Kernel - Sprat

2024-07-12

```r
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
```

```r
library(ggplot2)
```

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

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

sprat <- stom_df%>%filter(pred_taxa=="Sprattus sprattus")
```

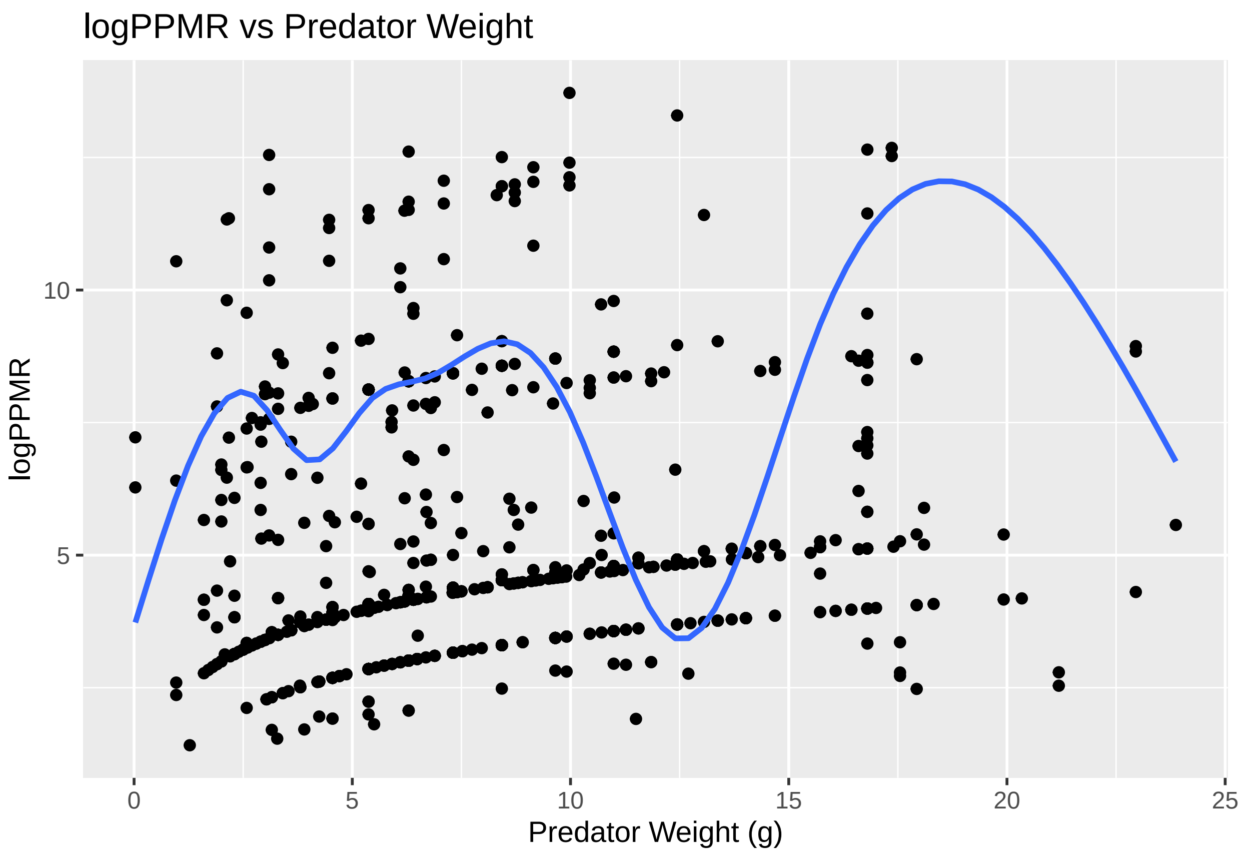
First lets check the PPMR predator size distribution.

