

Predicted Body Mass of Penguins

By

Huiting Wu
Claudio Hernandez Garcia
Michelle Joyce Pira
Thu Tran

Date

May 1st, 2023

1. Abstract:

In this analysis, we use “Palmer penguins” dataset created by Dr. Kristen Gorman’s dataset who is a Marine Biology’s professor. We use different statistic method to build a multiple linear model that help us to estimate the body mass of penguins living in Antarctica by using species, island, sex, and other measurements of penguins’ bill and flipper. To predict the body mass, we can use the variables such as species, year, sex, bill length and flipper length.

2. Problem and Motivation:

The body mass of a penguin is particularly important as it impacts their survival and reproductive success. Larger penguins are better able to dive deeper and catch more food, as well as survive harsh winter conditions and raise chicks. Understanding body mass is crucial for conservation efforts to ensure the long-term survival of penguin populations.

The motivation for the project was to better understand the ecology and behavior of these penguin species, as well as to monitor changes in their populations over time. Body mass is a particularly important variable, as it can affect a penguin’s ability to forage, reproduce, and survive harsh environmental conditions. Penguin body mass plays a crucial role in their survival and reproduction. According to Williams and Naveen (2019), body mass affects penguin behavior, including foraging and breeding success. Larger penguins are better able to dive deeper and stay underwater for longer periods of time, allowing them to catch more food. Additionally, heavier penguins have a greater chance of surviving harsh winter conditions and are better equipped to raise chicks. For example, in a study of Adélie penguins, larger individuals were found to have higher chick survival rates (Jenouvrier et al., 2019). Therefore, understanding the importance of body mass is essential for conservation efforts to ensure the long-term survival of penguin populations.

3. Data Description:

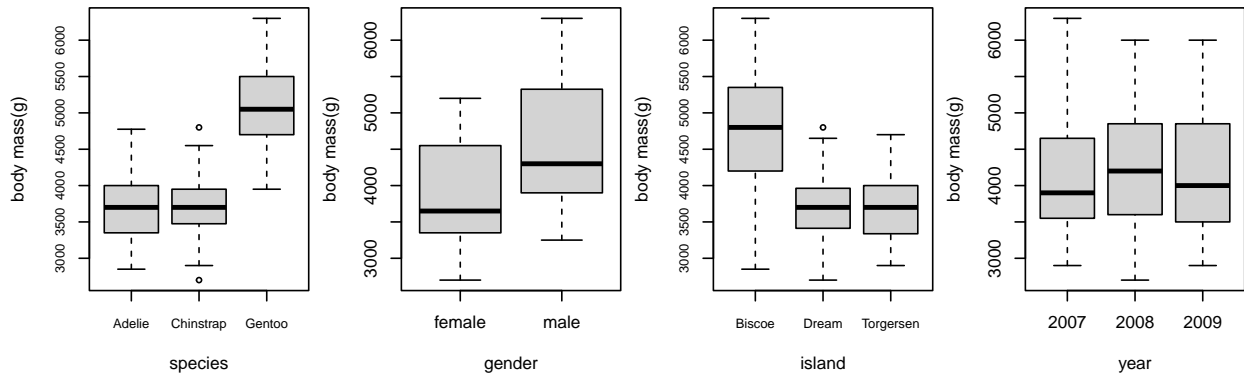
- **Data Sources & Data Description:** Dr. Kristen Gorman’s “palmer penguins” data set is a built-in data set in the R library. There are two data sets with this package. Penguins_raw which contains 17 variables, and a simplified version of penguins_raw, just called penguins has 8 variables. The project used the package of penguins. This data set contains information on three species of penguins - Adelie, Chinstrap, and Gentoo. This data was gathered from the Palmer Archipelago in Antarctica. The data set includes measurements of body mass, flipper length, bill length, and other characteristics for each penguin. The data set originally published in *Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins* by Gorman KB, Williams TD and Fraser WR in 2014. There are 344 penguins recorded and 8 variables. There are some missing values in the data set. To remove the missing data because it didn’t give enough information to predict the body mass of the penguin. We remove 11 penguins’ records.

