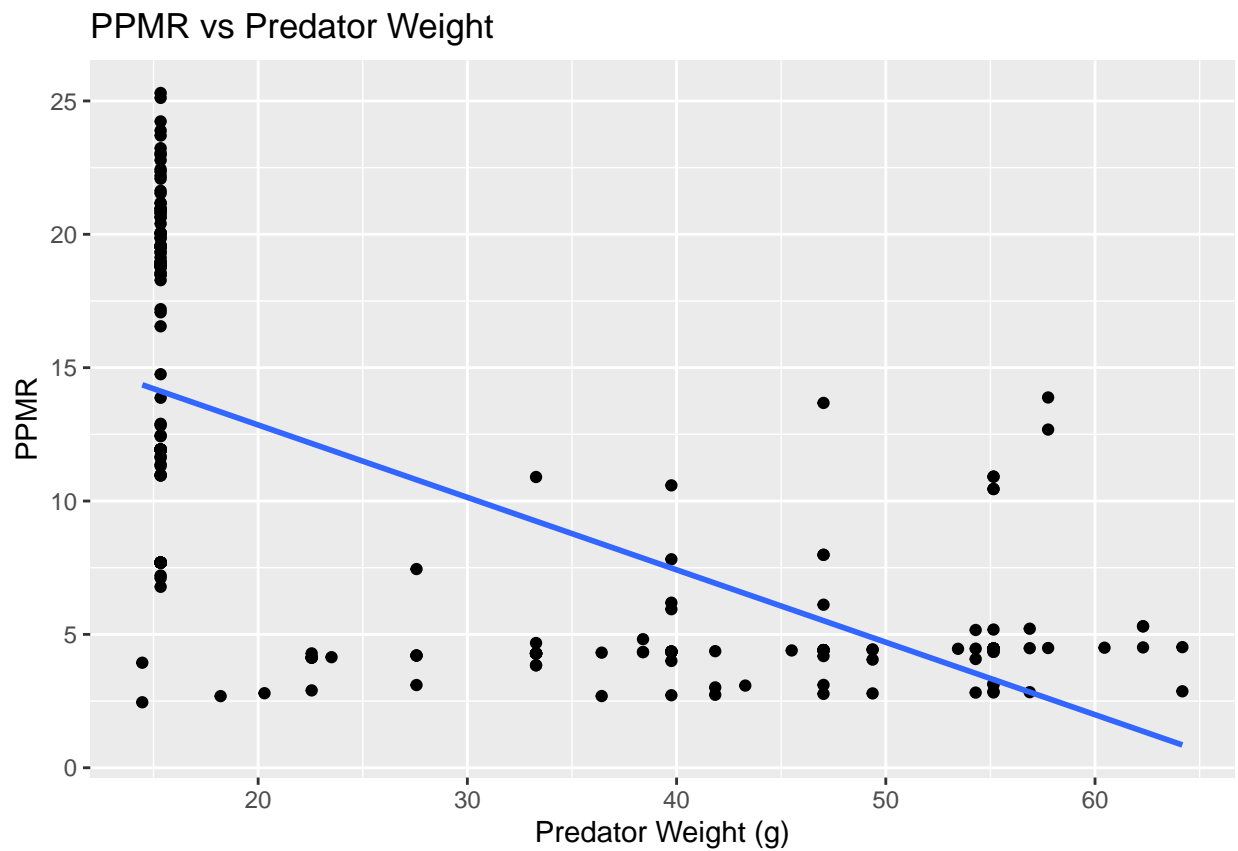
```
anchovy <- anchovy%>%mutate(ppmr=(wpredator/wprey))
```