```r
ggplot(sprat, aes(x=pred_weight_g, y=log(ppmr)))+
  geom_point()+
  geom_smooth(method="gam", se=FALSE, aes(weight = nprey_perpred))+
  labs(title="logPPMR vs Predator Weight", x="Predator Weight (g)", y="logPPMR")
```

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

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_smooth()‘).
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```



Not good. Will now plot the same but with the diet weighting

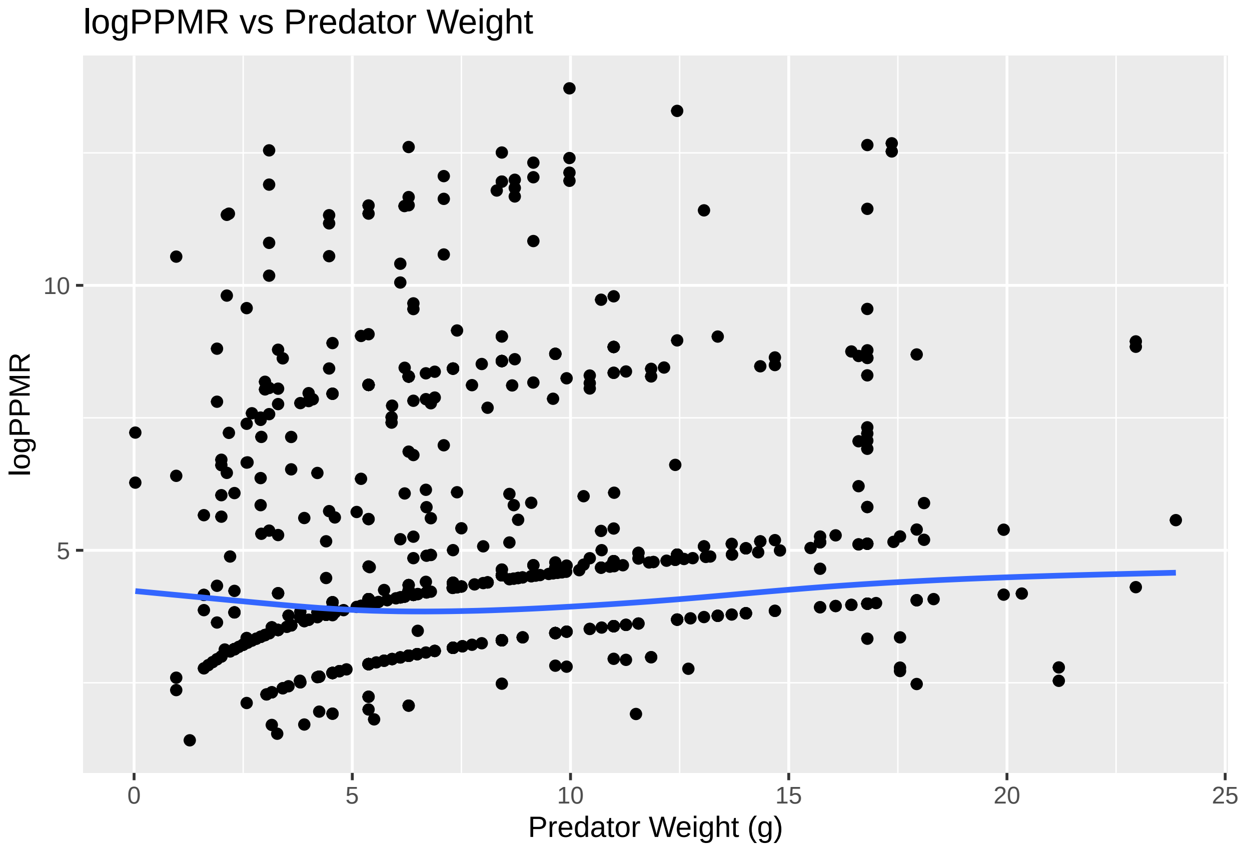
```
dig <- 2/3

ggplot(sprat, aes(x=pred_weight_g, y=log(ppmr)))+
  geom_point()+
  geom_smooth(method="gam", se=FALSE, aes(weight = nprey_perpred*sprat$prey_ind_weight_g^dig))+
  labs(title="logPPMR vs Predator Weight", x="Predator Weight (g)", y="logPPMR")
```

```
## Warning: Use of ‘sprat$prey_ind_weight_g‘ is discouraged.
## i Use ‘prey_ind_weight_g‘ instead.
```

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

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_smooth()‘).
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```

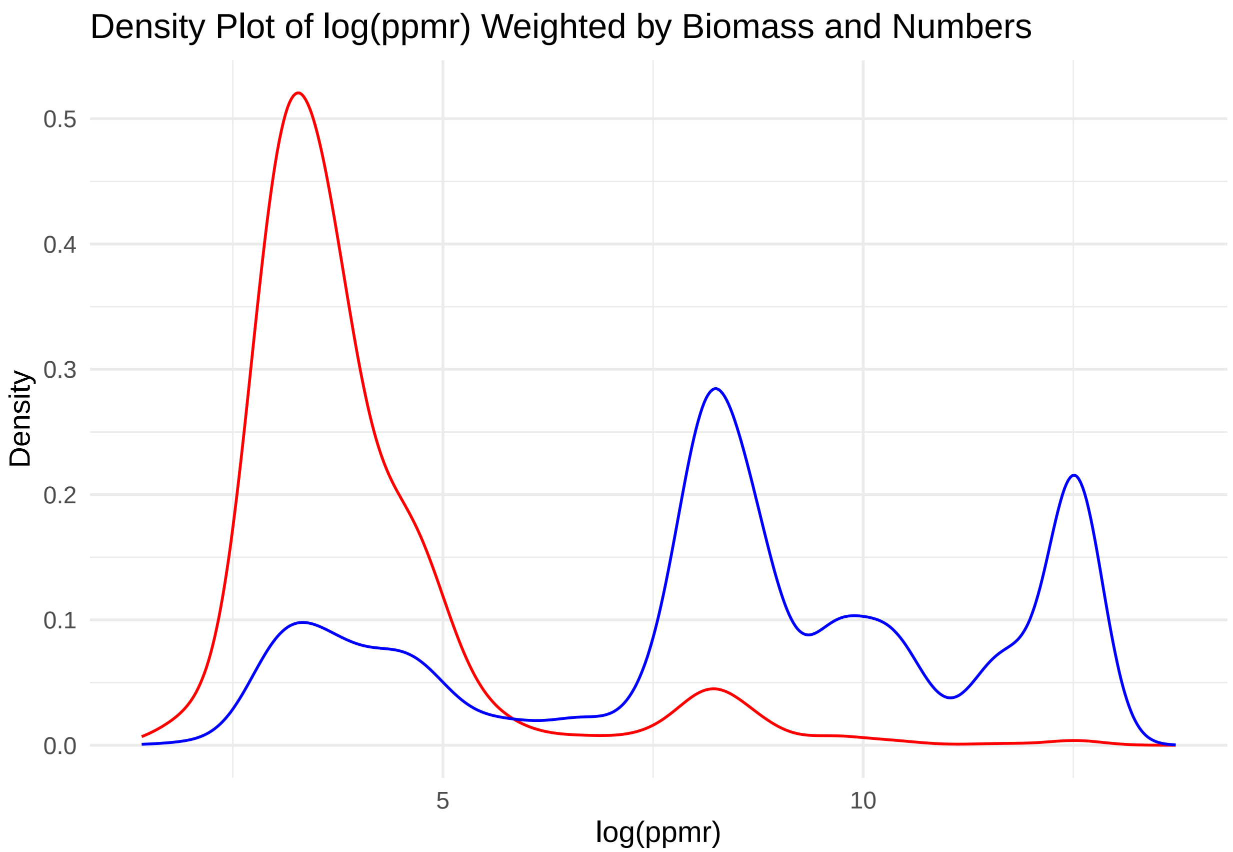


When looking at the contribution to the diet, the PPMR does not change across predator sizes.

So first, lets lot the density to see what might best fit.