Table 1: Description of the penguins dataset

Predictors	Data_Type	Description
Species	Character	a factor denoting penguin species (Adélie, Chinstrap and Gentoo)
Island	Character	a factor denoting island in Palmer Archipelago, Antarctica (Biscoe, Dream or Torgersen)
Bill Length_mm	Numerical	a number denoting bill length (millimeters)
Bill Depth	Numerical	a number denoting bill depth (millimeters)
Flipper Length	Numerical	an integer denoting flipper length (millimeters)

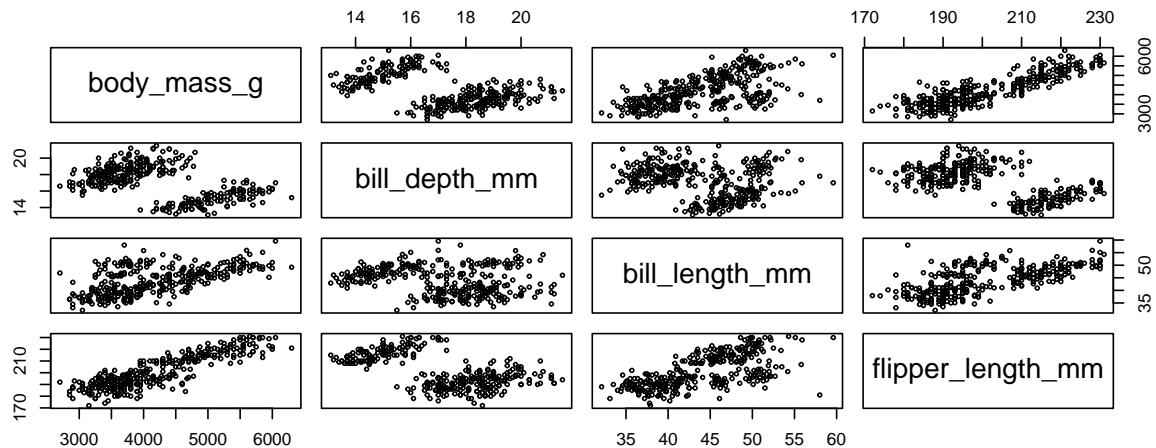
Predictors	Data_Type	Description
Body Mass	Numerical	an integer denoting body mass (grams)
Sex	Character	a factor denoting penguin sex (female, male)
Year	Integers	an integer denoting the study year (2007, 2008, or 2009)

- **Visualization of the variables:** There are some variables that are characteristic and some of them are numerical. Although the variable of Year is recorded as integer, we used it as a character variable since the data only contains from 2007 to 2009. The Box-Plot is useful to compare the Body Mass and the characteristic variables (*Graph 1: Box-plots of body mass on different character predictors*). The results from the box plots show the Gentoo species has a higher body mass than other species. Male penguins have a higher body mass than Female penguins. The penguins in Biscoe Island have higher body mass than other islands. In 2008, the penguin population has a higher body mass than in 2007 and in 2009.



Graph 1: Box-plots of body mass on different character predictors

Graph 3: Scatter plots of body mass on different numerical predictors show the relationships between numerical variables. Graph 3 shows that the Bill Depth has no linear relationship with Body Mass. The Bill Length and Flipper Length seem to have a positive linear relationship with Body Mass. Bill Depth neither has a linear relationship with Bill Length nor with Flipper Length. The Bill Length and Flipper Length seem to have a positive relationship with each other.