Now we have the dataset, first we should check that ppmr doesnt change across predator weights

```
#plotting the ppmr against predator weight
```

```
ggplot(anchovy, aes(x=wpredator, y=log(ppmr)))+  
  geom_point()+  
  geom_smooth(method="lm", se=FALSE)+  
  labs(title="PPMR vs Predator Weight", x="Predator Weight (g)", y="PPMR")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



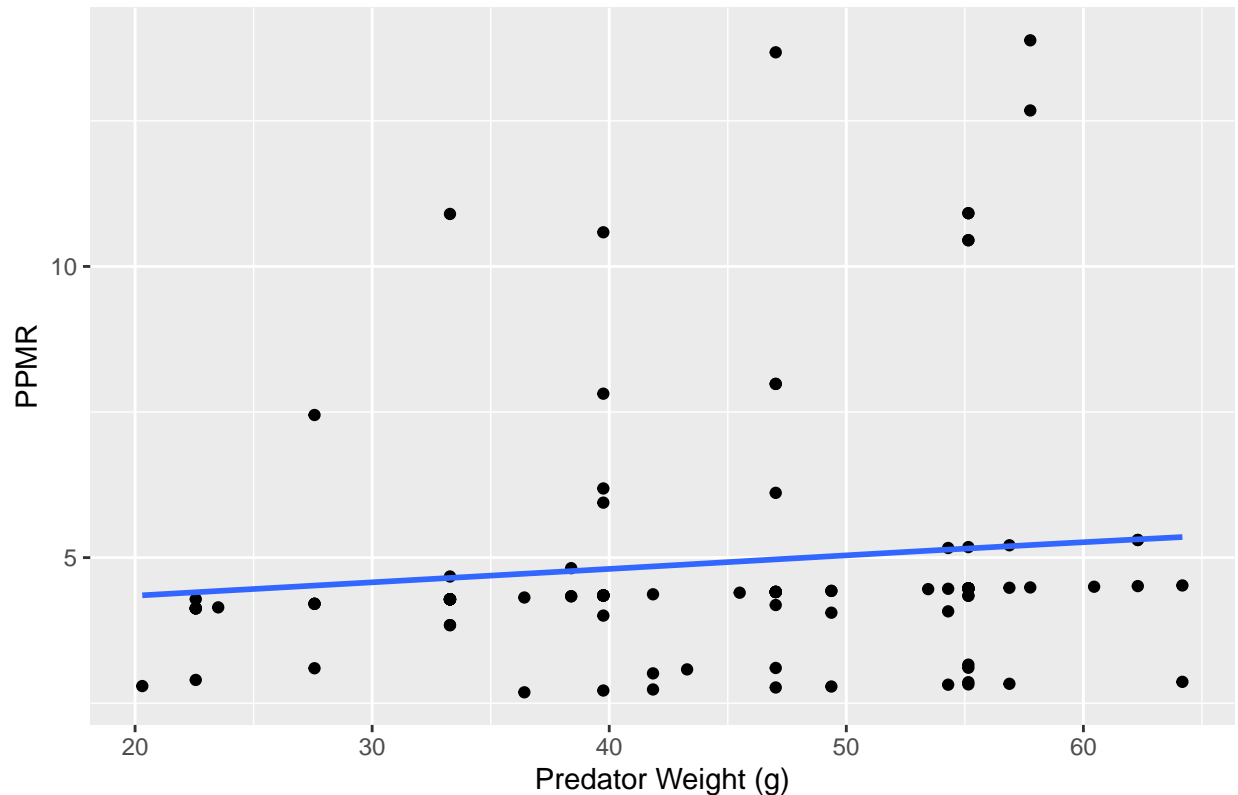
```
#the lowest sizes of anchovy apparently eat prey that is e^-10, which is very small,  
#so i will filter this out
```

```
#going to pplot only from 20g onwards
```

```
ggplot(anchovy%>%filter(wpredator>20), aes(x=wpredator, y=log(ppmr)))+  
  geom_point()+  
  geom_smooth(method="gam", se=FALSE)+  
  labs(title="PPMR vs Predator Weight", x="Predator Weight (g)", y="PPMR")
```

