Linking Diet and Reproductive Success in Northwest Atlantic Grey Seals

Max Henry-Adams (B00952295) and Peter Braithwaite (B00504978)

Data Analysis STAT 4620/5620 Winter 24-25"

Link to GitHub Repository: https://github.com/MaxHA17/STAT5620.Project

Abstract:

Understanding how top predators use different foraging strategies to navigate and adapt to patchy and unpredictable prey availability has important implications for individual fitness and population dynamics. Here we explore the relationship between diet and reproductive success in a capital breeding pinniped, the northwest Atlantic grey seal (*Halichoerus grypus*) through two research questions: 1. Does variation in diet influence maternal mass change over the foraging period leading up to parturition in Northwest Atlantic grey seals; and 2. Does female diet predict pup weaning mass in Northwest Atlantic grey seals? Using linear models (LMs), generalized linear models (GLMs), and generalized linear mixed effects models (GLMMs), we found that females who primarily fed on pollock experienced lower mass change and pup weaning masses than those who primarily fed on cod, sand lance, or redfish. While models for both questions were well fit, model predictive errors were relatively high, possibly indicating that missing covariates outside of diet are playing a significant role in explaining maternal mass change and pup weaning mass.

Keywords: Capital breeding strategy, Generalized Linear Models, Pinnipeds, Quantitative Fatty Acid Signature Analysis, Reproductive Ecology

Introduction:

For marine predators, prey distribution may vary unpredictably in time and space, undergoing seasonal, inter-annual, and longer-term changes in environmental conditions. Understanding how top predators use different foraging strategies to navigate and adapt to patchy and unpredictable prey availability has important implications for individual fitness and population dynamics, yet studies exploring the consequences of such decisions on fitness are relatively rare (Nathan et al., 2008).

Vertebrates fuel the costs of reproduction along a continuum of an income-capital breeding strategy, where animals that use an income breeding strategy rely on concurrent energy accumulation during the breeding period as opposed to a capital breeding strategy, where energy acquired and stored prior to the breeding period is used solely to finance major reproductive costs (Stephens et al., 2009; Stephens et al., 2014). A capital breeding strategy is thought to be advantageous for large animals capable of carrying large energy stores and offers energetic benefits in areas of patchy or unreliable food availability and allows parents to direct more resources (e.g. time and energy) towards offspring, as opposed to foraging, during the breeding and lactation period (Stephens et al., 2009; Stephens et al.,2014). Given the importance of pre-breeding and lactation period foraging success and energy acquisition in capital breeders, these species offer interesting model systems to study the relationships between diet and reproductive success.

Phocid seals are wide-ranging, large-bodied, and long-lived marine predators, with individuals of many species foraging across a wide range of habitats over large temporal scales (Bowen and Jonsen, 2022). Grey seals (*Halichoerus grypus*) are a long-lived, sexually dimorphic, phocid species and are capital breeders. In the Sable Island population, pregnant females haul out in late December or early January,give birth to a single pup and nurse that pup for 16-18 days, relying solely on energy accumulated during the foraging period prior to parturition to support the costs of lactation (Iverson et al., 1993; Stephens et al., 2009). At or near the abrupt weaning of that pup, females are mated and return to sea to replenish spent body reserves and, after a several month period of delayed implantation, to support gestation and most importantly preparation for the subsequent December or early January lactation period. As in all mammals, the costs of lactation,

1

and its contribution to reproductive success, far exceed the costs of gestation. Like most other phocids, grey seals are generalist predators (Bowen and Jonsen, 2022), but females tend to feed on a narrower range of energy rich species (predominantly sand lance, redfish, and other pelagic fishes) following the lactation period and expand their prey diversity in the several months leading up to the December-January breeding and lactation period (Becket al., 2007). Female grey seals reach sexual maturity at age 4-5 years and can continue to reproduce as late as into their early 40s (Bowen et al., 2006). During the brief 16–18-day lactation period, females utilize approximately 25-38% of parturition body mass and daily maternal mass loss is a significant predictor of pup growth rates during lactation, with heavier mothers at parturition weaning heavier pups (Iverson et al., 1993; Mellish et al., 1999). Pup survival is positively related to pup weaning mass up to the mean population weaning mass of 51.5 kg, as predicted by a bigger-is-better hypothesis for size selective mortality (Bowen et al., 2015). The phenomena of heavier females weaning larger offspring is evident even from primiparity where it is believed that heavier females have larger resource stores that they can mobilize during lactation to produce larger offspring at weaning (Iverson et al., 1993; Mellish et al., 1999; Bowen et al., 2006). Following the lactation period, parental care abruptly ends and the pup is left to fast and survive for several weeks on the energy acquired during lactation before undertaking its first foraging trip (Noren et al., 2008). Thus, answering questions relating to the diet of females prior to parturition will improve our understanding of the importance of habitat use and energy acquisition as they relate to life history characteristics, such as reproductive success, in these capital breeders.

Northwest Atlantic grey seals breeding on Sable Island, Nova Scotia, tend to be central place foragers and have high fidelity to Sable Island for breeding and as a haul-out site, making them excellent candidates for long-term studies (Bowen et al., 2015). Using diet estimations methods, such as quantitative fatty acid (FA) signature analysis (QFASA; Iverson et al., 2004, Beck et al., 2007), it becomes possible to explore the relationship between diet variation and reproductive success at the individual level. QFASA is used to estimate predator diet by comparing the FA signatures of predator adipose tissue to those of candidate prey species (Iverson, 2009; Karnovsky et al., 2012). Predator FA signatures alone have historically been used to provide a qualitative assessment of spatial and temporal patterns in diet diversity. However, since many FAs that are biosynthesized within unique prey species accumulate in predator adipose tissue over time, evaluating the amount of prey-specific FAs relative to the predator's overall FA stores can provide a quantitative estimate of diet, known as QFASA (Iverson etal., 2004). For accurate diet estimation, species-specific calibration coefficients must be experimentally derived to account for the altered incorporation of prey FAs into the predator adipose tissue due to lipid metabolic processes within the predator (Iverson et al., 2004; Karnovsky et al., 2012). When calibration coefficients are derived and the FA signature of many or all potential prey species in an ecosystem is known, as in the case of the Northwest Atlantic grey seal range (Iverson et al., 2004; Beck et al., 2007), QFASA can provide a statistical approach to accurately depicting diet and prey diversity over ecologically relevant time scales (e.g., weeks to several months). This study will test the hypothesis that differences in diet will affect an individual's ability to store energy prior to parturition, and thus influence their reproductive success. Specifically, we will explore two research questions:

1. Does variation in diet influence maternal mass change over the foraging period leading up toparturition in Northwest Atlantic grey seals?

2. Does female diet predict pup weaning mass in Northwest Atlantic grey seals?

Data Description

Response variables (organized by research question):

1. Maternal mass change during the pre-breeding season foraging period: continuous variable ranging from -45 – 88 kg. Calculated by subtracting the maternal recovery mass from her deployment mass. Insert plot of mass change distribution here?

2. Pup weaning mass: continuous variable ranging from 34.5 – 63 kg. Denotes the mass of the pup after the mother has terminated the lactation period and departed the pup and/or Sable Island. Insert plot of pup weaning mass distribution here?

1. Year: discrete variable with values ranging from 1996-2015. Denotes the year in which each observation was collected.

2. Maternal ID: discrete variable with a unique identifier for each individual female.

3. Pup sex: categorical; male or female.

4. Maternal age: discrete variable with values ranging from 7-40 years old.

5. Maternal deployment mass: continuous variable with values ranging from $116 - 250$ kg. Describes the female mass when she was located pregnant on Sable Island 3-6 months prior to parturition and prior to undergoing her pre-parturition foraging period.

6. Maternal mass at parturition: continuous variable ranging from $\sim 137 - 250$ kg. This is the mass of the female after returning to Sable Island and giving birth to her pup.

7. Maternal dominant prey species: discrete variable with 6 unique values. Represents the species making up the greatest proportion of the female diet.

8. Maternal dietary energy density: continuous variable ranging from $4.88 - 6.51$ kJ/g. Energy density (kJ/g) of the female diet is calculated from the average energy content per gram of prey body tissue.

9. Maternal diet diversity: continuous variable $0.18 - 0.58$, no specific units. Diversity is calculated using a standardized Shannon-Weaver diversity index to quantify individual diet diversity.

Methods

The initial step was data cleaning to ensure all readily apparent data errors were rectified prior to data analysis and model fitting. Most of the data errors discovered resulted from the absent of critical data (i.e. missing covariates) and as a result these observations were removed from the data-set. As a result, the first question data-set resulted in 56 observations (n=56) and the second question resulted in 76 observations (n=76). It should also be noted that additional observations were filtered from the data-set throughout the analysis process due to the identification of outliers with high leverage. A description of outlier identification and removal can be found in the analysis section below. Thus, the final data-set used for question one had 72 observation and the final data-set for question two had 50 observations.

