

# Spurdog

2024-07-21

```
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(bbmle)
```

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

```
## Loading required package: stats4
```

```
##
```

```
## Attaching package: 'bbmle'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

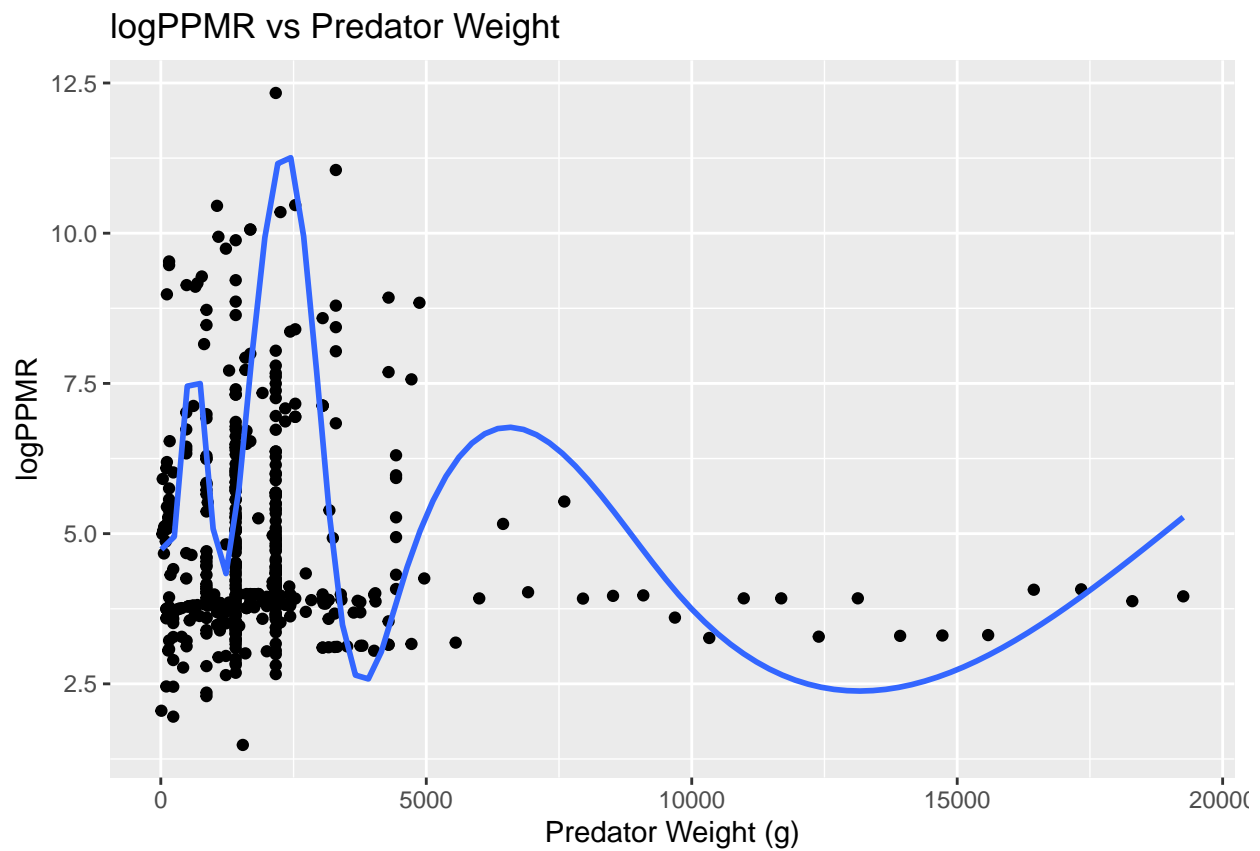
```
##      slice
```

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