```
## 'geom_smooth()' using formula = 'y ~ s(x, bs = "cs")'
```

PPMR vs Predator Weight



*#this looks a lot better now (i dont know when to use lm or gam, but gam looks better)
#therefore, we can say that ppmr stays the same across predator weights, so we can aggregate predator w*

It is interesting that we do not have to account for the biomass of the prey to get an equal ppmr across predator weights, maybe because anchovy already eats very small prey, and there isnt a great size variation.

Next step is to plot this all as a density distribution

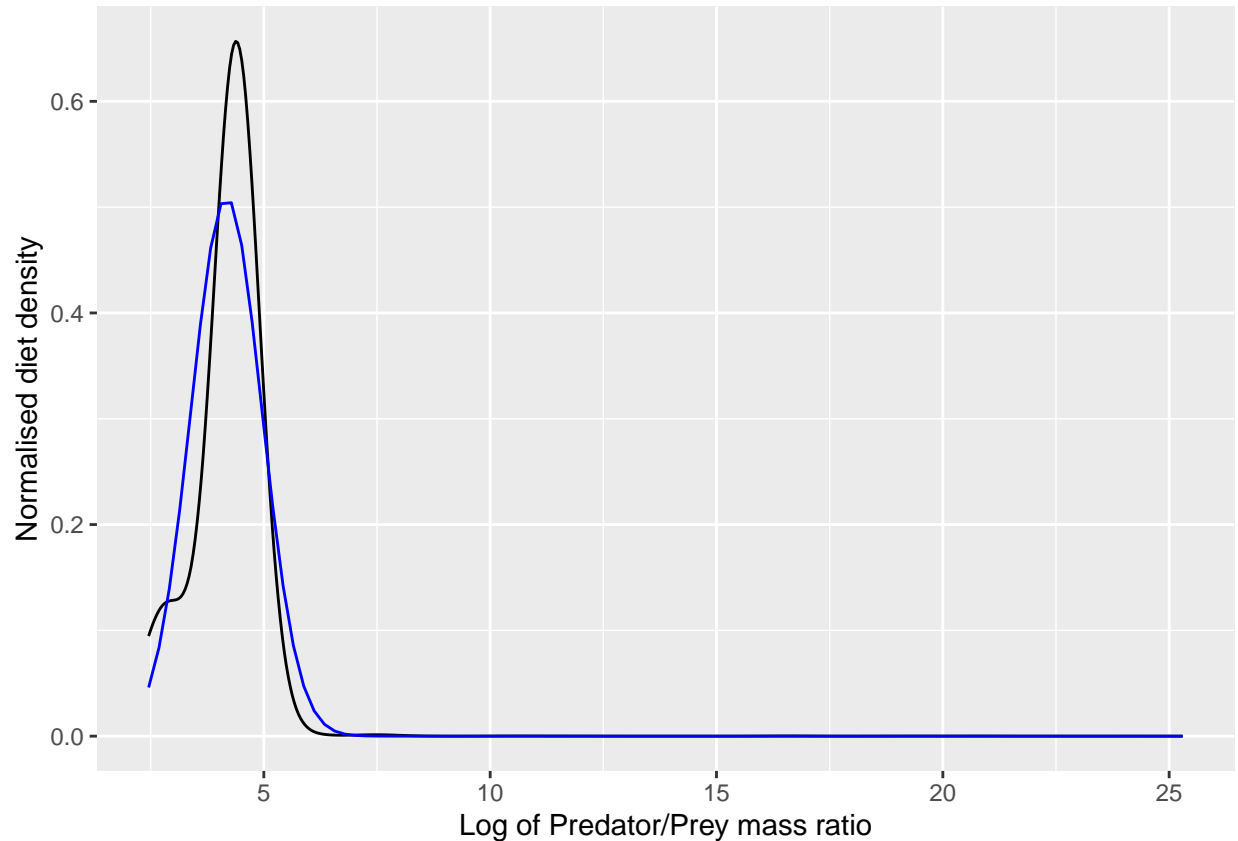
```
#defining the digestion coefficient
dig <- 2/3