Once cleaned and uploaded to RStudio, initial data plots were made using the 'flexplot' and 'plot_explore' R function we have included in our Git Hub. Description of 'plot_explore' function created in this R package can be found within the Vignettes in the STAT5620.Project package. These visualizations facilitated the exploration of variable distributions, specifically the response variables, and the apparent relationships between covariates and each response variable. From this it was determined that the response variables for both questions, Maternal Mass Change and Pup Weaning Mass, were Gaussian distributed. This normality allowed the exploration of the data through General Linear Models as well as more complex Generalized Linear Models and Mixed Models. Although our original analytical approach proposed the use of Generalized Linear Mixed Models, it felt prudent to explore the efficiency, interpretability and computational simplify of Linear Models if the assumption were met and the data suggested linearity.

As a result, both questions were analysed first as Linear Models (LM), then as Generalized Linear Models (GLM) with Gaussian distributions, and finally as Generalized Linear Mixed Models (GLMM) with random effects and fixed effects of covariates. Initially, Maternal ID was to be included as a random effect to account for repeated measures, individual heterogeneity, and allow for population level inferences to be made using these data. However, a lack of repeated measures (<10 per question) limited our ability to include Maternal ID as a random effect and Maternal ID was assumed to be fixed. For question one in particular, including a random slope for the effect of dominant prey species was explored due to the dominant prey species included in this data-set being a random draw from all possible prey species. However, the goal of this analysis was

to explore the direct effect of each prey species on maternal mass gain, resulting in dominant prey species being included as a fixed effect in the final model. With each model type, we began with all the predictor variables prior to using the backward and forward "Step" function to reduce the number of variables through elimination based on the p-values for Linear Models and AIC for GLM and GLMM.

Once we had established the best fit model for both questions using AIC, the performance of the best models were tested using Cross Validation, question one using 7 folds and question two using 11 folds. For question 1, all modelling and cross-validation was performed using functions included in base R.

Analysis

Research Question 1: Here we are exploring whether variation in diet influence maternal mass change over the foraging period leading up to parturition in Northwest Atlantic grey seals?

The response variable of interest is maternal mass change. We will first look at the distribution of this variable:

```
#Load the data
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.2
```

```
## here() starts at C:/Users/Maxhe/OneDrive - Dalhousie University/Documents/STAT5620.Project
```

```
library(readr)
```

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

```
here::here("Processed_Data", "Data_Q1.csv")
```

```
## [1] "C:/Users/Maxhe/OneDrive - Dalhousie University/Documents/STAT5620.Project/Processed_Data/Data_Q
```

```
Data_Q1 <- read.csv(here("Processed_Data", "Data_Q1.csv"))
#Data_Q1 <- read_csv("Processed_Data/Data_Q1.csv")
View(Data_Q1)

#Ensure the data types are correctly assigned
Data_Q1$MomID <- as.factor(Data_Q1$MomID)
Data_Q1$Year <- factor(Data_Q1$Year, levels = sort(unique(Data_Q1$Year)), ordered = TRUE)
Data_Q1$Dominant.prey.species <- as.character(Data_Q1$`Dominant.prey.species`)
Data_Q1$`Diet.diversity` <- as.numeric(Data_Q1$`Diet.diversity`)
Data_Q1$`Dietary.energy.density` <- as.numeric(Data_Q1$`Dietary.energy.density`)

summary(Data_Q1)
```

```
##       MomID          Year     Dietary.energy.density Diet.diversity
##   4269   : 3    2013   :13    Min.   :4.881          Min.   :0.1788
##   24     : 2    2011   :11    1st Qu.:5.621          1st Qu.:0.3061
##   146    : 2    2012   : 9    Median :5.754          Median :0.3560
##   829    : 2    2010   : 8    Mean   :5.736          Mean   :0.3606
```

```
## 3271   : 2   2009   : 6   3rd Qu.:5.893          3rd Qu.:0.4117
## 3616   : 2   2015   : 5   Max.   :6.517          Max.    :0.5782
## (Other):62   (Other):23
## Dominant.prey.species  Mass.change
## Length:75               Min.   :-40.50
## Class :character        1st Qu.:  8.00
## Mode  :character        Median : 23.50
##                         Mean   : 26.55
##                         3rd Qu.: 45.00
##                         Max.   : 88.00
##
```

```r
#rename columns
library(dplyr)
```

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

```
##
## Attaching package: 'dplyr'
```
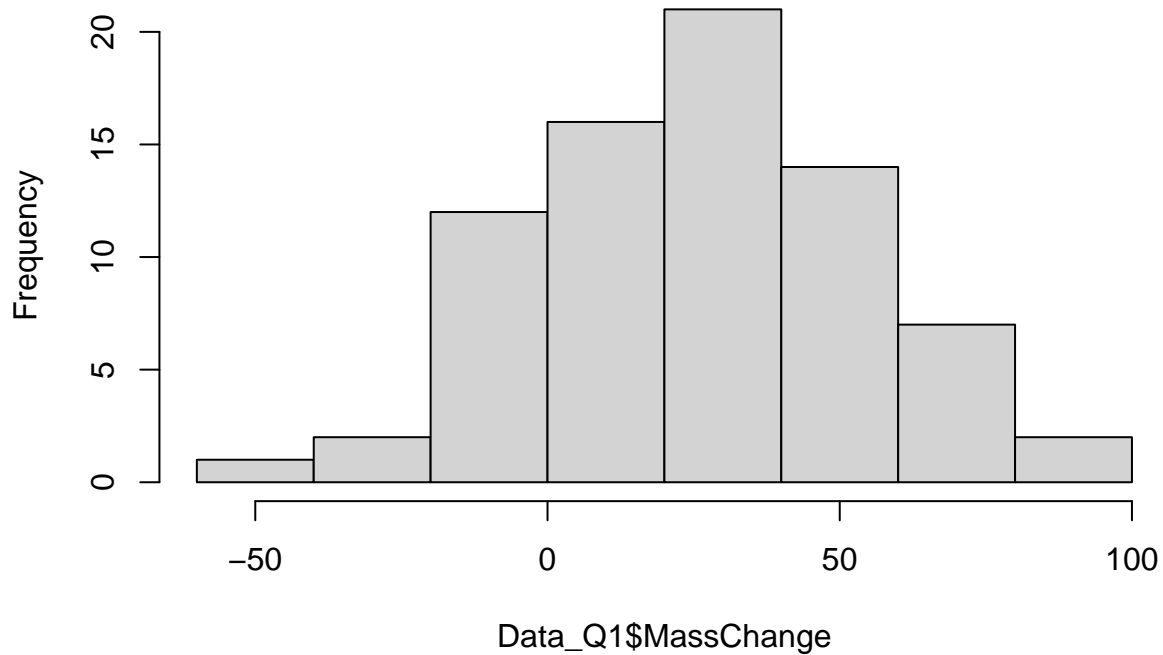
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
Data_Q1 <- Data_Q1 %>%
  rename(
    DietEngDen = `Dietary.energy.density`,
    DietDiv = `Diet.diversity`,
    DomSpp = `Dominant.prey.species`,
    MassChange = `Mass.change`)

hist(Data_Q1$MassChange)
```

Histogram of Data_Q1$MassChange

There are a total of n=76 observations and mass change appears to follow a normal distribution a Gaussian family distribution is a good starting point for an initial model.

One of the covariates, dominant prey species, is a categorical variable. We will first check the number of observations in each category:

```
table(Data_Q1$DomSpp)
```

```
##
##       AtlanticCod        Capelin NorthernSandlance          Pollock
##                 6              3                29                5
##           Redfish
##                32
```

With only one observation of white hake and 3 observations of capelin, there is likely not enough data to estimate the effect of these species on the response variable and we are unable to draw meaningful conclusions. These points also cause extreme residual values. Capelin and white hake observations will be removed from the data set:

```
Data_Q1 <- Data_Q1 %>%
  filter(DomSpp != "WhiteHake")

Data_Q1 <- Data_Q1 %>%
  filter(DomSpp != "Capelin")
```

The initial model seeks to model maternal mass change as a function of dominant prey species, diet diversity, dietary energy density, and deployment year. We will now explore the apparent relationships between each continuous and categorical covariate using the plot_explore function built and contained in the STAT5620.Project R package:

```r
library(STAT5620.Project)
plot_explore(Data_Q1, response = "MassChange",
                     continuous_vars = c("DietEngDen", "DietDiv"),
                     categorical_vars = c("Year", "DomSpp"))
```

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

## Warning: package 'gridExtra' was built under R version 4.3.2

##
## Attaching package: 'gridExtra'

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

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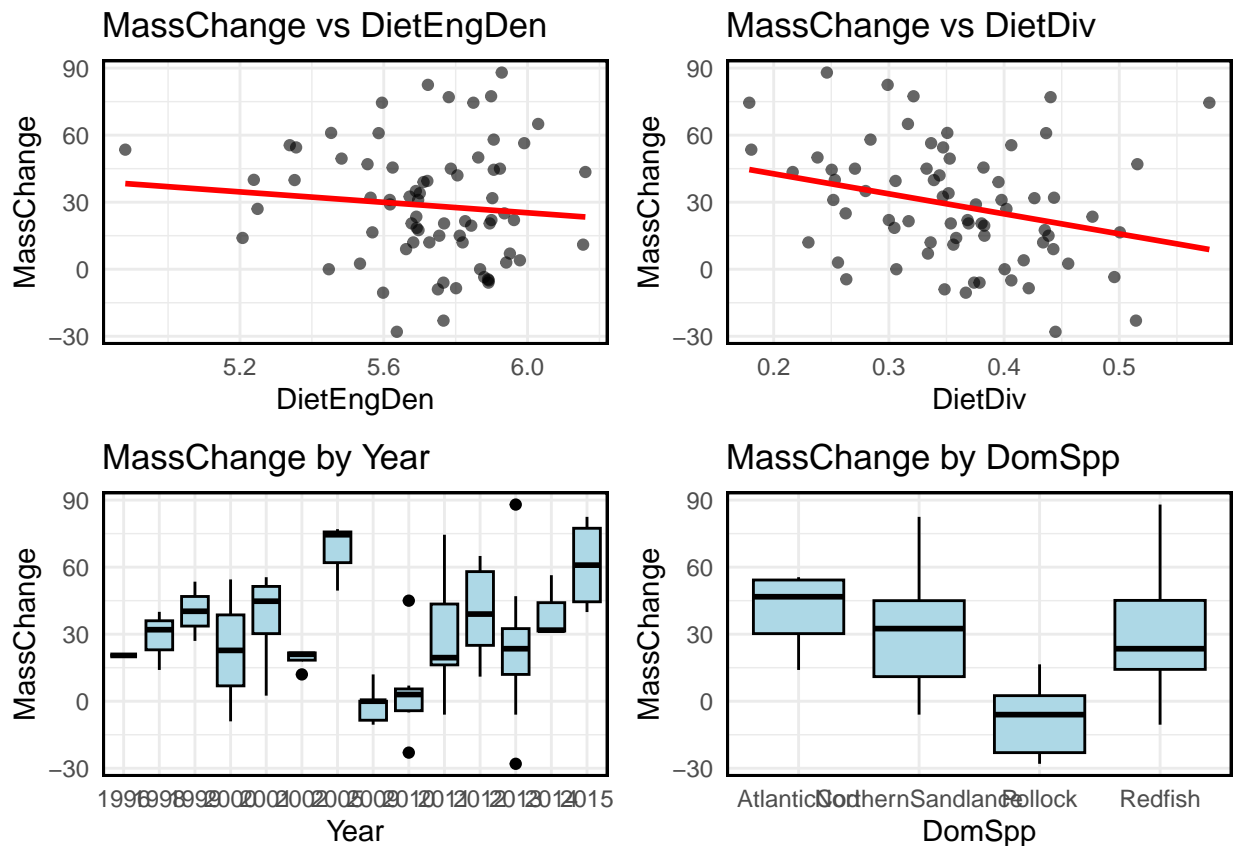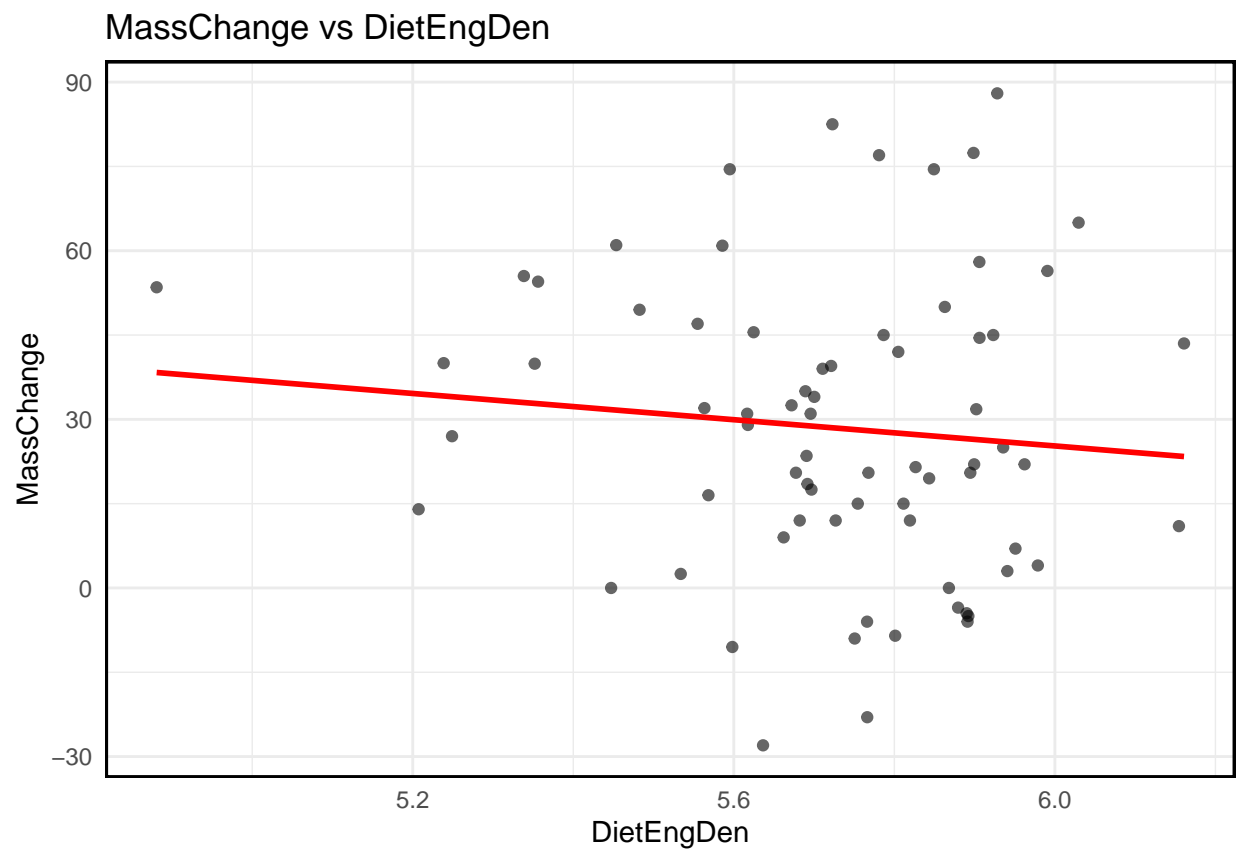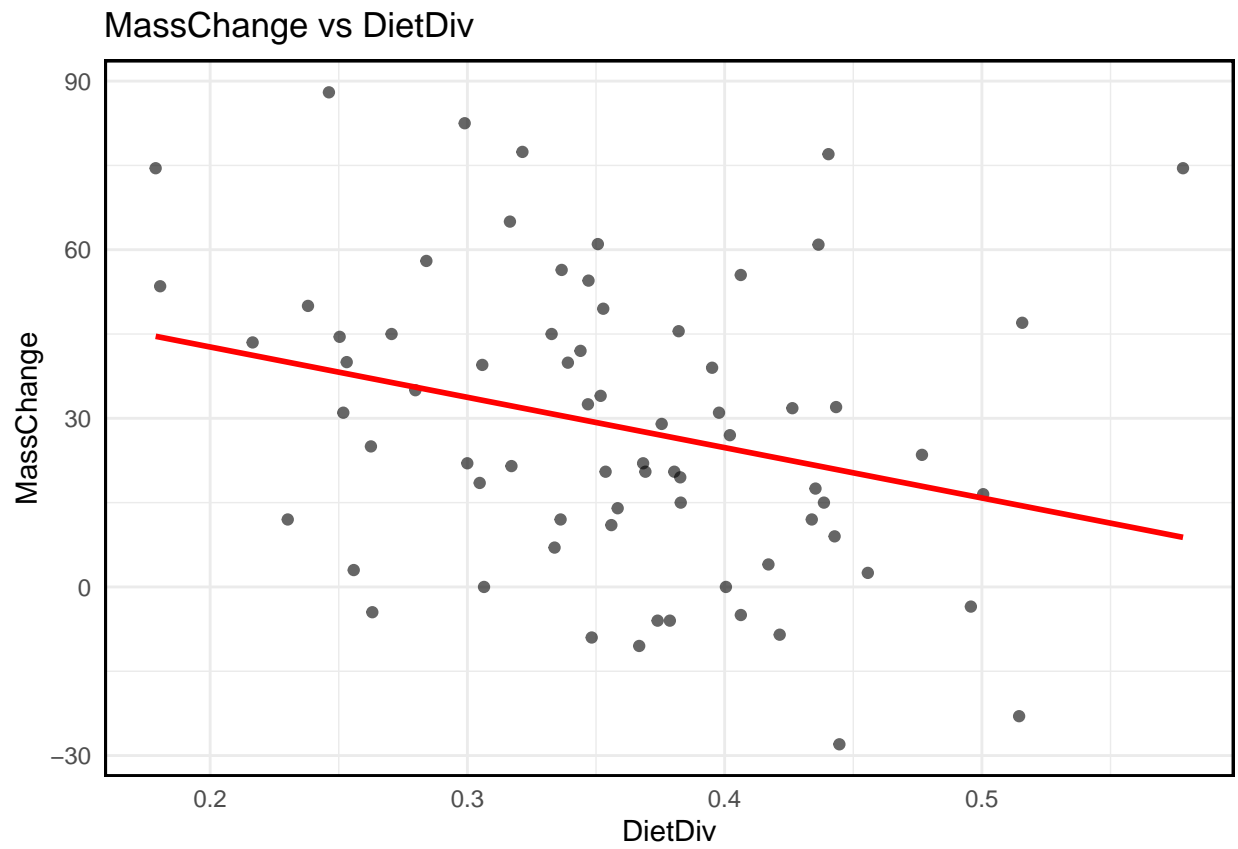
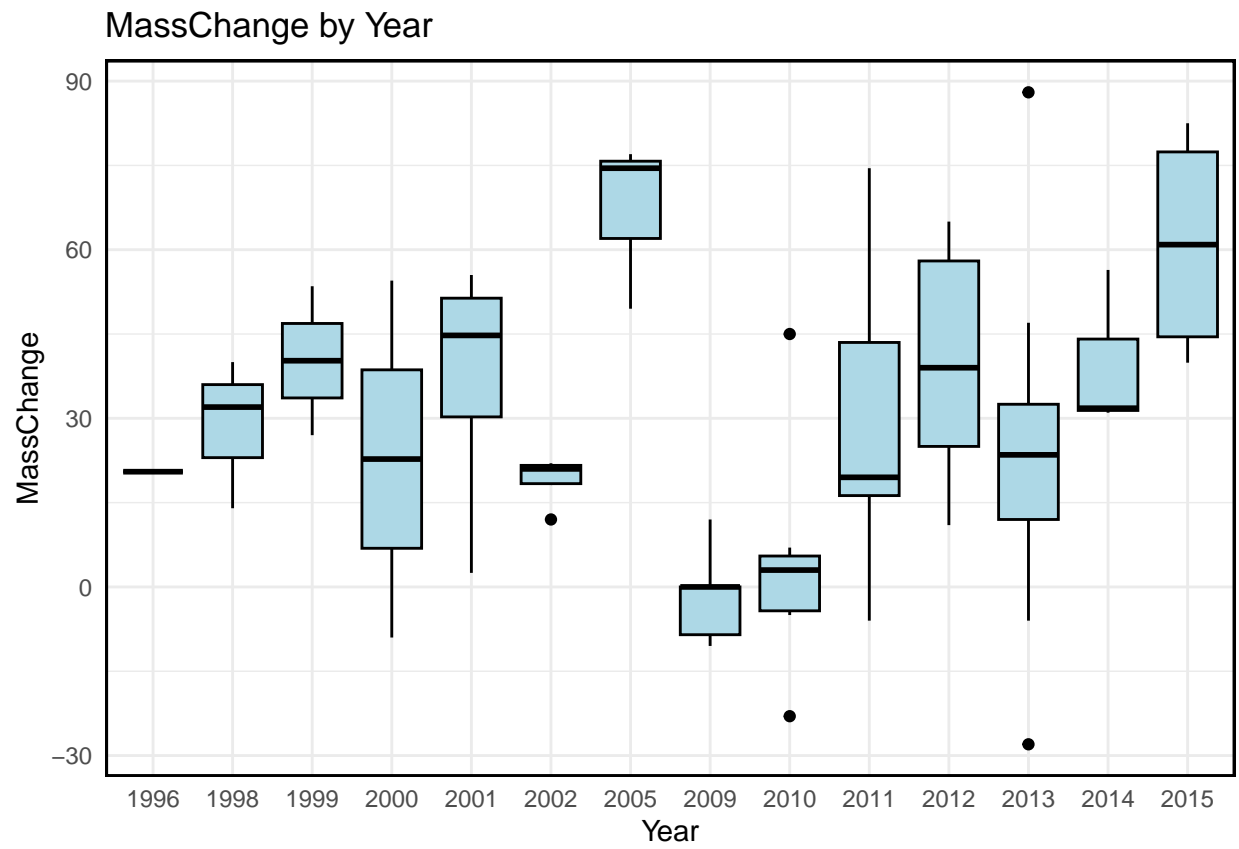## $DietEngDen

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

MassChange vs DietEngDen

##
$DietDiv

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

MassChange vs DietDiv