Graph 3: Scatter plots of body mass on different numerical predictors

4. Questions of Interest:

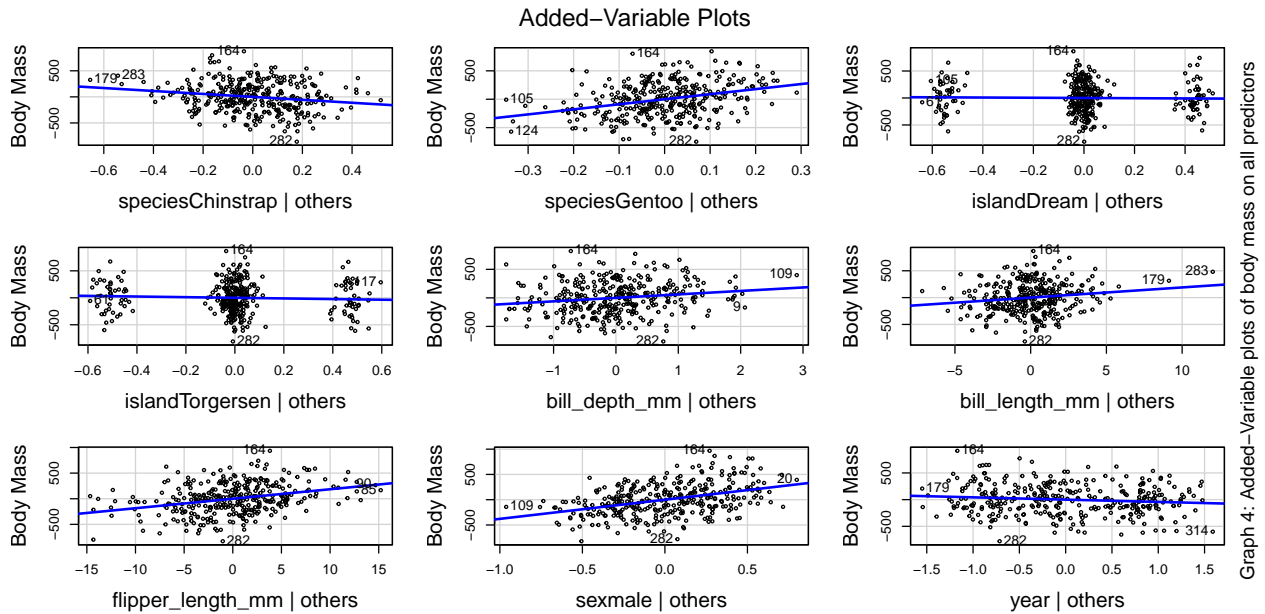
What factors are most strongly associated with the body mass of penguins, and how accurately can we predict their body mass using these predictors?

5. Regression Analysis, Results, and Interpretation:

- **Important Details of the Analysis & Exploratory Analysis:**

The first multiple linear regression model (*full model*) is Body Mass on Species, Island, Bill Length, Bill Depth, Flipper Length, Sex, and Year. The dataset contains 8 variables listed above. The full model simply contains all the variables which are Body Mass as response and the remaining 7 variables as predictors ($R^2 = 0.8734$). In order to have the most suitable model, the variable selection process is taken into the following analysis.

Checking the relationship between Body Mass and all other variables with the added-variable graph (Graph 4: Added-Variable plots mass on all predictors). Since The regression line is most the same as Body Mass equal 0 shown on the graph of different Islands verse Body Mass, there is no relationship between variables of Body Mass and Island. All other variables show a relationship with Body Mass in Graph 4 which supported our analysis in Graph 3.