```
sprat$weight_numbers <- sprat$nprey_perpred
sprat$weight_biomass <- sprat$nprey_perpred*sprat$prey_ind_weight_g^dig

ggplot() +
  geom_density(data = sprat, aes(x = log(ppmr), weight = weight_biomass), color = "red") +
  geom_density(data = sprat, aes(x = log(ppmr), weight = weight_numbers), color = "blue") +
  labs(title = "Density Plot of log(ppmr) Weighted by Biomass and Numbers",
       x = "log(ppmr)",
       y = "Density") +
  theme_minimal()
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_density()‘).
## Removed 6 rows containing non-finite outside the scale range
## (‘stat_density()‘).
```



Not sure which fit would be best here. It seems to be triple modal feeding patterns, but only one of the modes contributes to the diet.

As sprat should be planktivorous, our data should probably be recent so that the the technology allows for.

I will now plot the PPMR/pred weight scatterplot but with the years plotted also.

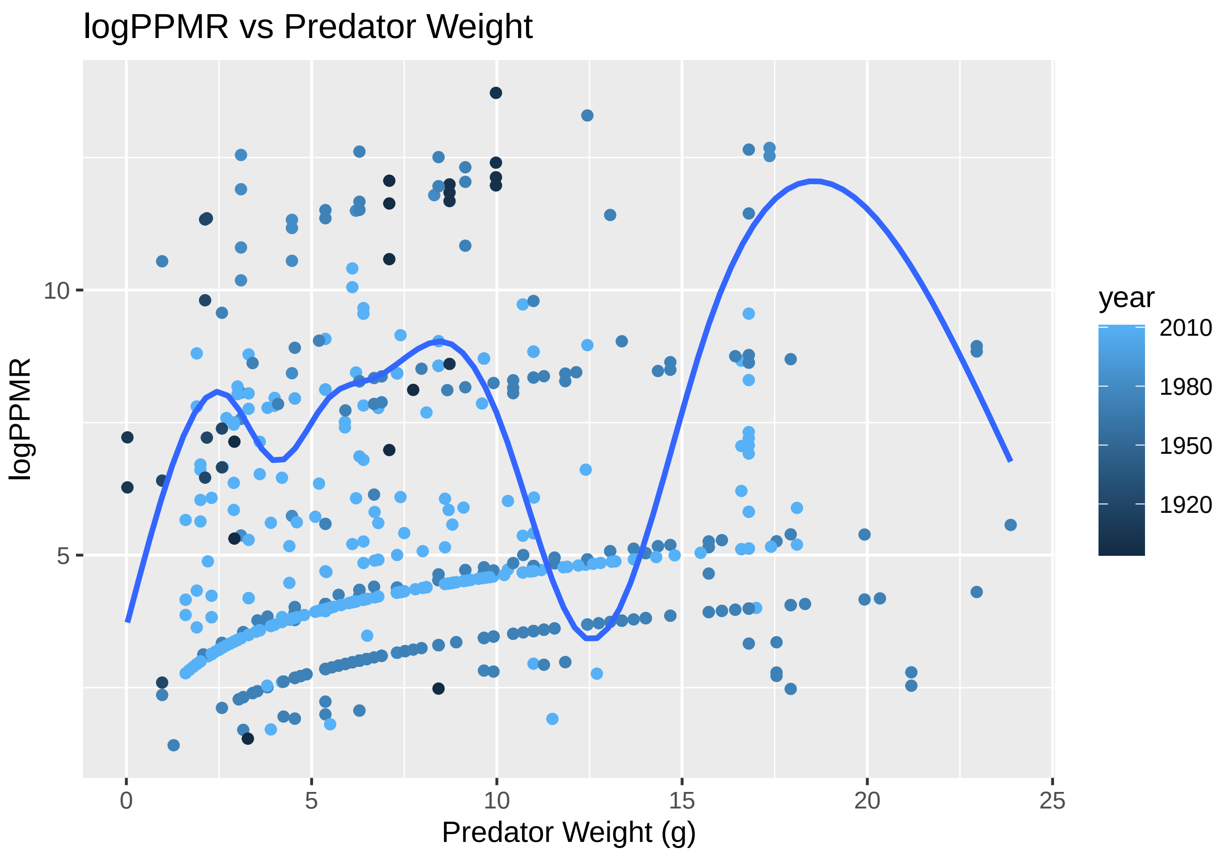
```r
ggplot(sprat, aes(x=pred_weight_g, y=log(ppmr), color=year))+
  geom_point()+
  geom_smooth(method="gam", se=FALSE, aes(weight = nprey_perpred))+
  labs(title="logPPMR vs Predator Weight", x="Predator Weight (g)", y="logPPMR")
```

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

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



There doesn't seem to be a pattern in the years and PPMRs taken, so I assume that we can use all the data.

Regardless, this is difficult to model, potentially, a triple mixture gaussian model may fit each of the 3 feeding modes. From some light googling, sprat has seasonal feeding patterns, feeding on fish, zooplankton and small crustaceans. This may be the 3 feeding modes that we see.

Additionally, sprat moves to the surface at night, so one of the feeding modes may be what sprat can find in the day, the other mode what it can find at night, and the third feeding mode what it can find out of season?

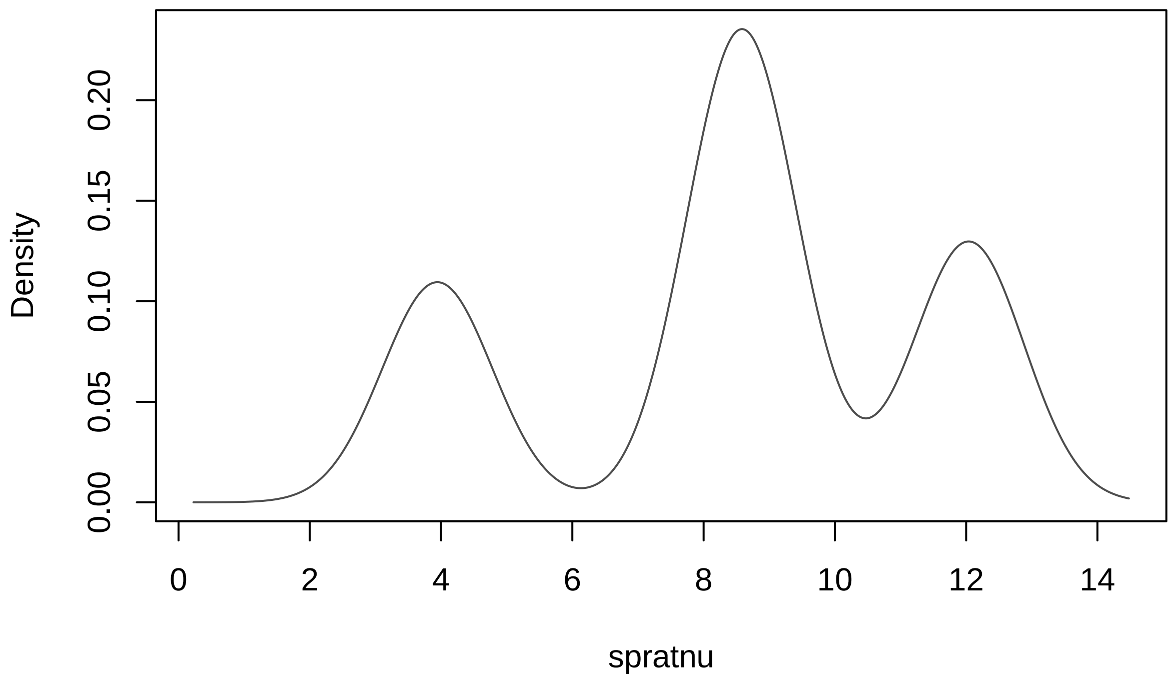
First, lets try the gaussian mixture model with 3 modes.