weighted.sd <- function(x, w) { sqrt(sum(w * (x - weighted.mean(x, w))^2)) }

weight <- anchovy$Nprey * (anchovy$wprey/anchovy$wpredator)^dig
weight <- weight / sum(weight)

est_mean <- weighted.mean(log(anchovy$ppmr), weight)
est_sd <- weighted.sd(log(anchovy$ppmr), weight)

ggplot(anchovy) +
  geom_density(aes(log(ppmr), weight = Nprey*wprey^dig), bw = 0.5) +
  stat_function(fun = dnorm, args = list(mean = est_mean, sd = est_sd), colour = "blue") +
  xlab("Log of Predator/Prey mass ratio") + ylab("Normalised diet density")
```



Ok, the distribution for it seems to fit. it needs a slightly lower variance however. I will now check it for the number distribution

```
stomach <- anchovy

grid <- seq(0, max(log(stomach$ppmr)), length = 100)

weighted.sd <- function(x, w) { sqrt(sum(w * (x - weighted.mean(x, w))^2)) }

weight <- anchovy$Nprey * (anchovy$wprey/anchovy$wpredator)^dig
weight <- weight / sum(weight)

n_prey <- anchovy$Nprey
weight_numbers = n_prey / sum(n_prey)

est_mean <- weighted.mean(log(stomach$ppmr), weight)
est_sd <- weighted.sd(log(stomach$ppmr), weight)

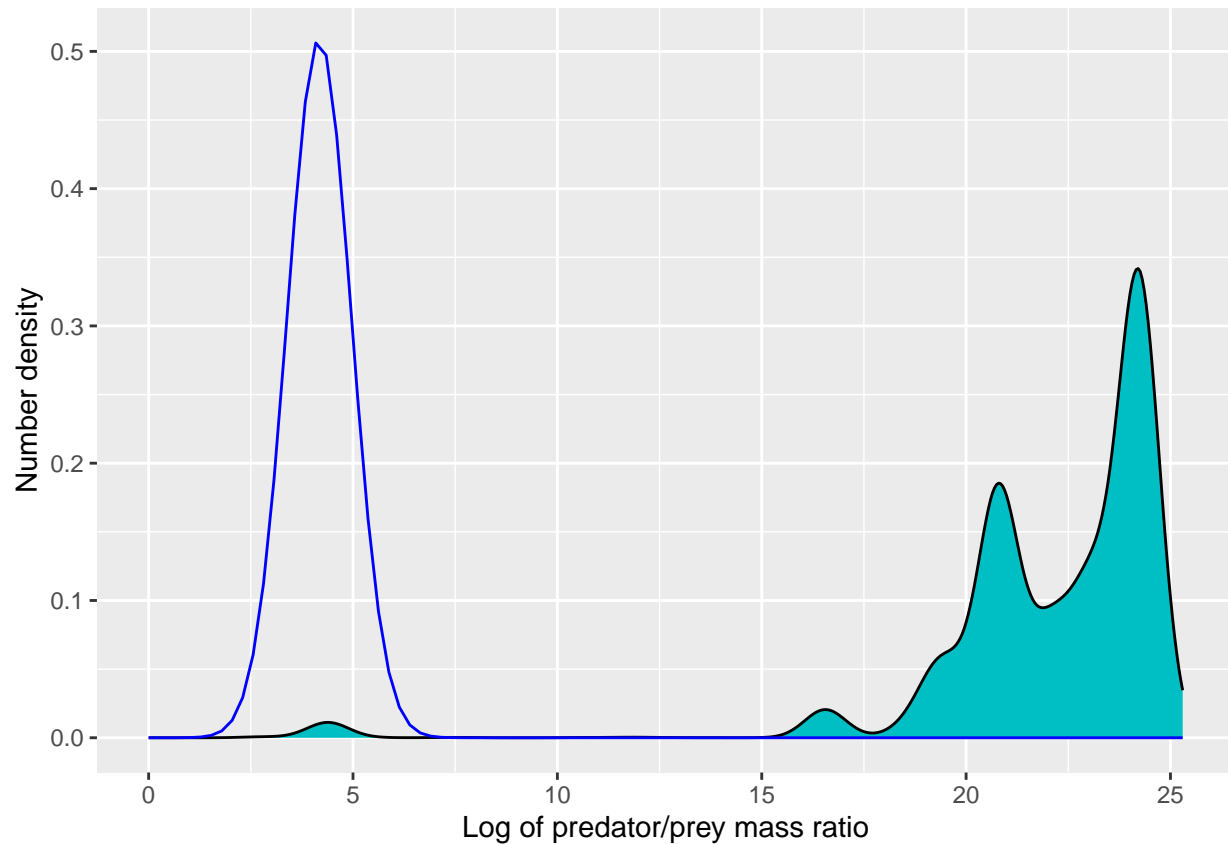
shifted_normaldens <- data.frame(
  log_ppmr = grid,
  density = dnorm(grid, est_mean, est_sd)
)