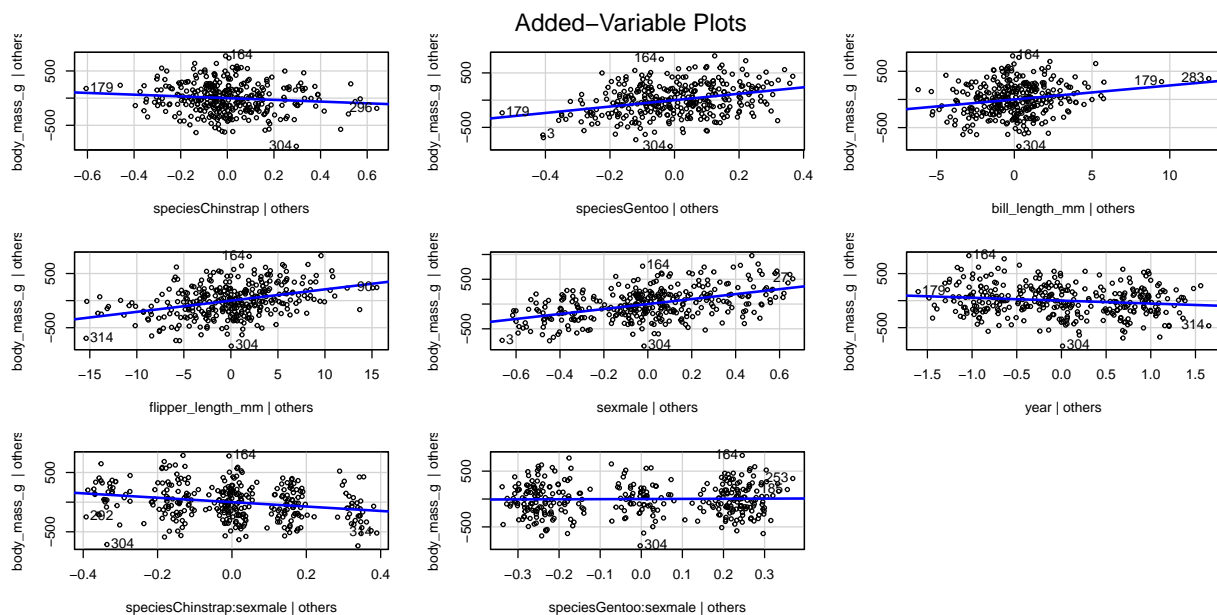
Graph 4: Added-Variable plots of body mass on all predictors

To remove the Island variable with the result of the ANOVA table (*Null hypothesis: model removed the Island variable; Alternative hypothesis: the full model which contains all the variables; P-value = 0.6168* $> \alpha = 0.05$, *F-value = 0.484, fail to reject the null hypothesis, keep the reduced model*) and the results of t-test for the measurement of the predictor Island. (*Null hypothesis: the measurement of the predictor Island = 0; alternative hypothesis: measurement of the predictor Island $\neq 0$. T-test for island of Dream: t-value = -0.363, p-value = 0.71704* $> \alpha = 0.05$, and *t-test for island of Torgersen: t-value = -0.966, p-value = 0.33482* $> \alpha = 0.05$. Both fail to reject the null hypothesis) All levels of Island are not significant. Remove the variable of Island.

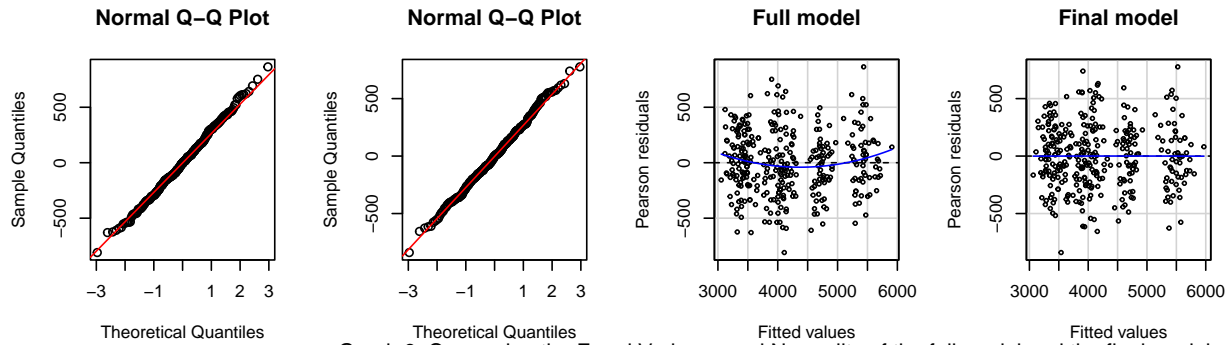
To remove the Bill Depth variable with the result of comparing the adjusted R^2 of the model remove Island only ($R^2 = 0.8738$) and the model remove Island and Bill Depth together ($R^2 = 0.8706$) which doesn't change a lot. It indicates that the both model has a similar variability of Body Mass can explained by all the predictors collectively. We want a model with fewer predictors therefore we choose to remove both Island and Bill Depth.