```
## 
## $Year
```

MassChange by Year


```
##
## $DomSpp
```

MassChange by DomSpp
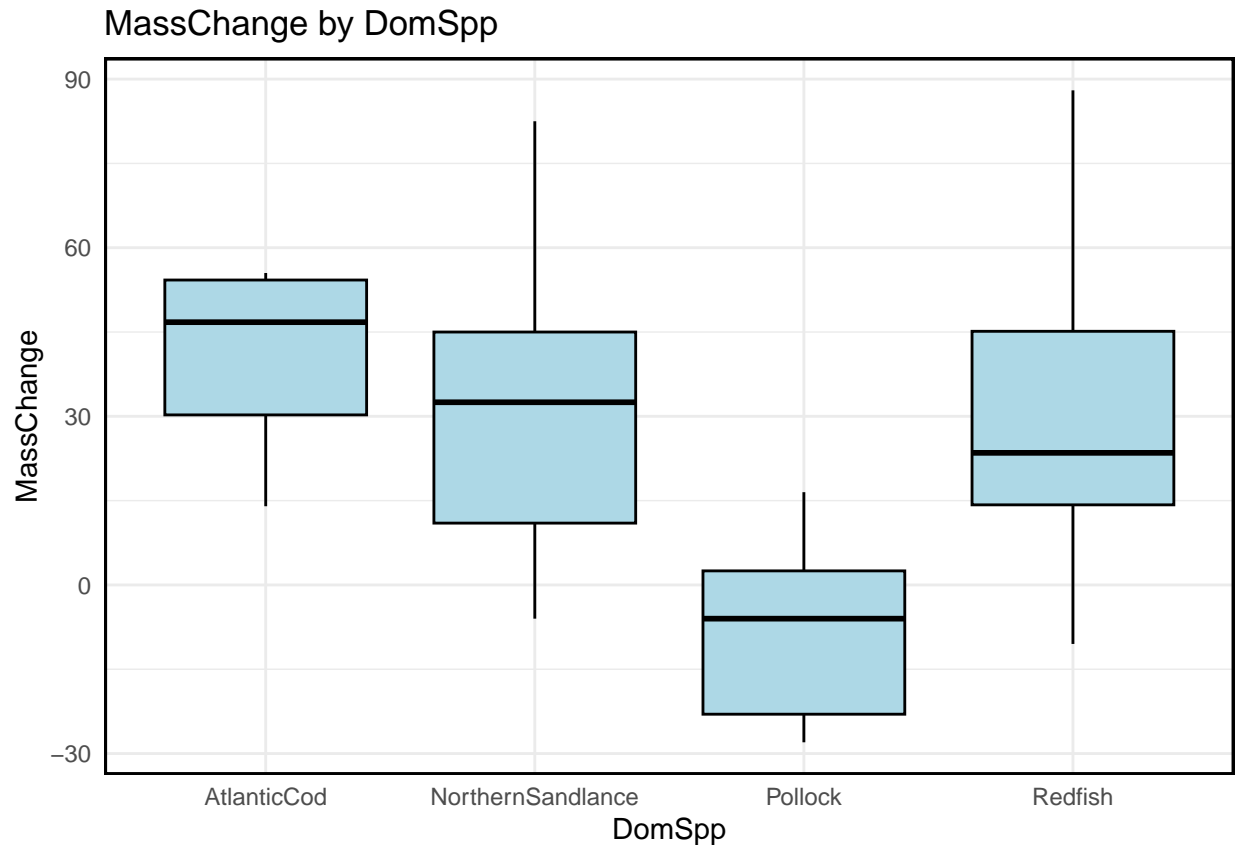
Initially, it appears that diet diversity is negatively related to mass change, with energy density possibly having a slight negative relationship. It also appears that animals with cod, redfish or sand lance as the dominant prey species gain more mass during foraging than females who primarily prey on Pollock. There does not appear to be any clear trends in mass change over time.

We will begin with an initial model with all covariates of interest. Based on the distribution of mass change, we will begin with a simple model with a gaussian distribution and an identity link function:

```
library(lme4)
```

```
## Loading required package: Matrix
```
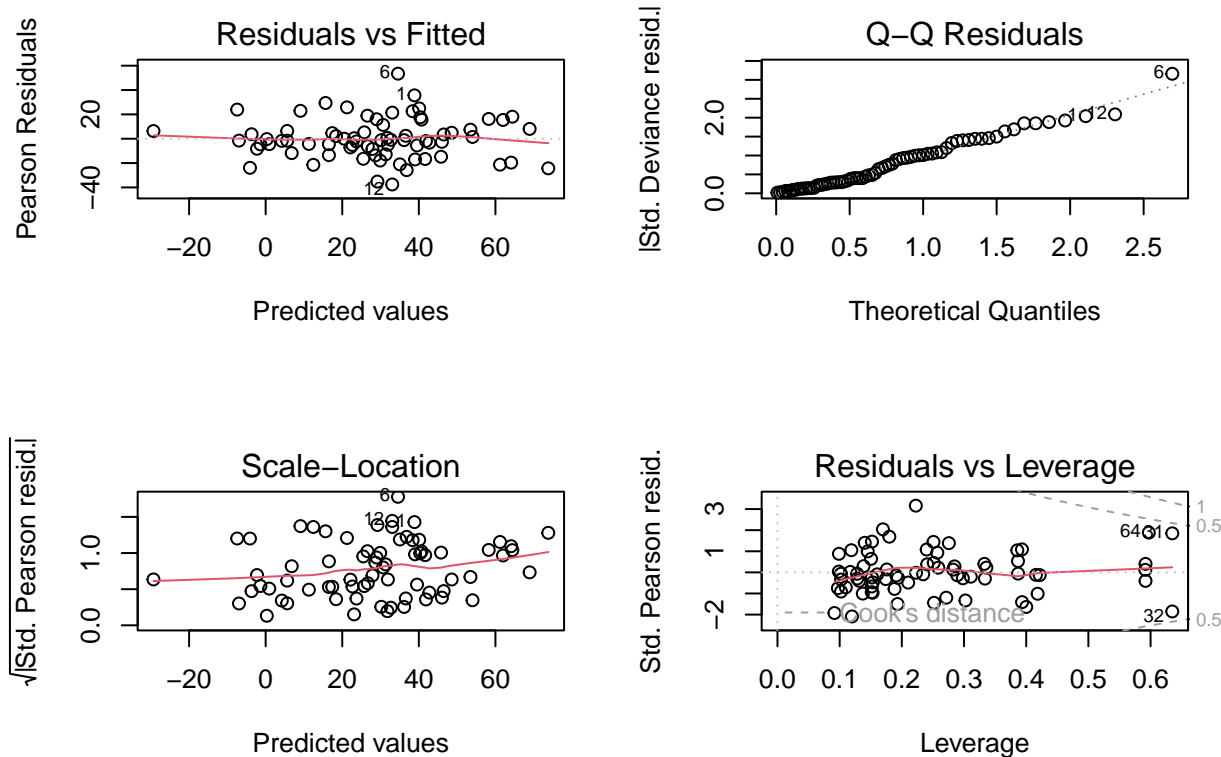
```
mom_mod <- glm(data=Data_Q1, family = gaussian(link="identity"), formula = MassChange ~ DietDiv + DietEn
```

```
par(mfrow = c(2, 2))
plot(mom_mod)
```

```
## Warning: not plotting observations with leverage one:
##    45
```

Residuals vs Fitted

Pearson Residuals

6
1

12

Predicted values

Q–Q Residuals

|Std. Deviance resid.|

6

12

Theoretical Quantiles

Scale–Location

√|Std. Pearson resid.|

6
12 1

Predicted values

Residuals vs Leverage

Std. Pearson resid.

1
0.5
64 1
Cook's distance
32
0.5

Leverage

The residual plots from this initial model show good homoscedasticity and normality. There is one observation (observation 6) that appears to be an extreme value. This observation comes from a female that experiences an 88 kg increase in mass over the pre-breeding period foraging period, which is well above the mean mass change value of 28.5. However, none of the covariate values associated with this observation are abnormal and this point does not have high leverage. As a results, the decision is to retain this data point since it is not highly influential and knowledge of the system suggests it is a valid data point. Since the model residuals look good, there is no need to add complexity by using a different link function to generalize the model.

We can now use step selection to preform repeated hypothesis tests for variable selection to determine which covariates best explain maternal mass change while minimizing model complexity and colinearity between covariates.

```
step(mom_mod, direction = "both")
```

```
## Start:  AIC=647.55
## MassChange ~ DietDiv + DietEngDen + Year + DomSpp
##
##               Df Deviance    AIC
## - DietEngDen   1    19508 645.67
## <none>              19478 647.55
## - DietDiv      1    20858 650.48
## - DomSpp       3    23417 654.82
## - Year        13    38977 671.50
##
## Step:  AIC=645.67
```
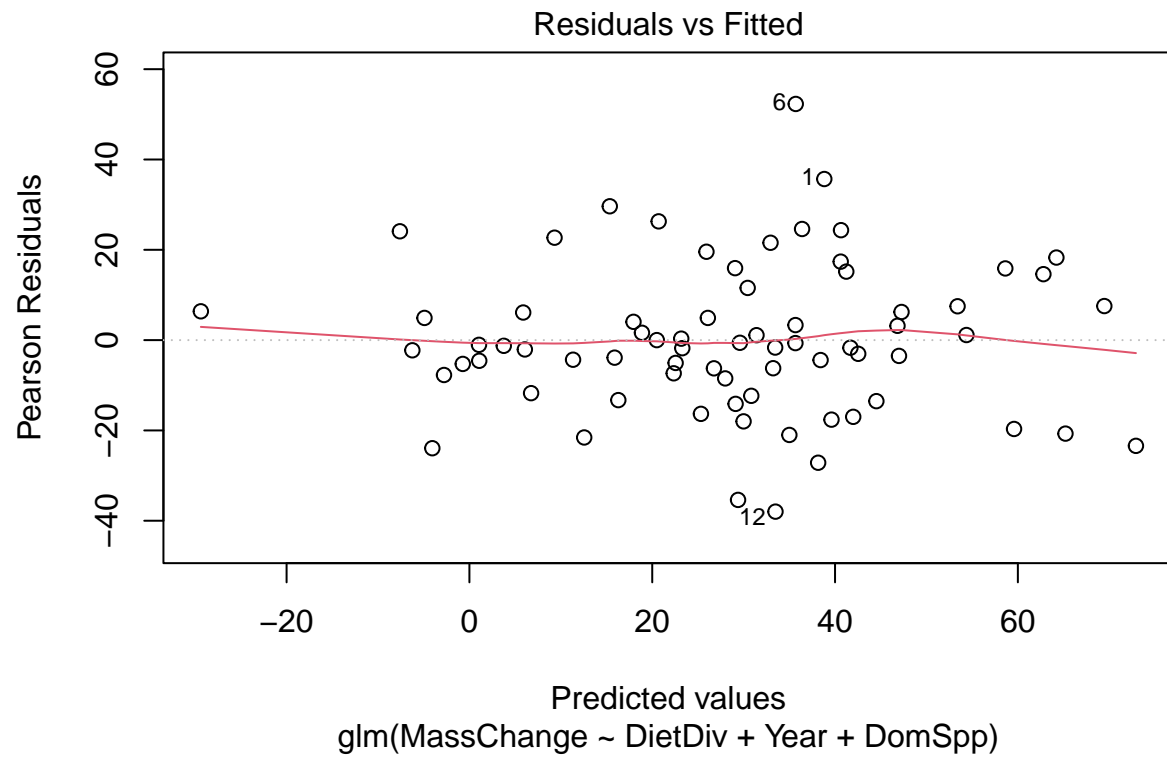
```
## MassChange ~ DietDiv + Year + DomSpp
##
##              Df Deviance    AIC
## <none>           19508 645.67
## + DietEngDen  1   19478 647.55
## - DietDiv      1   20891 648.60
## - DomSpp       3   23602 653.38
## - Year        13   39769 670.95
```
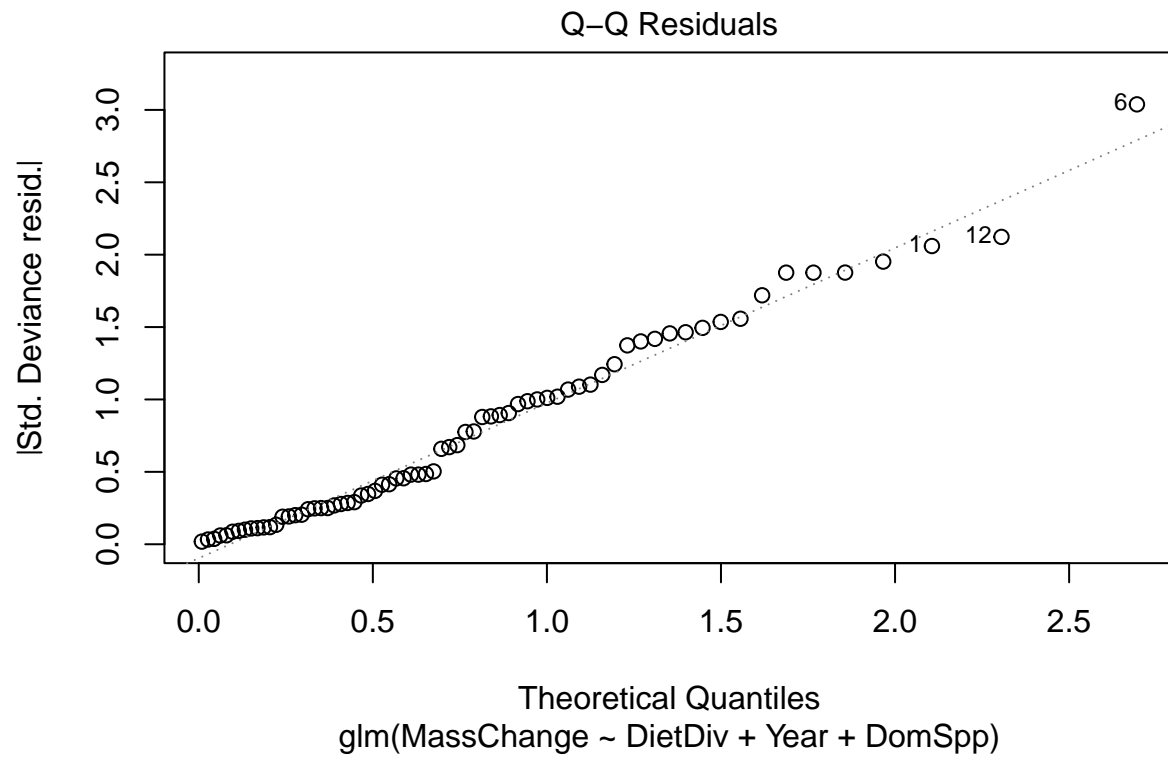
```
##
## Call:  glm(formula = MassChange ~ DietDiv + Year + DomSpp, family = gaussian(link = "identity"),
##     data = Data_Q1)
##
## Coefficients:
##           (Intercept)                 DietDiv                  Year.L
##                70.599                 -63.415                  28.367
##                Year.Q                  Year.C                  Year^4
##                12.387                  18.307                  15.184
##                Year^5                  Year^6                  Year^7
##               -15.549                   3.305                  23.361
##                Year^8                  Year^9                 Year^10
##                 2.392                 -23.277                 -11.186
##               Year^11                 Year^12                 Year^13
##                13.261                  21.328                 -39.502
## DomSppNorthernSandlance          DomSppPollock           DomSppRedfish
##               -18.212                 -47.479                 -20.303
##
## Degrees of Freedom: 71 Total (i.e. Null);  54 Residual
## Null Deviance:       48570
## Residual Deviance: 19510     AIC: 645.7
```
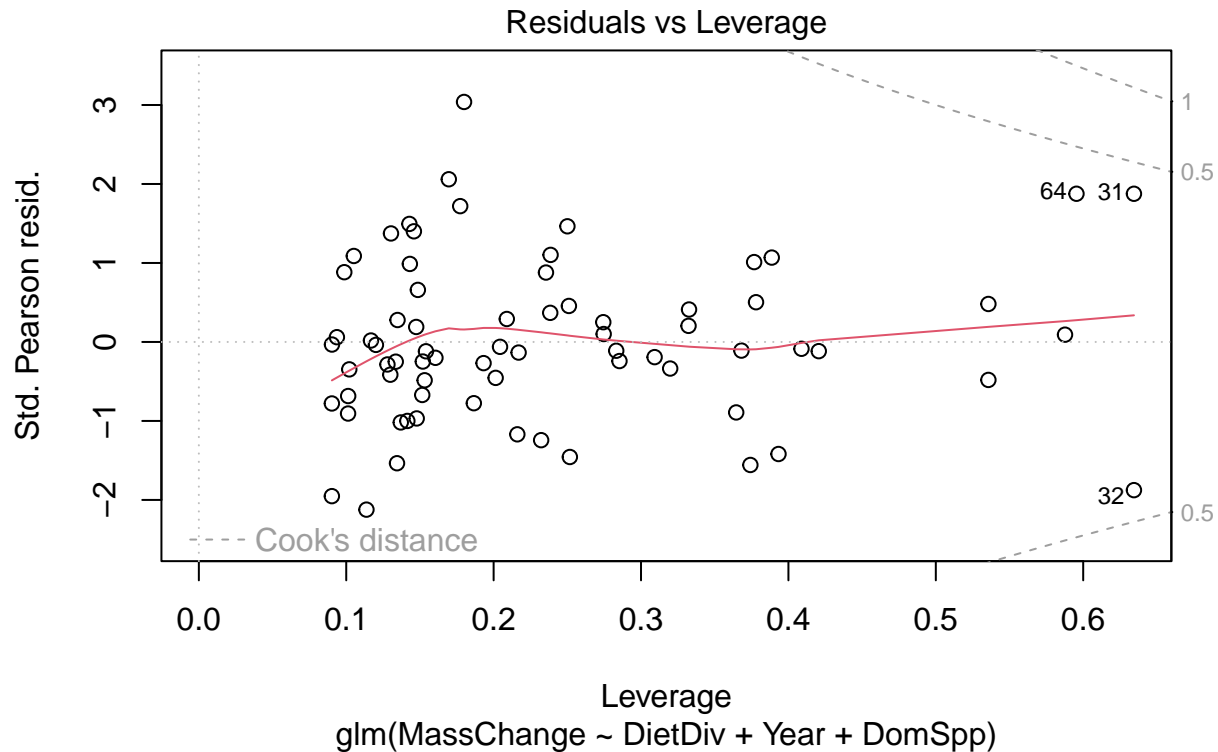
The lowest AIC model drops the dietary energy density covariate. We will update the initial model to drop the energy density covariate:

```r
mom_mod2 <- glm(formula = MassChange ~ DietDiv + Year + DomSpp, family = gaussian(link = "identity"),
                data = Data_Q1)
plot(mom_mod2)
```

```
## Warning: not plotting observations with leverage one:
##   45
```

Residuals vs Fitted

glm(MassChange ~ DietDiv + Year + DomSpp)

Q–Q Residuals

Theoretical Quantiles
glm(MassChange ~ DietDiv + Year + DomSpp)

Scale−Location

√|Std. Pearson resid.|

Predicted values
glm(MassChange ~ DietDiv + Year + DomSpp)

16

Residuals vs Leverage

glm(MassChange ~ DietDiv + Year + DomSpp)

All model residuals still look good. We can proceed to interpreting model output:

```
summary(mom_mod2)
```

```
##
## Call:
## glm(formula = MassChange ~ DietDiv + Year + DomSpp, family = gaussian(link = "identity"),
##     data = Data_Q1)
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       70.599     16.241   4.347 6.15e-05 ***
## DietDiv          -63.415     32.417  -1.956 0.055621 .
## Year.L            28.367     15.657   1.812 0.075584 .
## Year.Q            12.387     12.620   0.982 0.330717
## Year.C            18.307     13.274   1.379 0.173517
## Year^4            15.184     14.787   1.027 0.309065
## Year^5           -15.549     12.334  -1.261 0.212857
## Year^6             3.305     10.924   0.303 0.763373
## Year^7            23.361     10.525   2.219 0.030674 *
## Year^8             2.392     10.067   0.238 0.813123
## Year^9           -23.277     10.057  -2.314 0.024475 *
## Year^10          -11.186     10.451  -1.070 0.289261
## Year^11           13.261      9.842   1.347 0.183483
## Year^12           21.328      8.874   2.403 0.019708 *
## Year^13          -39.502      9.864  -4.005 0.000191 ***
```

```
## DomSppNorthernSandlance  -18.212      14.318  -1.272 0.208824
## DomSppPollock            -47.479      16.434  -2.889 0.005550 **
## DomSppRedfish            -20.303      13.938  -1.457 0.150992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 361.2663)
##
##     Null deviance: 48567  on 71  degrees of freedom
## Residual deviance: 19508  on 54  degrees of freedom
## AIC: 645.67
##
## Number of Fisher Scoring iterations: 2
```

Cross validation We can perform cross-validation to to test this models predictions. In order to have each fold represent about 10% of the data, we will perform 7 fold cross validation:

```
set.seed(123)

#Number of folds
k <- 7

# Split the data into folds
DatCV <- Data_Q1 %>% mutate(fold = sample(rep(1:k, length.out = n())))

# Create a vector to store RMSE for each fold
rmse_values <- numeric(k)

#Perform cross validation
for (i in 1:k) {
  train_data <- DatCV %>% filter(fold != i)
  test_data  <- DatCV %>% filter(fold == i)

  # Fit the model on training data
modCV <- glm(MassChange ~ DietDiv + Year + DomSpp, family = gaussian(link = "identity"), data = Data_Q1)

  # Predict on the test data
  predictions <- predict(modCV, newdata = test_data)

  # Compute RMSE for this fold
  actuals <- test_data$MassChange
  rmse_values[i] <- sqrt(mean((predictions - actuals)^2))
}

rmse_values
```

```
## [1] 22.44760 11.88442 18.05650 11.05129 18.98355 15.02906 14.37965
```

```
mean_rmse <- mean(rmse_values)
cat("Average RMSE across", k, "folds:", round(mean_rmse, 2), "\n")
```

```
## Average RMSE across 7 folds: 15.98
```

The average root mean squared error is 15.98, indicating the model generally predicts values within 16 kg of the true value.

Results:

Question1: **Does variation in diet influence maternal mass change over the foraging period leading up to parturition in Northwest Atlantic grey seals?**

After using step selection methods to determine the best model using diet to explain variation in maternal mass change during the foraging period leading up to parturition, the lowest AIC model retains diet diversity, year, and dominant prey species as significant covariates. Looking at the summary output, even though several years have mass change values that significantly differ from the expected mass change value, the effect of time is not of direct interest to our research question and is only present to account for changes in prey dynamics over time. As a result, the interpretation of this output is not relevant. However, we see that females who feed primarily on pollock experience significantly lower ($p<0.05$) mass change than the reference level, with the reference level in this case being females who primarily feed on cod. Being an identity link function, the model is essentially a simple linear model and there is no transformations applied to the response variable. As such, coefficient estimates can be interpreted directly. Individuals who feed on pollock experienced 47.5% lower mass change than individuals who primarily feed on cod, while females who fed primarily on redfish or sand lance experienced similar mass change to females who fed on cod. This effect is presented in figure 1.