```
sprat <- stom_df %>% filter(pred_taxa=="Squalus acanthias")
```

```
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")'
```



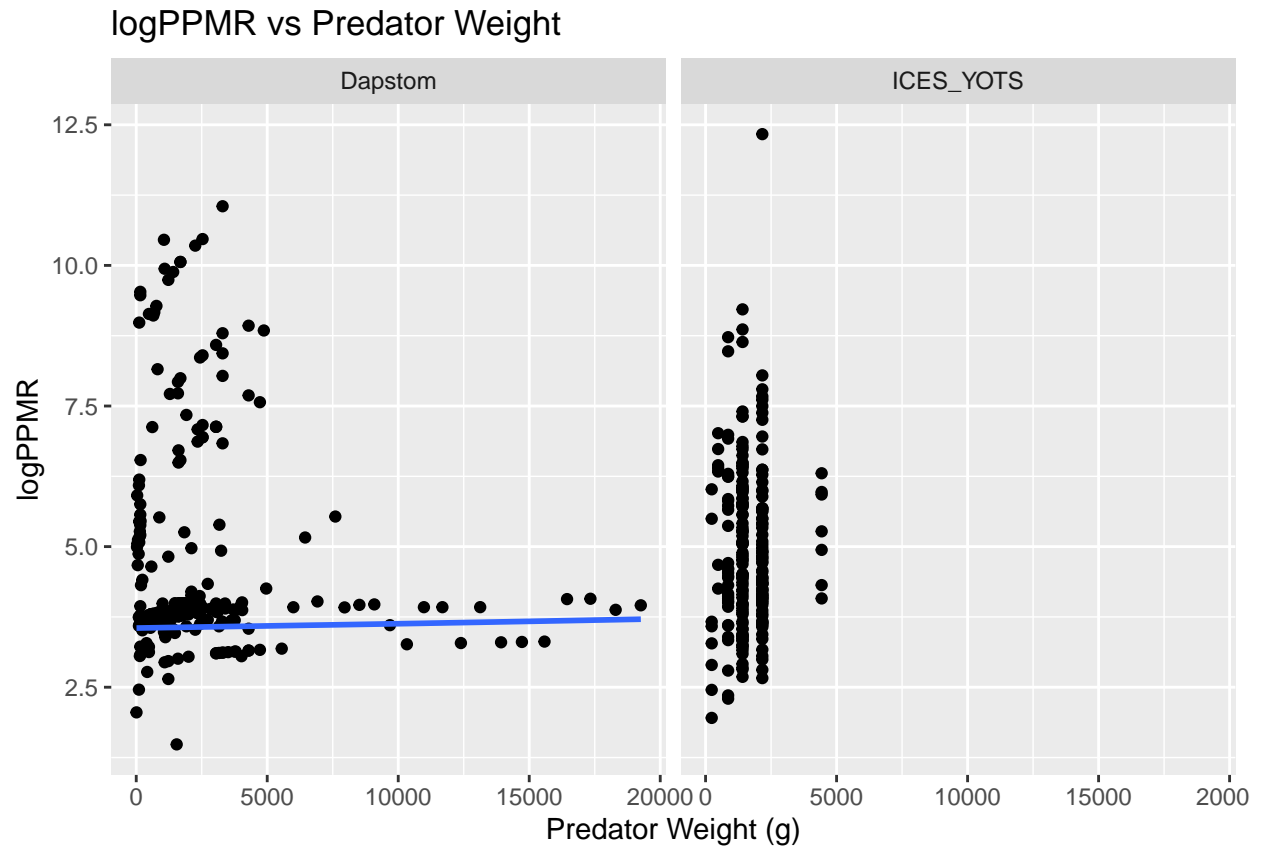
```
dig <- 1

ggplot(sprat, aes(x=pred_weight_g, y=log(ppmr)))+
  geom_point()+
  facet_wrap(~data)+
  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.
```

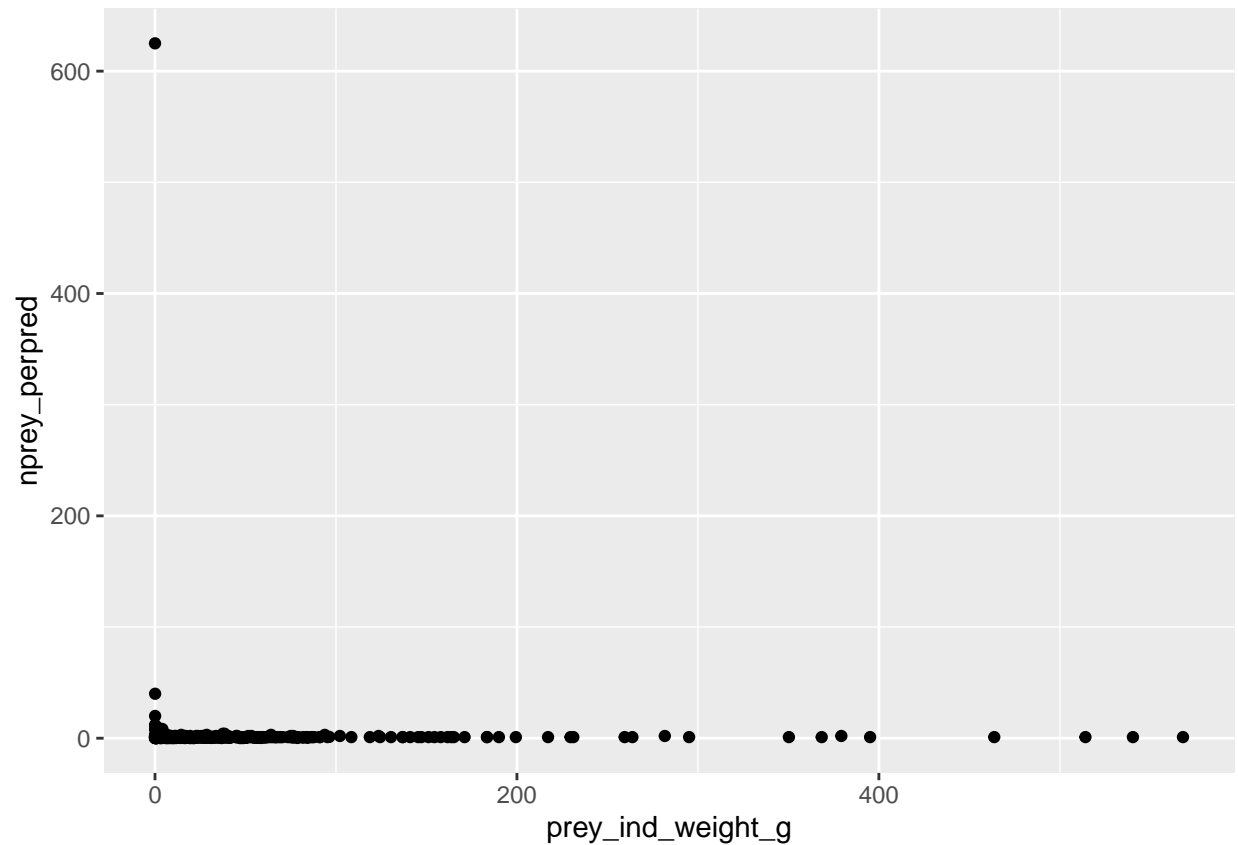
```
## 'geom_smooth()' using formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Failed to fit group -1.
## Caused by error in 'smooth.construct.cr.smooth.spec()':
## ! x has insufficient unique values to support 10 knots: reduce k.
```



In the dapstom data, it is very straight.