Since the t-test of the measurement of Year is slightly important (*null hypothesis: measurement of Year = 0; alternative hypothesis: measurement of Year $\neq 0$; t-value = -2.042, P-value = 0.04194 $< \alpha = 0.05$, reject the null hypothesis*) which indicates it should be included in the model. We double check it with comparing the adjusted R^2 of the model remove Island, Bill Depth, and Year ($R^2 = 0.8685$) and the model remove Island and Bill Depth ($R^2 = 0.8706$). It is decreasing if we remove Year which indicated the less variability of Body Mass can be explained by all the predictors collectively. Therefore, we keep the variable of Year.

The ***final model*** with response as Body Mass and the predictors are Species, Bill Length, Flipper Length, Sex, Year, and the interaction of Species and Sex.



- Independent: The sample of penguins is randomly selected.
- Normality: As comparing the full model and the final model, we can obviously see more points falling along the quantile-quantile line (*Graph 9: Comparing the the Equal Variance and Normality of the full model and the final model*).
- Equal variance: The variance of the residuals is roughly constant across all values of the fitted values. The assumption of equal variance is met based on the residual plot showing that most of the points evenly distributed around the line of residual equals 0. There is a pattern in the residual plot because the model contains some characteristic variables.



Graph 9: Comparing the Equal Variance and Normality of the full model and the final model

- **Interpretation:** The equation to predict body mass: $E(\hat{BodyMass}) = 106016 - 159.4 \text{ SpeciesChinstrap} + 588.5 \text{ SpeciesGentoo} + 24.85 \text{ Bill Length} + 20.74 \text{ Flipper Length} + 501 \text{ SexMale} - 53.52 \text{ Year} - 371.2 \text{ SpeciesChinstrap} * \text{SexMale} + 23.46 \text{ SpeciesGentoo} * \text{SexMale}$.

Table 20 (*Table 20: the measurements of all the predictors of the final model in Appendix 2*) shows that most of the measures of the predictions are important, but the interaction between Gentoo species and male penguins is not important. However, the interaction between Chinstrap species and male penguins is important, so we decided to keep the interaction between Species and Sex of penguins in their analysis.

This decision is supported by the box-plot of species v.s. body mass, which shows that Gentoo penguins have a higher mean weight than other species. Additionally, the numeric variables such as Bill Length and Flipper Length have a positive linear correlation with body mass, meaning that as these measurements increase, so does body mass.

We also observed a negative relationship between study year and body mass. This indicates that the weight of penguins has decreased over time, which is concerning and suggests that penguin populations may be at risk.

6. Conclusions:

This project used the “palmer penguins” data set of Dr. Kristen Gorman to predict the body mass of penguins by utilizing multiple linear regression with Species, Bill Length, Flipper Length, Sex, Year, and the interaction of Species and Sex as predictors. The final multiple linear models states that there is 87.91% variability in body mass can be explained by Species, Bill Length, Flipper Length, Sex, Year, and the interaction of Species and Sex collectively.