```
ggplot(Data_Q1, aes(x = DomSpp, y = MassChange, fill = DomSpp)) +
  geom_boxplot() +
  stat_summary(
    fun = "mean", geom = "errorbar",
    aes(ymax = ..y.., ymin = ..y..),
    width = 0.75, color = "red", size = 1.2
  ) + geom_text(x = "Pollock", y = 20, label = "*",
                aes(x = x, y = y, label = label),
                color = "black", size = 6, fontface = "bold"
  ) +
  labs(
    title = "Figure 1. Median (black line) and mean (red line) mass change by dominant prey species",
    x = "Prey species",
    y = "Mass change (kg)", fill = "Dominant prey species"
  ) +
  theme_minimal() + theme(
    axis.line = element_line(color = "black", size = 1),  # Add axis lines
    axis.text.x = element_text(size = 12, angle = 45, hjust = 1)
  ) +
  scale_y_continuous(
    breaks = seq(0, max(Data_Q1$MassChange), by = 25) - 50
  )
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Warning: The 'size' argument of 'element_line()' is deprecated as of ggplot2 3.4.0.
## i Please use the 'linewidth' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Warning: The dot-dot notation ('..y..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(y)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```


Figure 1. Median (black line) and mean (red line) mass change by dominar
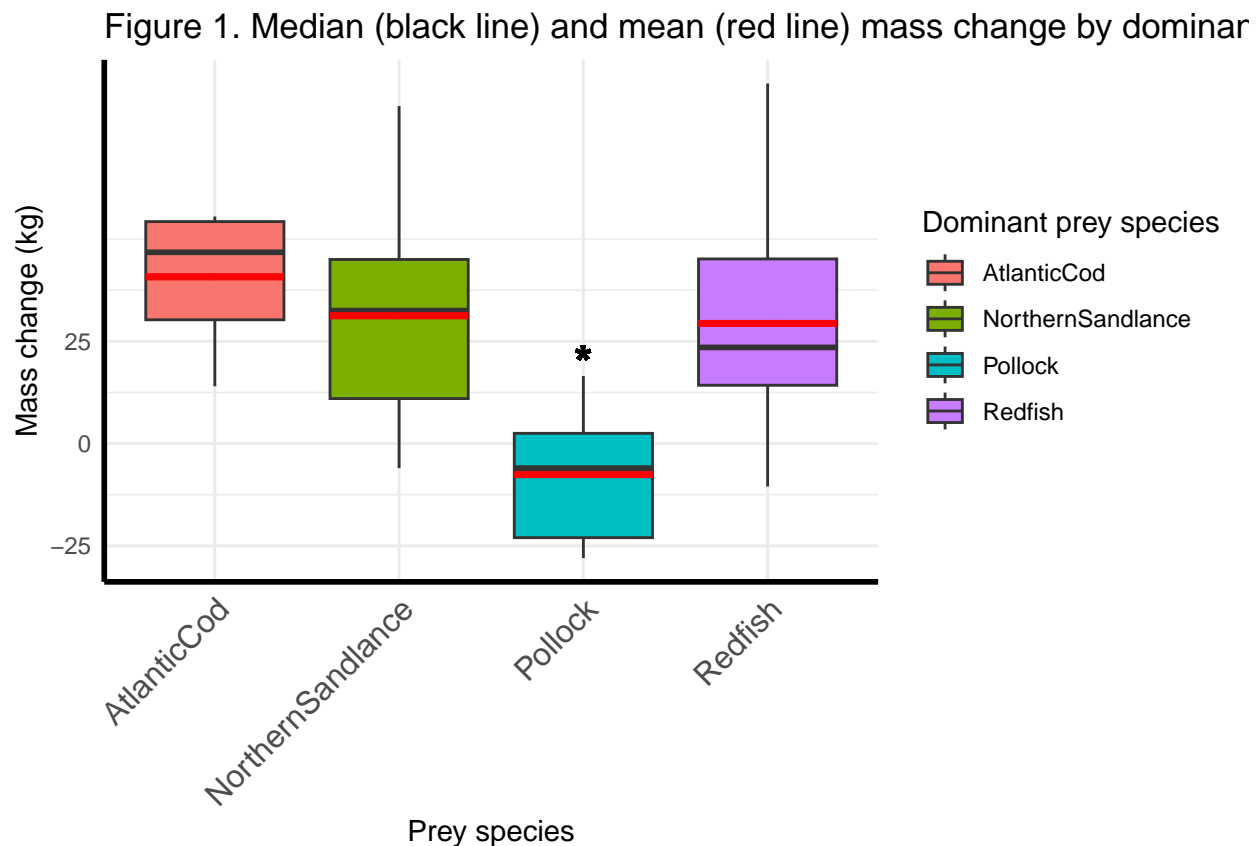
It may also be worth noting that diet diversity is almost significant at p = 0.0556. While this p-value is not significant, the negative relationship between diet diversity and mass change may support the observed effects of dominant prey species, where individuals who primarily prey on cod, sand lance, or redfish are focusing efforts on these few profitable species, as reflected by more successful mass gain than individuals who feed on a more diverse array of prey species which include less profitable species. As a result, reduced diet diversity may reflect focus on highly profitable prey species and increased diet diversity may reflect a greater proportion of less profitable prey species.

Cross-validation was used to varify the predictive power of the final model. The average root mean squared error (RMSE) after 7-fold cross-validation is 15.98, indicating the model generally predicts values within 16 kg of the true value. We can check the standard deviation to compare the average RMSE to:

```
sd(Data_Q1$MassChange)
```

```
## [1] 26.15433
```

Compared to the standard deviation of mass change, the RMSE value is about 40% lower than the standard deviation, meaning the model predictions are better than using the population mean alone to predict mass change. However, compared to the overall mean mass change within our dataset (28.5 kg), there is relatively large prediction error. We can also visualize prediction accuracy with a fitted vs observed value plot (figure 2):

```r
fit_mom <- fitted(mom_mod2)

# Get the observed response variable (actual values)
obs_mom <- Data_Q1$MassChange

# Plot the fitted vs. observed values
ggplot(data = Data_Q1, aes(x = fit_mom, y = obs_mom)) +
  geom_point(color = "black") +
  geom_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed") +  # Add 1:1 line
  labs(
    title = "Figure 2. Fitted vs. Observed Values with 1:1 line",
    x = "Fitted Values",
    y = "Observed Values"
  ) +
  theme_minimal()
```


Figure 2. Fitted vs. Observed Values with 1:1 line
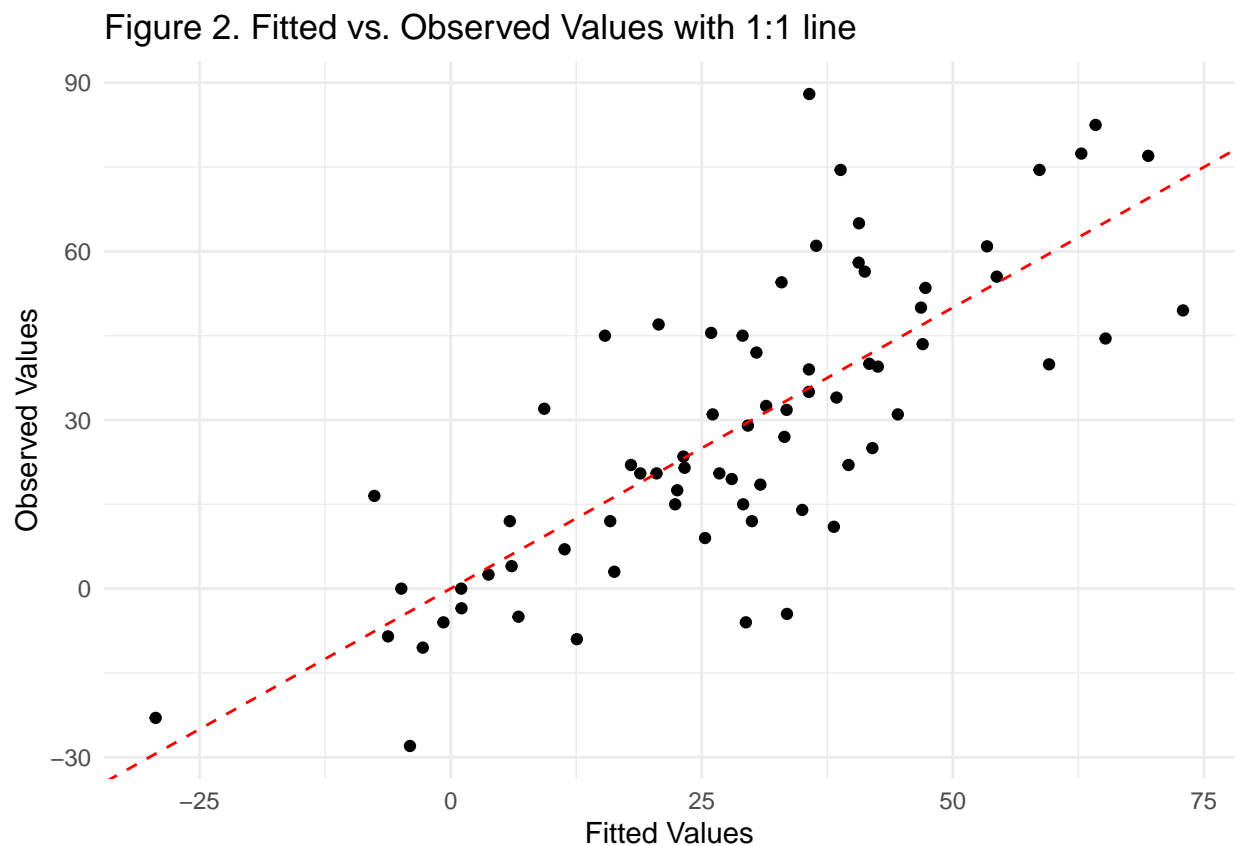
While figure 2 shows that the model predictions have consistent accuracy across the range of fitted values, as suggested by the average RMSE value obtained from cross-validation, there is reasonably large prediction