```
ggplot(data=sprat)+  
  geom_point(aes(x=prey_ind_weight_g, y=npred_perpred))
```



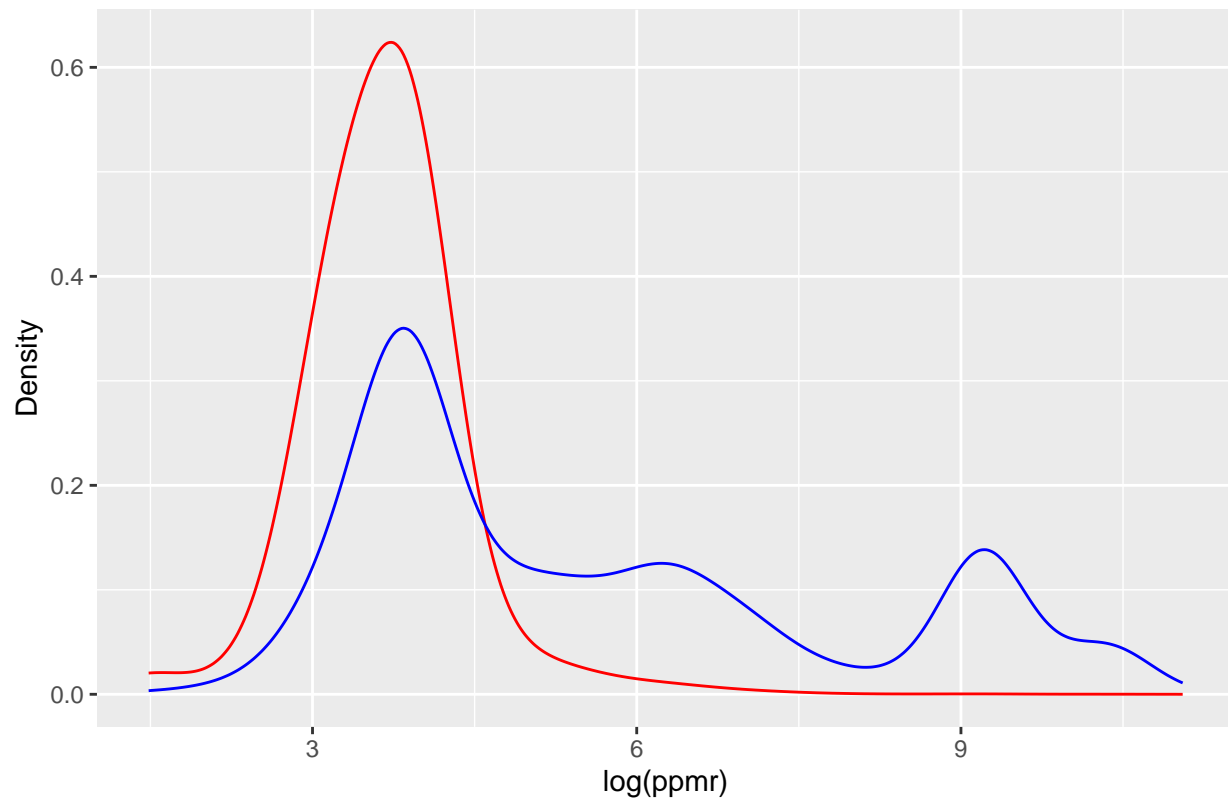
```
sprat <- sprat%>%filter(npred_perpred<200)
```

```
sprat$weight_numbers <- sprat$npred_perpred
```

```
sprat$weight_biomass <- sprat$npred_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")
```

Density Plot of log(ppmr) Weighted by Biomass and Numbers



I am not sure how to model this one, maybe exponential? Will try a double gaussian as well.