The final model can be used to predict the body mass of penguins based on the values of the predictors. Such predictions can be helpful in various fields, such as conservation biology, where understanding the body mass of penguins can assist scientists in making informed decisions about their management and conservation. During the analysis, a decline in the body mass of penguins was observed over the study years. This decline can be an indicator of potential threats to the penguin population, such as changes in food availability, climate change, or pollution. By utilizing this information, scientists can identify such threats and develop strategies to mitigate them and protect the penguin population.

In conclusion, this project demonstrates the potential of multiple linear regression to predict body mass and its significance in informing conservation efforts. The findings can serve as a starting point for further research and as a tool for decision-makers in the field of conservation biology.

7. Appendices:

Appendix 1 : R Code

```
#DataSet input
library(palmerpenguins)

library(tidyverse)

# Print the descriptions of the penguins data as a table
Predictors <- c("Species", "Island", "Bill Length_mm", "Bill Depth",
               "Flipper Length", "Body Mass", "Sex", "Year")
Description <- c("a factor denoting penguin species (Adélie, Chinstrap and Gentoo)",
                 "a factor denoting island in Palmer Archipelago, Antarctica
                 (Biscoe, Dream or Torgersen)",
                 "a number denoting bill length (millimeters)",
                 "a number denoting bill depth (millimeters)",
                 "an integer denoting flipper length (millimeters)",
                 "an integer denoting body mass (grams)",
                 "a factor denoting penguin sex (female, male)",
                 "an integer denoting the study year (2007, 2008, or 2009)")
Data_Type <- c("Character", "Character", "Numerical",
               "Numerical", "Numerical", "Numerical",
               "Character", "Integers")
table1 <- data.frame(Predictors, Data_Type, Description)
library(pander)
pander(table1, caption = "Description of the penguins dataset")

# remove the null value of the dataset
penguins1 <- na.omit(penguins)
# attache the dataset after removing the null value
attach(penguins1)
# full model
model <- lm(body_mass_g ~ species + island + bill_depth_mm
            + bill_length_mm + flipper_length_mm + sex + year, penguins1)
# null model
null.model <- lm(body_mass_g ~ 1, penguins1)
# summary of the full model
summary(model)

## Visulization of the Dataset
### Character Predictors with Response
par(mfrow = c(1,4), mar = c(8,4,4,1))
plot(species, body_mass_g, cex.axis = 0.6, gap.axis = -1,
     xlab = "species", ylab = "body mass(g)")
plot(body_mass_g ~ sex, xlab = "gender", ylab = "body mass(g)",
     cex.axis = 1, penguins1)
# Box-plot of Character Predictors versus Response
mtext("Graph 1: Box-plots of body mass on different character predictors",
     side = 1, line = 6, font = 1, cex = 1)
plot(island, body_mass_g, cex.axis = 0.6, gap.axis = -1,
     xlab = "island", ylab = "body mass(g)")
```

```
fyear <- as.factor(penguins1$year)
plot(fyear, penguins1$body_mass_g, xlab = "year", ylab = "body mass(g)")
```

```
# Sample size of different levels of the character predictors
Species <- penguins %>%
  count(species)
Island <- penguins %>%
  count(island)
Sex <- penguins %>%
  count(sex)
Table2 <- data.frame(Species, Island, Sex)
pander(Table2,
  caption = "Sample size of different levels of character predictors")
```

```
# Means of numerical predictors of different levels of character predictors
library(dplyr)
p <- penguins %>%
  group_by(species) %>%
  summarize(across(where(is.numeric), mean, na.rm = TRUE))
i <- penguins %>%
  group_by(island) %>%
  summarize(across(where(is.numeric), mean, na.rm = TRUE))
s <- penguins %>%
  group_by(sex) %>%
  summarize(across(where(is.numeric), mean, na.rm = TRUE))
pander(p, caption = "Mean of Numerical predictors of different Specifics",
  split.table = Inf)
pander(i, caption = "Mean of Numerical predictors of different Islands",
  split.table = Inf)
pander(s, caption = "Mean of Numerical predictors of different Sex")
```

```
library(car)
```

```
# Multi-collinearity of numerical predictors
## By correlation matrix
### delete delete the character predictors columns from the data
penguins2 <- penguins1[, c(-1,-2,-6,-7,-8)]
### correlation between predictors
correlation <- round(cor(penguins2), 2)
pander(correlation, caption = "Correlation Matrix of Predictors")
```

```
## By scatter plot matrix
scatterplotMatrix(penguins2)
mtext("Graph 2: Scatter plots of different numerical predictors",
  side = 1, line = 4, font = 1, cex = 1)
```

```
### Relationships between all predictors and between numerical response and predictors
pairs(~ body_mass_g + bill_depth_mm + bill_length_mm + flipper_length_mm,
  penguins1, cex = 0.5)
mtext("Graph 3: Scatter plots of body mass on different numerical predictors",
  side = 1, line = 4, font = 1, cex = 1)
```



```

par(cex.lab = 0.8)
avPlots(model, cex = 0.5)
mtext("Graph 4: Added-Variable plots of body mass on all predictors",
      side = 1, line = 4, font = 1, cex = 0.8)