```
library(mclust)
```

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

```
## Package 'mclust' version 6.1.1
## Type 'citation("mclust")' for citing this R package in publications.
```

```
repeat_elements <- function(data, weights) {

    valid_indices <- !is.na(data) & !is.na(weights)
  data <- data[valid_indices]
  weights <- weights[valid_indices]

  final_vector <- c()
  for (i in seq_along(data)) {

    rounded_weight <- round(weights[i])

    repeated_values <- rep(data[i], times = rounded_weight)

    final_vector <- c(final_vector, repeated_values)
  }

  return(final_vector)
}

spratnu <- repeat_elements(log(sprat$ppmr), sprat$nprey_perpred)

gmm <- densityMclust(spratnu, G=3)
```

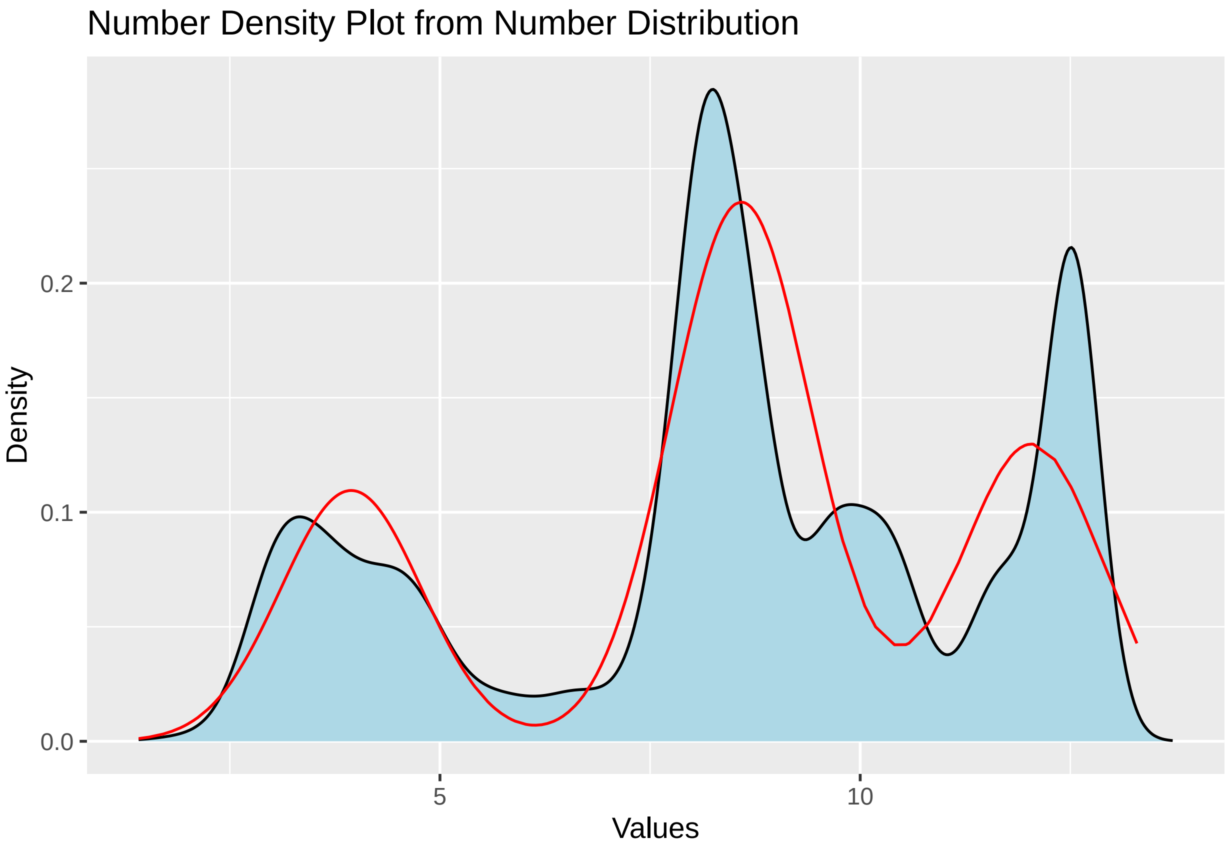


```
ngmm <- gmm
```

Ok, now lets plot it over the number density.

```
dplot <- data.frame(x=gmm[["data"]], density=gmm[["density"]])


(numbfit <- ggplot() +
    geom_density(data=sprat, aes(log(ppmr), weight=weight_numbers), fill="lightblue")+
  geom_line(data=dplot, aes(x = x, y = density), color="red") +
  labs(x = "Values", y = "Density") +
  ggtitle("Number Density Plot from Number Distribution"))
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_density()‘).
```

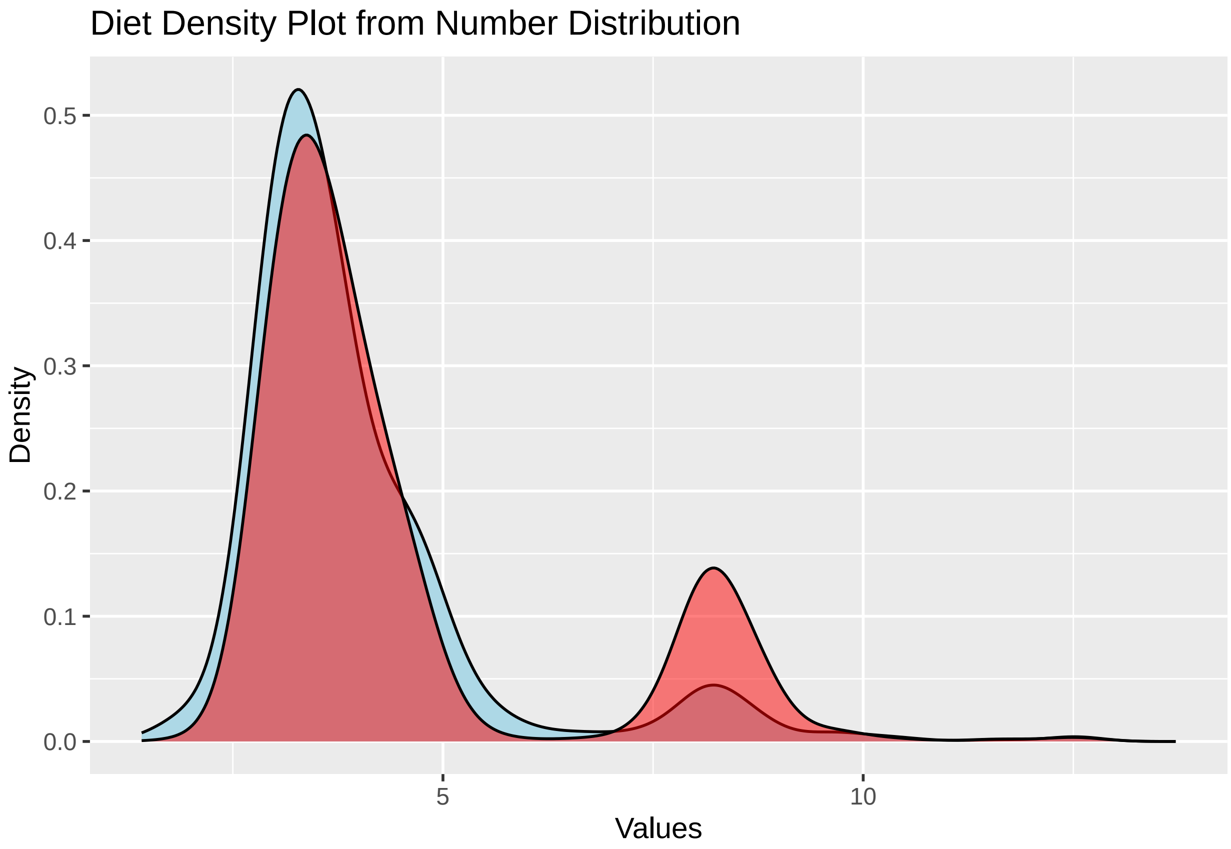


This is not bad, this works fairly well for how hard it is to model.

But will it move across to the biomass distribution well?