```
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)  
  }  
}
```

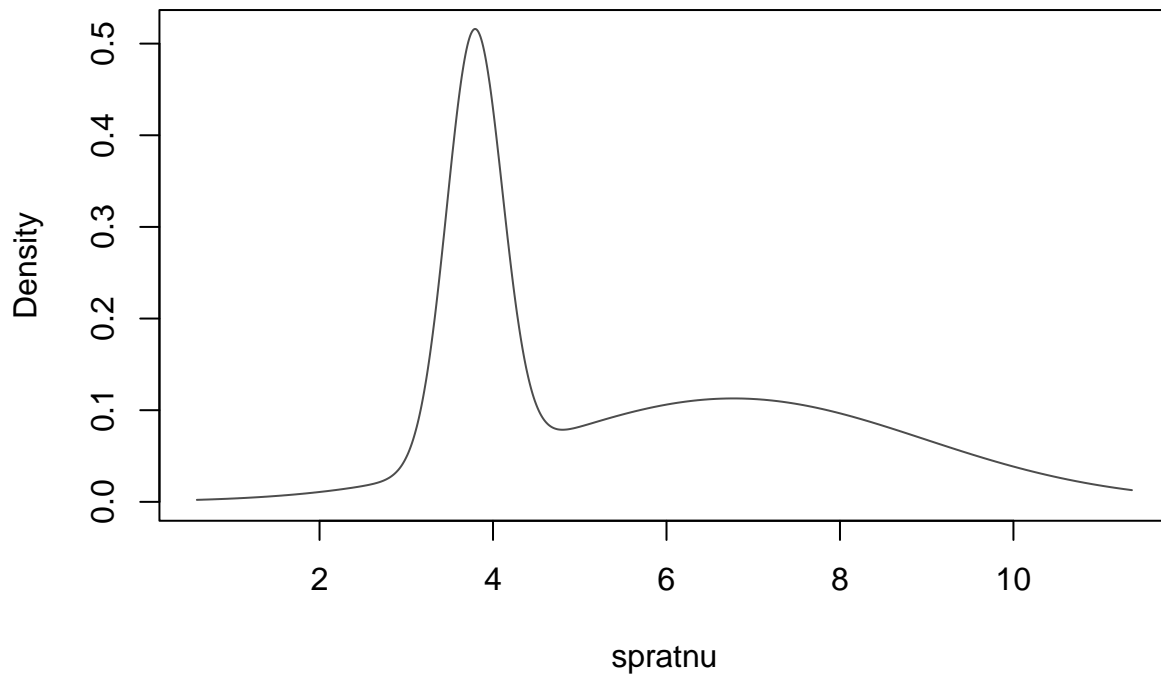
```

    return(final_vector)
  }

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

  gmm <- densityMclust(spratnu, G=2)

```



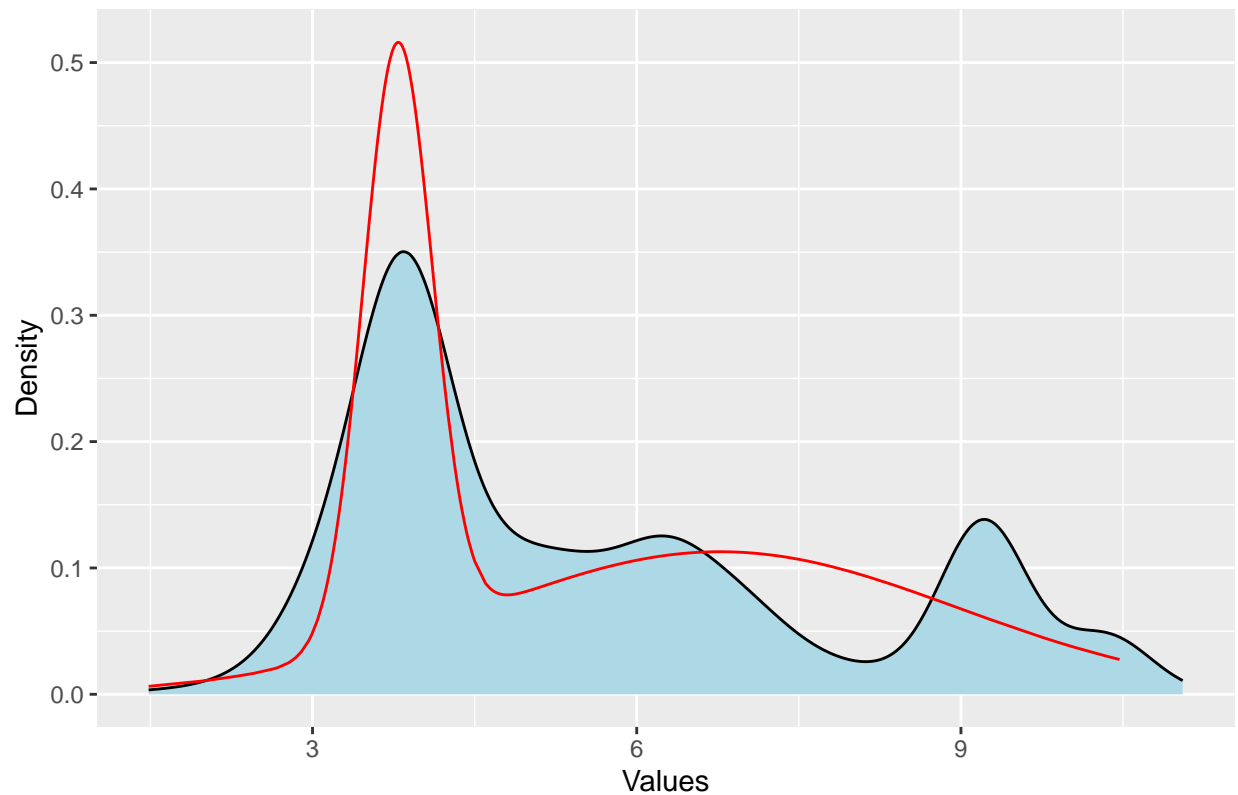
```