ggplot(stomach) +
  geom_density(aes(log(ppmr), weight = weight_numbers),
```

```

    fill = "#00BFC4", bw = 0.5) +
  xlab("Log of predator/prey mass ratio") +
  ylab("Number density") +
  geom_line(aes(log_ppmr, density), data = shifted_normaldens,
    colour = "blue")

```



Okay, the distribution fit here isn't good. But when I do the same without including the data from the anchovy and sardines file, it looks a lot better.

I am going to plot them now separately.

```

#I am going to plot both, but as separate plots

#need to match the column names to this data.

stomach <- anchovy

colnames(stomach) <- c("w_pre", "w_pred", "n_pre", "pred_species", "log_ppmr")

stomach <- stomach%>%mutate(log_ppmr=log(w_pred/w_pre))

grid <- seq(0, max(stomach$log_ppmr), length = 100)
normaldens <- plyr::ddply(stomach, "pred_species", function(stomach) {
  weight <- stomach$n_pre * (stomach$w_pre / stomach$w_pred)^dig
  weight <- weight / sum(weight)

```

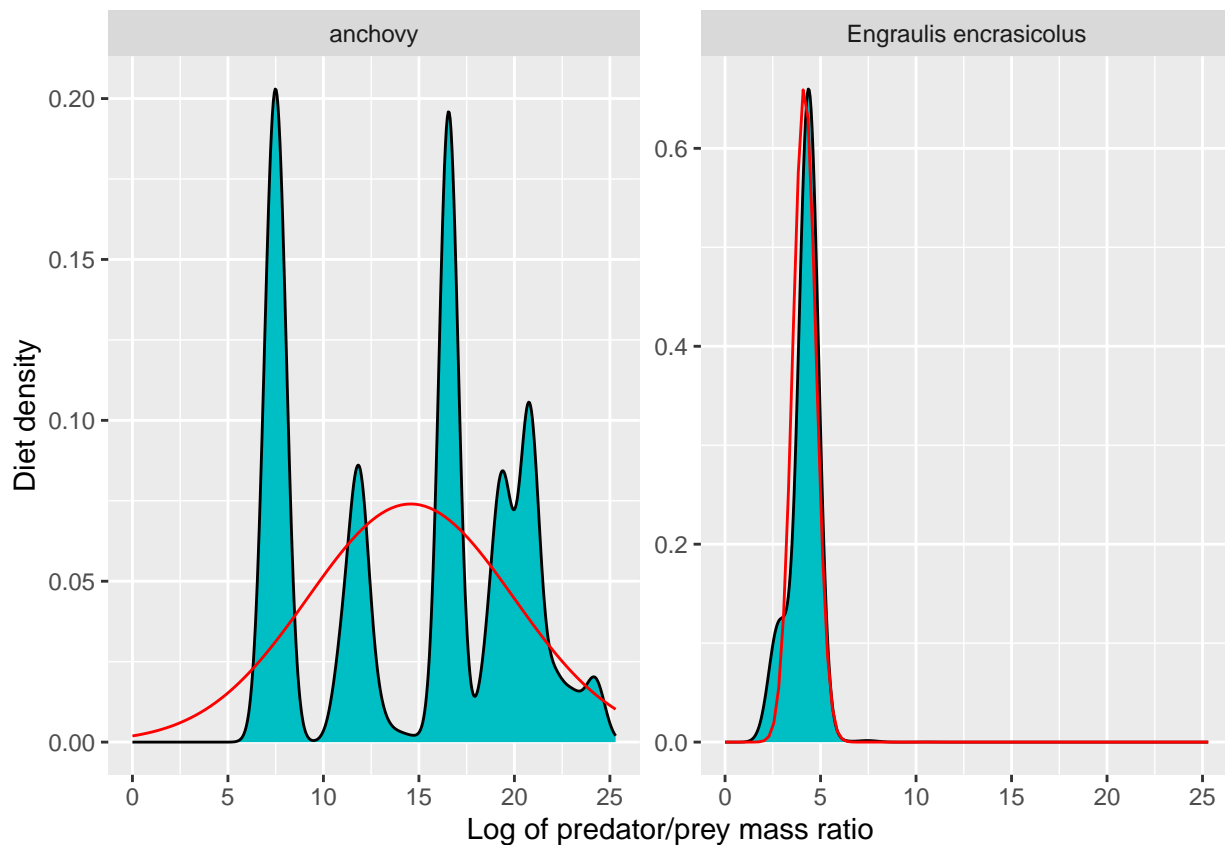
```