error. Given that the model residuals suggest good model fit, the relatively large prediction error suggest the set of covariates used in this model are not sufficient in predicting maternal mass change. From this we can conclude that females who feed on redfish, sand lance, or cod experience greater mass change than individuals who feed on Pollock, but other factors outside of maternal diet, such as individual physiology or characteristics associated with individual foraging behavior, are likely responsible for explaining a large proportion of variation in maternal mass change.

Question 2: **Does female diet predict pup weaning mass in Northwest Atlantic grey seals?**

Although it was proposed that many predictor variables (2 continuous + 3 categorical for question one and 2 continuous + 5 categorical for question two) would allow greater computational potential to establish a well fit model, the added complexity of these variables was observed to significantly increase the AIC and potentially overfit the model. It was found, particularly in question two, that the simplicity of a Linear Model with a Least Square approach resulted in a simpler linear model with a significantly lower AIC.

In question 2 it was found that best fit Linear Model for Pup Wean Mass included the predictor variables Dietary Energy Density and Diet Diversity with an AIC = 178.496. When a Generalized Linear Model was explored with a Gaussian Family and Identity Link Function it was also found that the best fit model included Dietary Energy Density and Diet Diversity predictor variable but the AIC was much less preferable at 322.39. The reasons for this decreased predicted fit is likely based on the more complex Maximum Likelihood Estimation used in Generalized Linear Models rather than the Linear Least Square method used in the Linear Models. Generalized Linear Mixed Models (GLMM) were also explored using the Dominant Prey Species as the Cluster Variable because it was assumed that this variable would likely have a effect on the prey energy density as well as the diet diversity based on feeding location ecosystems. The best fit GLMM was found to include Diet Diversity and Dietary Energy Density as fixed effects predictor variables and Dominant Prey Species as an intercept random effect variable which had a AIC of 324.3899. As a result it was determined that the best fit model to determine Pup Wean Mass was the Linear Model and this model was tested using Cross Validation. The correlation of predicted values to actual values showed a value of 0.6596914 meaning the Linear Model could be expected to predict Pup Wean Mass with about a 66 percent accuracy.

Conclusion:

In both question one, that explored the effect of diet on Maternal Mass Change, and question two, that explored the predictive ability of maternal diet on pup wean mass at partition, we found relatively large predictive errors in our models (cross-validation revealed RMSE of +/- 16kg for question 1 and 33% for question two). Although the data-set provided many predictive variables, many of these were categorical data and only two, Dietary Energy Density and Diet Diversity, where numeric in nature. In addition to this the number of observation in the dataset were limited (question one n=76, and question two n=56) which resulted in many rows of data but few columns with ultimately resulted in added complexity in our models and the need to reduce the number of predictor variable to prevent overfitting the model.

In question one the residuals suggest good model fit but the relatively large prediction error suggest the set of covariates used in this model are not sufficient in predicting maternal mass change or more observation would be required to fit a more reliable model. It was found from the data that females who fed on redfish, sand lance, or cod experience greater mass change than individuals who feed on Pollock, but other factors outside of maternal diet, such as individual physiology or characteristics associated with individual foraging behavior are likely responsible for a large proportion of variation in maternal mass change. Diet diversity appeared to have negative relationship with maternal mass change, possibly indicating a narrow diet focusing on more profitable prey species is beneficial in increasing mass gain, although these results were not quite statistically significant.

In question two it was determined through both AIC and Cross Validation that the simpler Linear Model had a higher accuracy for predicting Pup Wean Mass with the two available numeric predictor variables. Access to additional observation could increase the predictive ability of the Linear Model as well as allow the fitting of more complex model such as a Mixed Model without overfitting. Although we feel the models

provided fair accuracy in predicting Pup Wean Mass at partition, the lack of number of observations may have significantly impacted the reliability and validity of the statistical analysis.

References:

Data source: Sara Iverson laboratory and DFO Grey Seal Program.

Beck, C.A., Iverson, S.J., Bowen, W.D., and Blanchard, W. 2007. Sex differences in grey seal diet reflect seasonal variation in foraging behaviour and reproductive expenditure: evidence from quantitative fatty acid signature analysis. J. Anim. Ecol. 76(3): 490–502.

Bowen, W.D., Iverson, S.J., McMillan, J.I., and Boness, D.J. 2006. Reproductive performance in grey seals: age-related improvement and senescence in a capital breeder. J. Anim. Ecol. 75(6): 1340–1351.

Bowen, W.D., den Heyer, C.E., McMillan, J.I., and Iverson, S.J. 2015. Offspring size at weaning affects survival to recruitment and reproductive performance of primiparous gray seals. Ecol Evol. 5(7): 1412-1424.

Bowen, W.D. and Jonsen, I.D. 2022. Foraging Ecology and Behavior. In: Costa, D.P., McHuron, E.A. (eds) Ethology and Behavioral Ecology of Phocids. Ethology and Behavioral Ecology of Marine Mammals. Springer, Cham.

Iverson, S.J., Bowen, W.D., Boness, D.J., and Oftedal, O.T. 1993. The Effect of Maternal Size and Milk Energy Output on Pup Growth in Grey Seals (Halichoerus grypus). Physiol. Zool. 66(1): 61–88.

Iverson, S.J., Field, C., Bowen, W.D., and Blanchard, W. 2004. Quantitative fatty acid signature analysis: a new method of estimating predator diets. Ecol. Monogr. 74(2): 211-235.

Iverson, S.J. 2009. Tracing aquatic food webs using fatty acids: from qualitative indicators to quantitative determination. In Lipids in aquatic ecosystems. Edited by M. Arts, M.T. Brett, and M. Kainz. Springer, New York. pp. 281–308.

Karnovsky, N., Hobson, K.A., and Iverson, S.J. 2012. From lavage to lipids: Estimating diets of seabirds. Mar. Ecol. Prog. Ser. 451: 263-284.

Mellish, J.E., Iverson, S.J., and Bowen, W.D. 1999. Variation in milk production and lactation performance in grey seals and consequences for pup growth and weaning characteristics. Physiol. Biochem. Zool. 72(6): 677–690.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., and Smouse, P.E. 2008 A movement ecology paradigm for unifying organismal movement research. Proc. Natl. Acad. Sci. USA. 105(49): 19052–19059.

Noren, S.R., Boness, D.J., Iverson, S.J., McMillan, J.I., and Bowen, W.D. 2008. Body condition at weaning affects the duration of the postweaning fast in gray seal pups (Halichoerus grypus). Physiol. Biochem. Zool. 81: 269–277.

Stephens, P.A., Boyd, I.L., McNamara, J.M. and Houston, A.I. 2009. Capital breeding and income breeding: their meaning, measurement, and worth. Ecology. 90(8): 2057-2067. Stephens, P.A., Houston, A.I., Harding, K.C., Boyd, I.L. and McNamara, J.M. 2014. Capital and income breeding: the role of food supply. Ecology. 95(4): 882-896.

Add all R packages used to the description file

```r
used_pkgs <- c("lme4", "here", "readr", "usethis")


library(usethis)
for (pkg in used_pkgs) {
  try(use_package(pkg, type = "imports"), silent = TRUE)
}
```

```
## v Setting active project to 'C:/Users/Maxhe/OneDrive - Dalhousie
## University/Documents/STAT5620.Project'
```