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"))

```

## Number Density Plot from Number Distribution



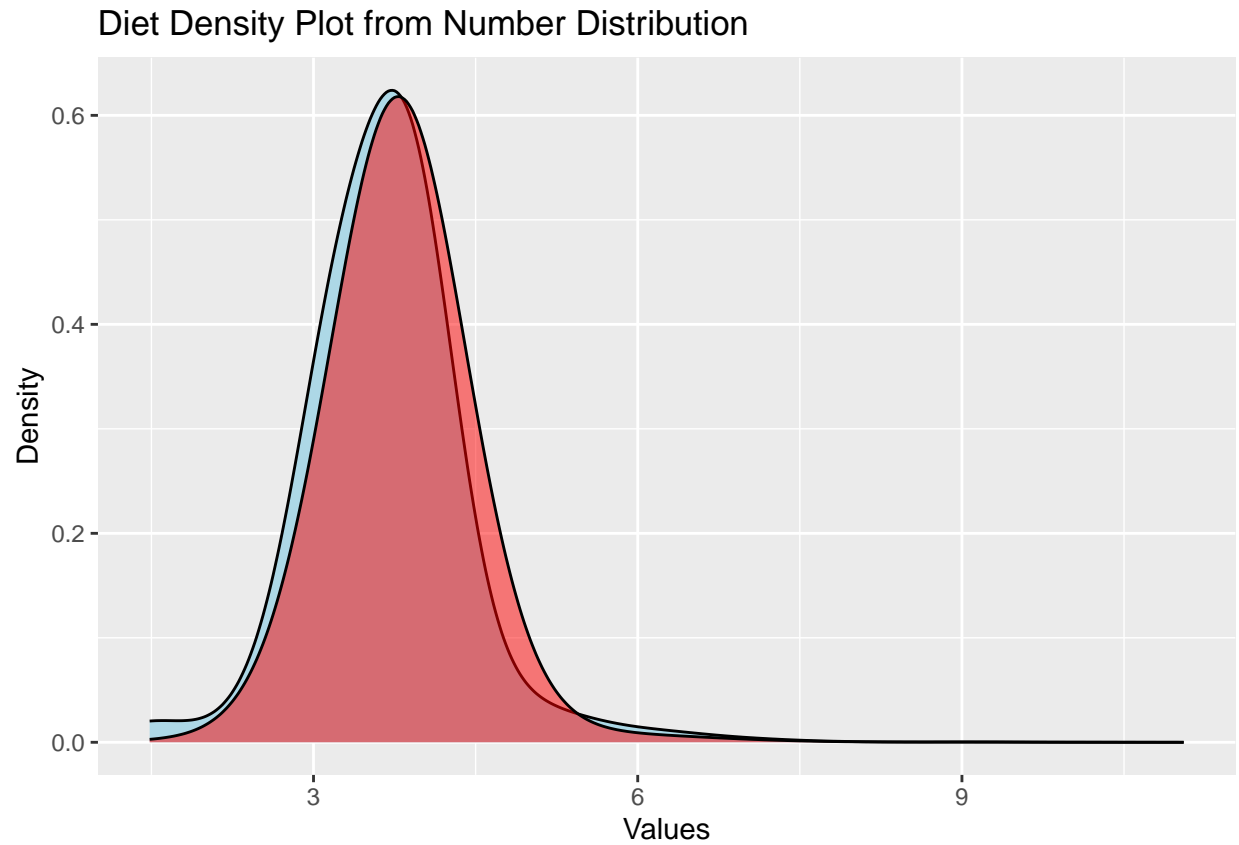
It is pretty good.

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

ngmm <- gmm

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

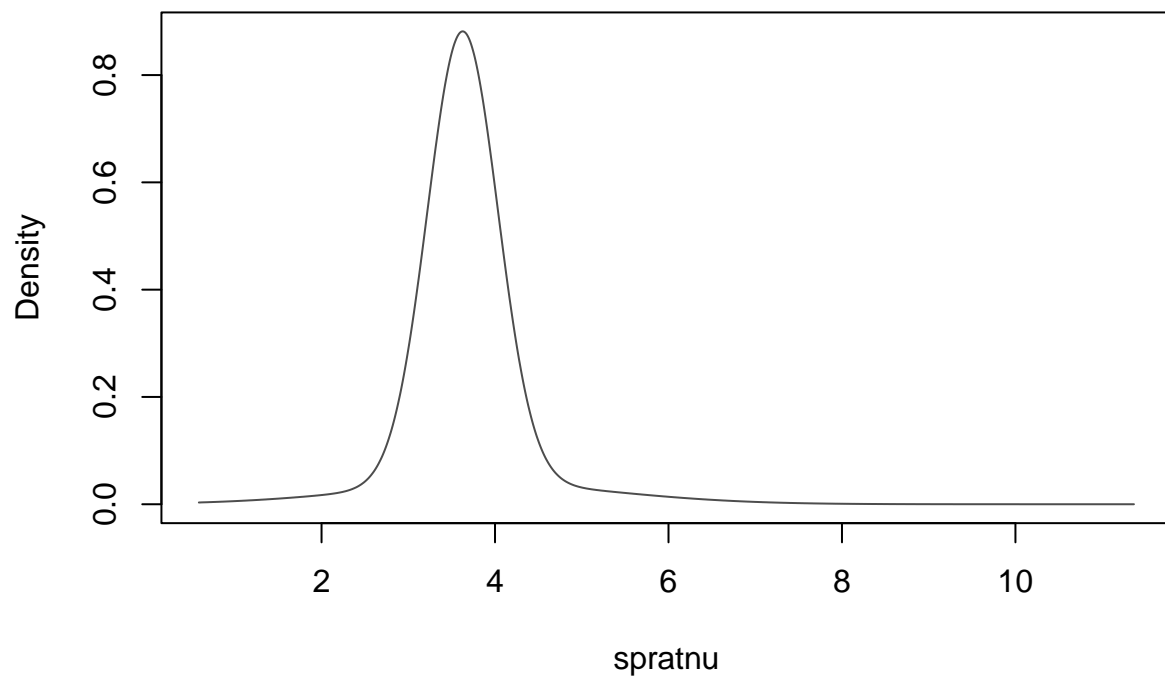
(numbfittbio <- ggplot() +
  geom_density(data=sprat, aes(log(ppmr), weight=weight_biomass), fill="lightblue")+
  geom_density(data=dplot, aes(x, weight=shifted_pdf_normalized), fill="red", alpha=0.5)+
  #geom_line(data=dplot, aes(x = x, y = density), color="red") +
  labs(x = "Values", y = "Density")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



This works quite well. the fit to the number density isnt great, but?