est_mean <- weighted.mean(stomach$log_ppmr, weight)
est_sd <- weighted.sd(stomach$log_ppmr, weight)
data.frame(
  log_ppmr = grid,
  density = dnorm(grid, est_mean, est_sd)
)
})

stomach <- stomach |>
  mutate(weight_diet = (n_prey * (w_prey / w_pred)^dig) /
    sum(n_prey * (w_prey / w_pred)^dig))

ggplot(stomach) +
  geom_density(aes(log_ppmr, weight = weight_diet),
    fill = "#00BFC4", bw = 0.5) +
  facet_wrap(~pred_species, scales = "free_y", ncol = 4) +
  xlab("Log of predator/prey mass ratio") +
  ylab("Diet density")+
  geom_line(aes(log_ppmr, density), data = normaldens,
    colour = "red")

```



and now doing the same for the number density

```

shifted_normaldens <- plyr::ddply(stomach, "pred_species", function(stomach) {
  weight <- stomach$n_prey * (stomach$w_prey / stomach$w_pred)^dig
  weight <- weight / sum(weight)

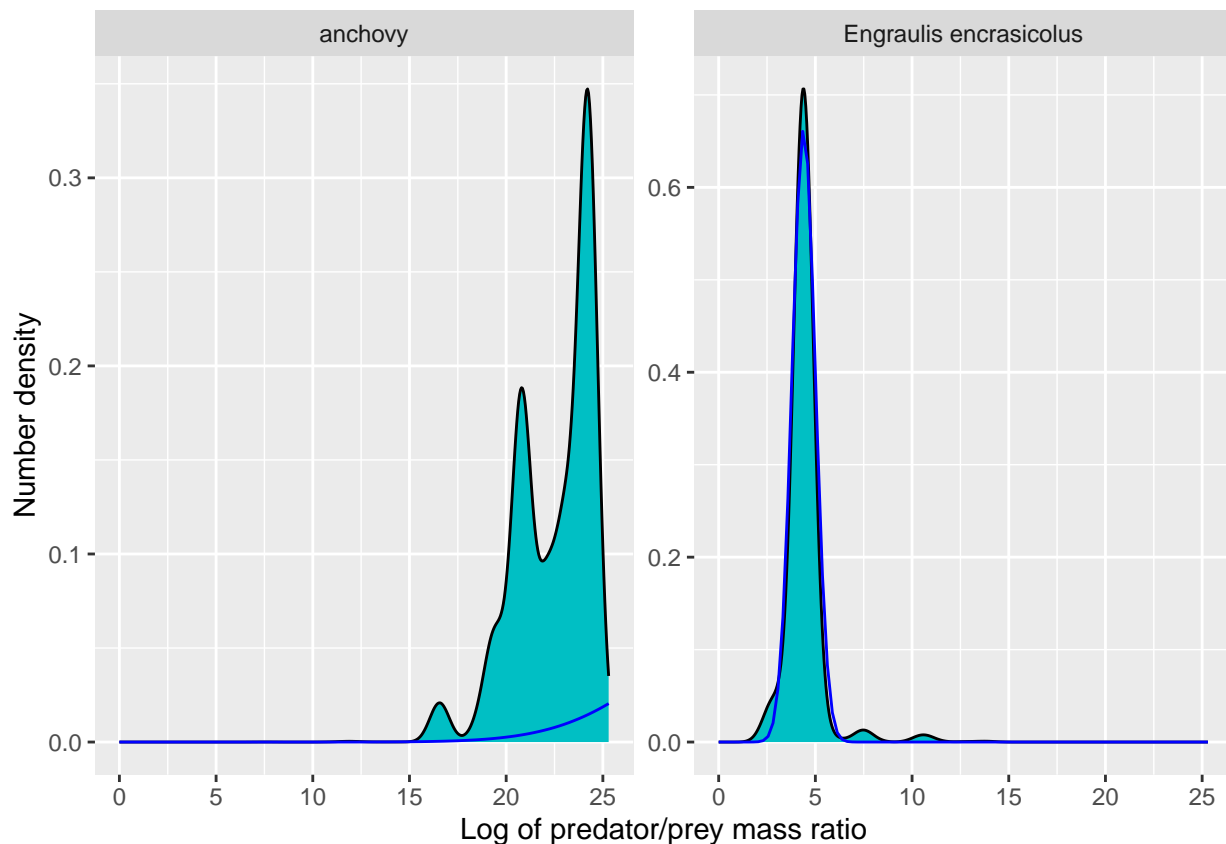
```

```

est_mean <- weighted.mean(stomach$log_ppmr, weight)
est_sd <- weighted.sd(stomach$log_ppmr, weight)
data.frame(
  log_ppmr = grid,
  density = dnorm(grid, est_mean+dig*est_sd^2, est_sd)
)
})

ggplot(stomach) +
  geom_density(aes(log_ppmr, weight = weight_numbers),
    fill = "#00BFC4", bw = 0.5) +
  facet_wrap(~pred_species, scales = "free_y", ncol = 4) +
  xlab("Log of predator/prey mass ratio") +
  ylab("Number density") +
  geom_line(aes(log_ppmr, density), data = shifted_normaldens,
    colour = "blue")

```



It works very well for the stom_df dataset, but not at all for the anchovy dataset. Not sure why.

I think the problem is that Nprey from the anchovy dataset is the probability of the prey item to be in a stomach, whereas this code assumes that Nprey is the number of prey found for those sizes of pred and prey. - thinking about it this shouldn't matter, as it is both a weight of number of observations. - something else must be up with this data.

So I am not sure how to convert this to what we want (will have to redo code to weight it in that fashion)

For now I am just going to use the stomach dataset, and this next section I will run a mixed effects linear

model to see the variation not due to region/year

```
#setting up the model

anchovy <- stom_df%>%filter(pred_species=="Engraulis encrasicolus")%>%
  mutate(wprey=prey_weight_g/nprey_perpred)

#model <- lmer(data=anchovy, wprey~(1|ices_rectangle))

#ok to this model states that 99% percent of the wprey is not due to the region or year
#therefore it must all be due to the predator selection
```

I think we need more data, as currently there isn't too much, and if we can figure out how to use the anchovies and sardines datafile, we can include that too, but currently it completely different.

Still, the normal distribution fits the anchovy predation kernel very well, so I think we can use this.