```
#sorting out the column names and cleaning the data
sprat$l <- log(sprat$ppmr)
sprat <- sprat[!is.na(sprat$l),]
x_vals <- seq(min(sprat$l), max(sprat$l), length.out = 1000)

#now creating the new density

#shifting and normalising the new density
newspratdens <- gmm[["density"]]*exp(-dig*gmm[["data"]])
shifted_pdf_normalized <- newspratdens / sum(newspratdens)

#making new dataframe
dplot <- data.frame(x=gmm[["data"]], density=shifted_pdf_normalized)
```

```
(numbfitbio <- ggplot() +
    geom_density(data=sprat, aes(log(ppmr), weight=weight_biomass), fill="lightblue")+
    geom_density(data=dplot, aes(x, weight=shifted_pdf_normalized), fill="red", alpha=0.5)+
  #geom_line(data=dplot, aes(x = x, y = density), color="red") +
  labs(x = "Values", y = "Density") +
  ggtitle("Diet Density Plot from Number Distribution"))
```

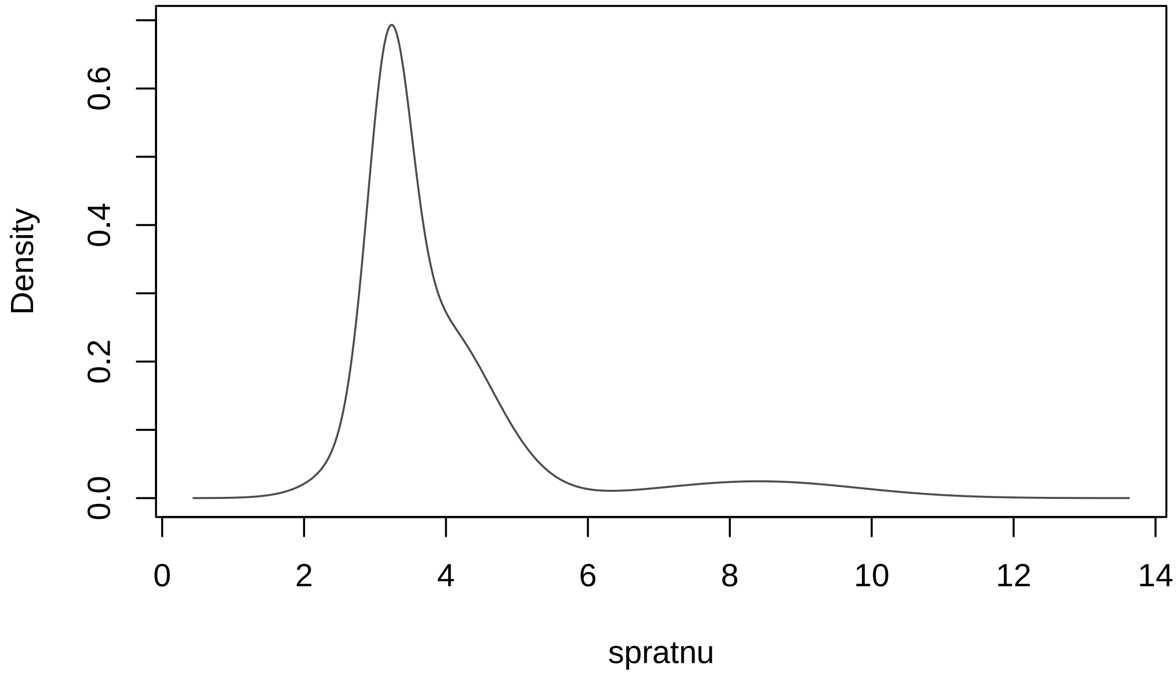


Red is the shifted density from fitting the model to the number distribution, blue is the plotted density from the sprat data, when weighting to diet contribution.

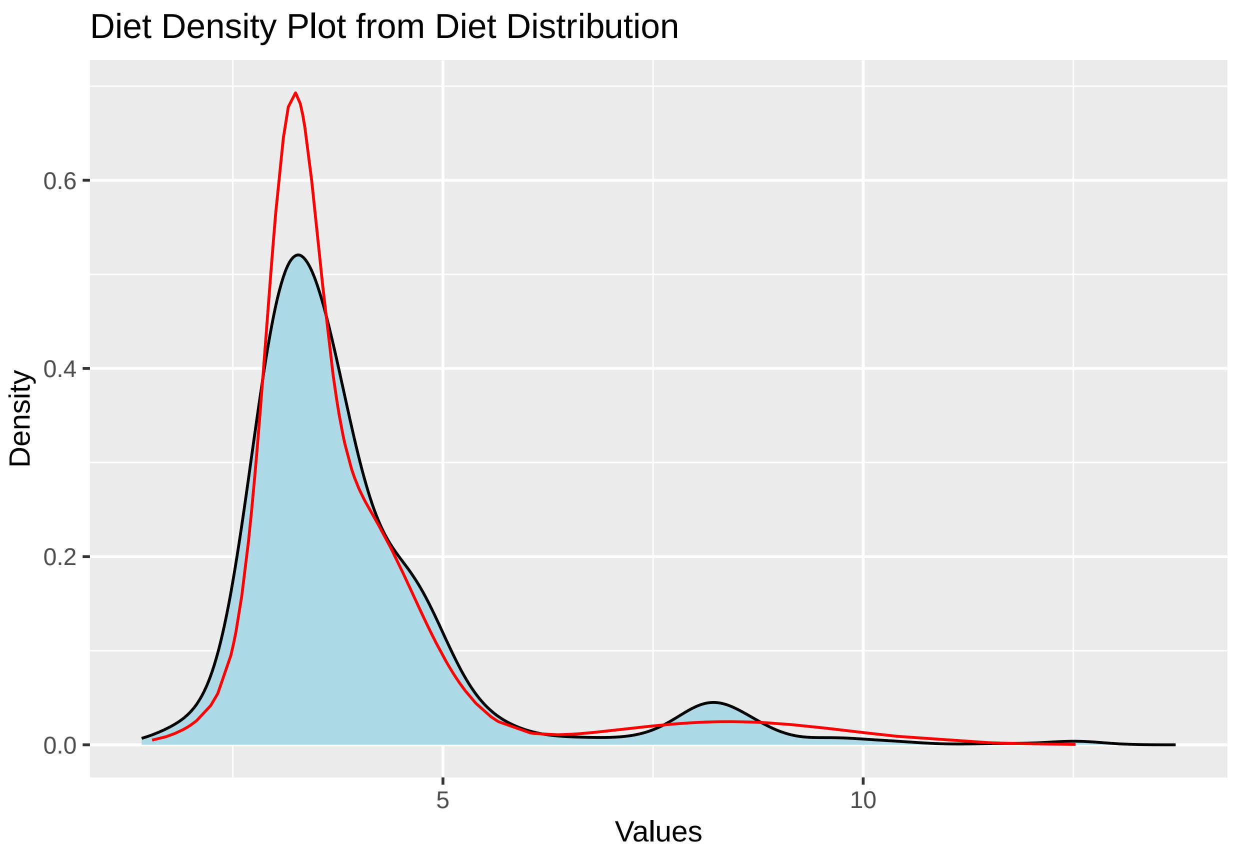
This works quite well, it overestimates one of the feeding modes, however, the locations of the PPMR of the feeding modes works well.

Just for interest, and to see if it is better, I will fit the biomass distribution, then see if it works for the number distribution.

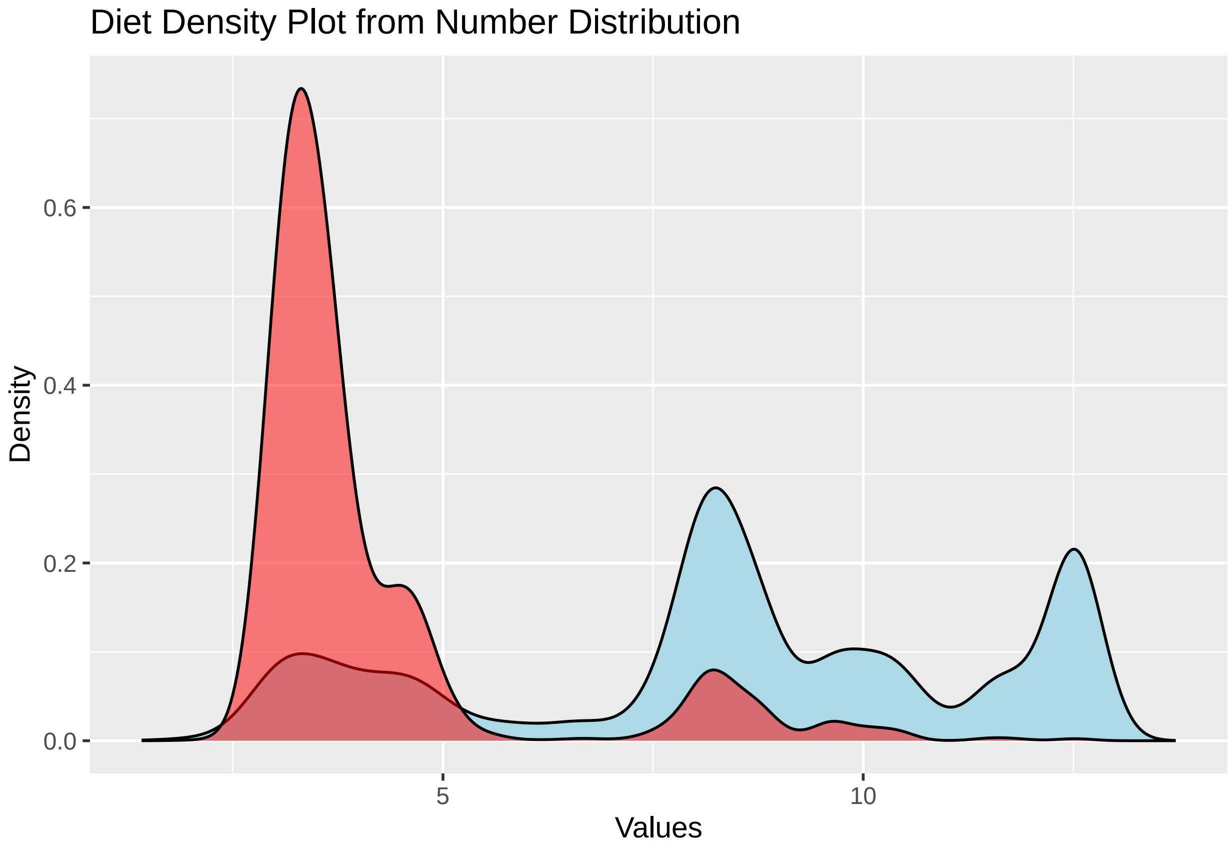
```
spratnu <- repeat_elements(log(sprat$ppmr), sprat$nprey_perpred*sprat$prey_ind_weight_g^dig)
gmm <- densityMclust(spratnu, G=3)
```