```
spratnu <- repeat_elements(log(sprat$ppmr), sprat$npred_perpred*sprat$prey_ind_weight_g^dig)
gmm <- densityMclust(spratnu, G=2)
```

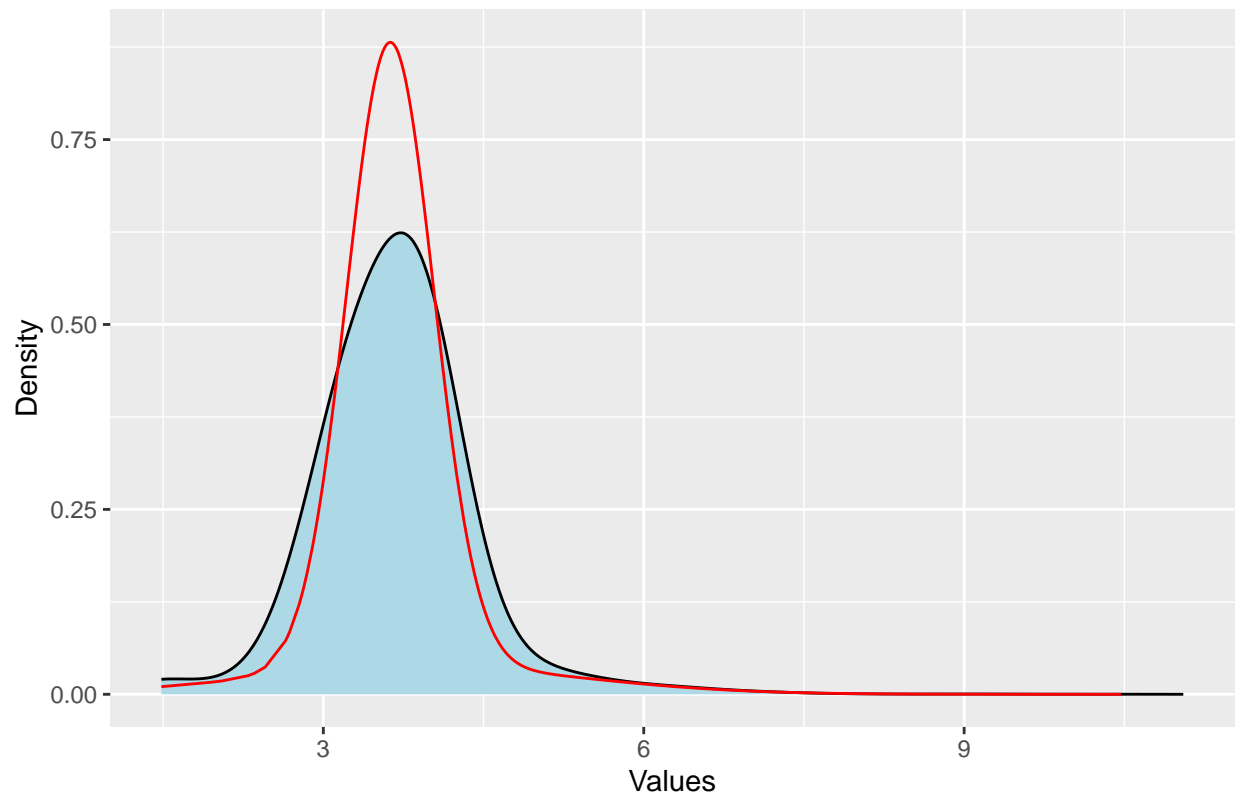




```
dplot <- data.frame(x=gmm[["data"]], density=gmm[["density"]])

(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"))
```

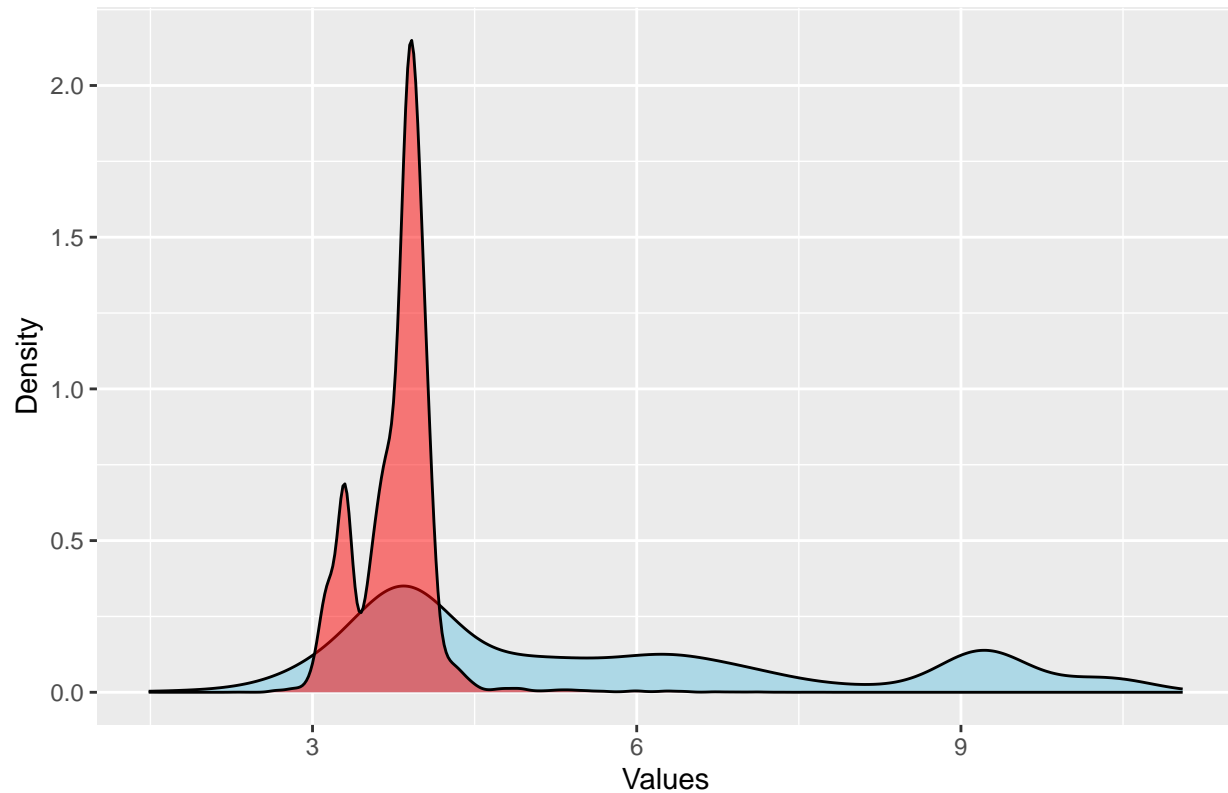
Diet Density Plot from Diet Distribution