```
dplot <- data.frame(x=gmm[["data"]], density=gmm[["density"]])

(biofit <- ggplot() +
    geom_density(data=sprat, aes(log(ppmr), weight=weight_biomass), fill="lightblue")+
  geom_line(data=dplot, aes(x = x, y = density), color="red") +
  labs(x = "Values", y = "Density") +
  ggtitle("Diet Density Plot from Diet Distribution"))
```

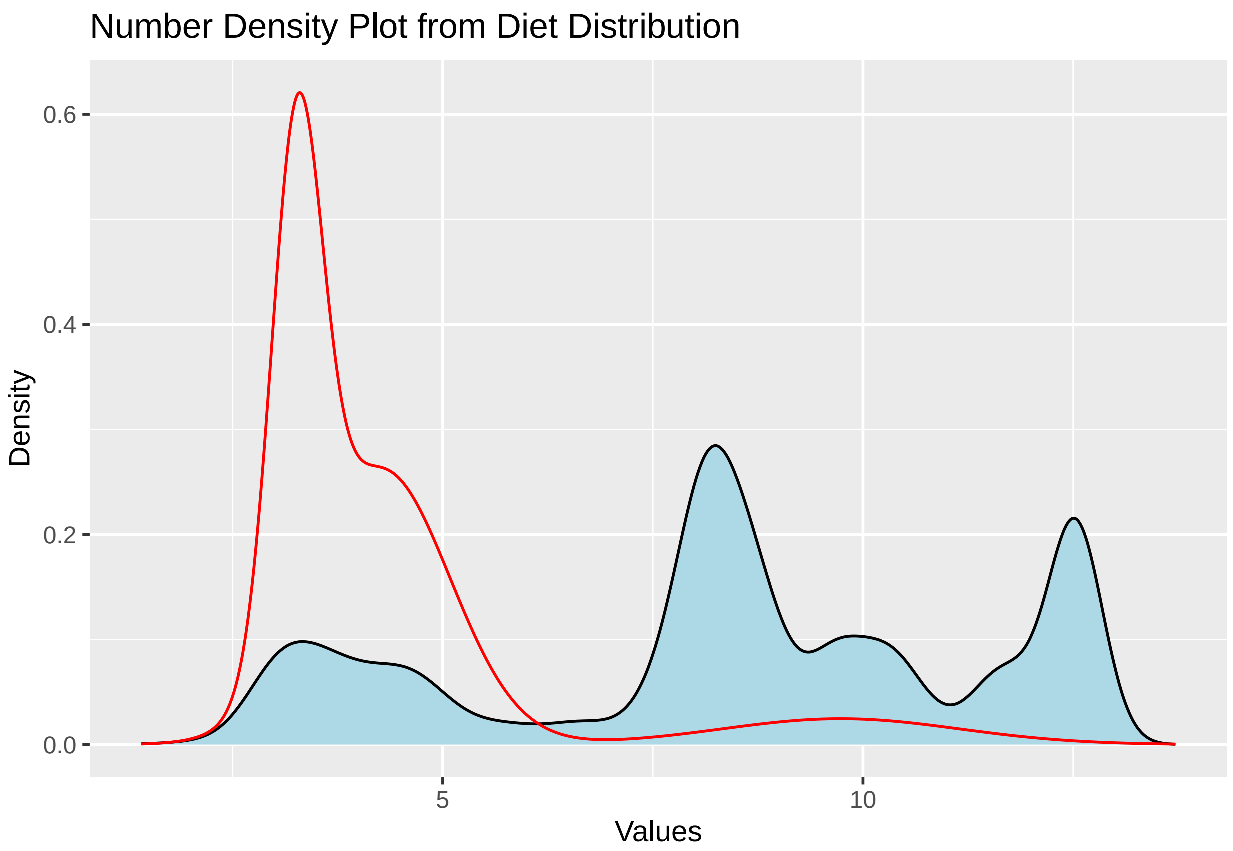


```
newspratdens <- gmm[["density"]]*exp(dig*gmm[["data"]])

shifted_pdf_normalized <- newspratdens / sum(newspratdens)

dplot <- data.frame(x=gmm[["data"]], density=shifted_pdf_normalized)

(numbfitbio <- ggplot() +
    geom_density(data=sprat, aes(log(ppmr), weight=weight_numbers), fill="lightblue")+
    geom_density(data=dplot, aes(x, weight=shifted_pdf_normalized), fill="red", alpha=0.5)+
  #geom_line(data=dplot, aes(x = x, y = density), color="red") +
  labs(x = "Values", y = "Density") +
  ggtitle("Diet Density Plot from Number Distribution"))
```



```
#I dont think I have done it right here, so I will do it in another way

shifted_fit <- gmm
shifted_fit[["parameters"]][["mean"]] <- shifted_fit[["parameters"]][["mean"]]+
  (2/3)*shifted_fit[["parameters"]][["variance"]][["sigmasq"]]

#generating the density values
shifted_pdf <- sapply(x_vals, function(x) {
  sum(shifted_fit$parameters$pro * dnorm(x, mean = shifted_fit$parameters$mean, sd = sqrt(shifted_fit$pa
})

plot_data <- data.frame(x = x_vals, shifted_pdf = shifted_pdf)

(biofitnum <- ggplot() +
    geom_density(data=sprat, aes(log(sprat$ppmr), weight=weight_numbers), fill="lightblue")+
  geom_line(data=plot_data, aes(x = x_vals, y = shifted_pdf), color="red") +
  labs(x = "Values", y = "Density") +
  ggtitle("Number Density Plot from Diet Distribution"))
```

```
## Warning: Use of ‘sprat$ppmr‘ is discouraged.
## i Use ‘ppmr‘ instead.
```



I don't think this fits very well. The other is much better. I also do not think that the exponential distribution or one gaussian would fit this PPMR distribution, so this is probably our best bet.

(Final note - I am not sure if I have done something wrong when shifting the diet density function to fit the number density, as it seems wildly wrong. I followed the same process as before, but the shifted number density does not fit in the slightest)

```
print(ngmm$parameters)
```

```
## $pro
## [1] 0.2307388 0.4959397 0.2733215
##
## $mean
##         1         2         3
##  3.943746  8.584915 12.039642
##
## $variance
## $variance$modelName
## [1] "E"
##
## $variance$d
## [1] 1
##
## $variance$G
## [1] 3
##
```

```
## $variance$sigmasq
## [1] 0.7068256
```