```
newspratdens <- gmm[["density"]]*exp(dig*gmm[["data"]])
shifted_pdf_normalized <- newspratdens / sum(newspratdens)
dplot <- data.frame(x=gmm[["data"]], density=shifted_pdf_normalized)

(numbfittbio <- 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"))
```

Diet Density Plot from Number Distribution



```
sprat$l <- log(sprat$ppmr)
sprat <- sprat[!is.na(sprat$l),]
x_vals <- seq(min(sprat$l), max(sprat$l), length.out = 1000)

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

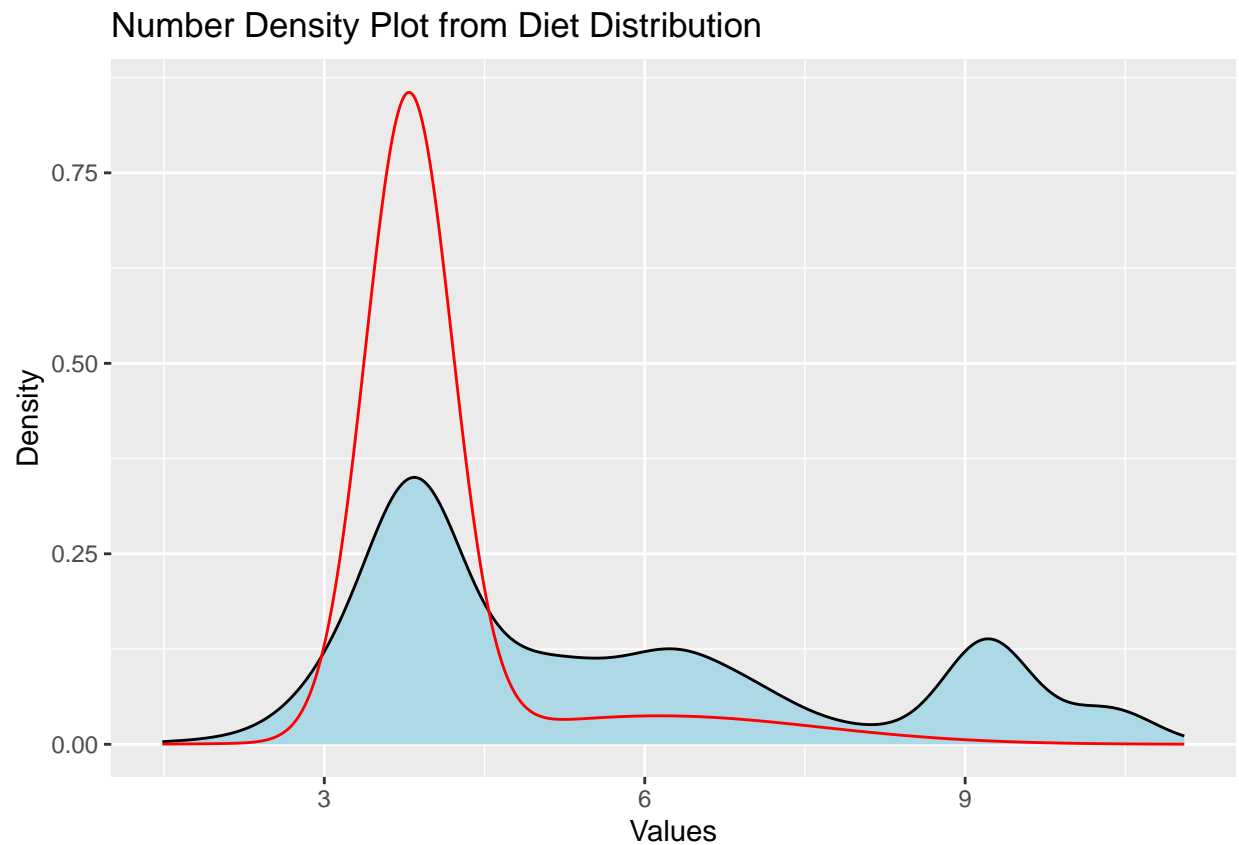
shifted_fit <- gmm
shifted_fit[["parameters"]][["mean"]] <- shifted_fit[["parameters"]][["mean"]] +
  (1)*shifted_fit[["parameters"]][["variance"]][["sigmasq"]]

#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$parameters$variance$pro * shifted_fit$parameters$variance$sigmasq)))
})

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.
```



This doesnt work though.

Here is parameters

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

```
## $pro
## [1] 0.377416 0.622584
##
## $mean
##      1      2
## 3.787757 6.772024
##
## $variance
## $variance$modelName
## [1] "V"
##
## $variance$d
## [1] 1
##
## $variance$G
## [1] 2
##
## $variance$sigma_sq
## [1] 0.1022199 4.8450106
##
```

```
## $variance$scale  
## [1] 0.1022199 4.8450106
```