```

```

plot(body_mass_g~bill_depth_mm,penguins1, mar = c(2,2,2,2))
mtext("Graph 5: Scatter plots of body mass on predictor Bill Depth",
      side = 1, line = 4, font = 1, cex = 1)

```

Variable Selection by manually

```

fit <- lm(body_mass_g ~ species + bill_depth_mm + bill_length_mm +
          flipper_length_mm + sex + year, penguins1)
anova(fit, model)

```

```

fit1 <- lm(body_mass_g ~ species + bill_length_mm + flipper_length_mm +
           sex + year, penguins1)
Anova_Table <- anova(fit1, fit)
pander(Anova_Table, caption = "ANOVA Test on Model **with Bill Dength** and **without Bill Dength**")

```

```

fit2 <- lm(body_mass_g ~ species + bill_length_mm + flipper_length_mm + sex, penguins1)
anova(fit2, model)

```

Compare the 3 models's R^2

```

Name <- c("Model remove Island", "Model remove Island and Bill Depth",
          "Model remove Island, Bill Depth, and Year", "Full model")
Radj_squared <- c(summary(fit)$adj.r.squared, summary(fit1)$adj.r.squared,
                  summary(fit2)$adj.r.squared, summary(model)$adj.r.squared )
Rtable <- data.frame(Name, Radj_squared)
pander(Rtable, caption = "Goodness of Fit of different models")

```

Variable selection by automatically:

```

n <- nrow(penguins1)
back.bic <- step(model, direction = "backward", trace = 0, k = log(n))
back.bic

```

```

back.aic <- step(model, direction = "backward", trace = 0)
back.aic

```

```

forward.aic <- step(model, direction = "forward", trace = 0,
                   scop = list(lower=null.model, upper = model))
forward.aic

```

```

forward.bic <- step(model, direction = "forward", trace = 0,
                   k = log(n), scop = list(lower=null.model, upper = model))
forward.bic

```

Compare the two models

fit 2 models

```

model.bbic <- lm(body_mass_g ~ species + bill_length_mm + bill_depth_mm + flipper_length_mm
                 + sex, penguins1)
model.baic <- lm(body_mass_g ~ species + bill_length_mm + bill_depth_mm + flipper_length_mm

```

```

+ sex + year, penguins1)
# compare two models
anova(model.bbic, model.baic)
# compare with island(full model) and without island(reduced model)
anova(model.baic, model)
# compare the  $R^2$  of full model, model without island
c(summary(model)$adj.r.squared, summary(model.baic)$adj.r.squared)

```

Assumptions Checking

linearity

```

pairs(body_mass_g ~ bill_length_mm + flipper_length_mm, penguins1)
mtext("Graph 6: Linearity assumptions checking by Scatter plots
      between Body Mass, Bill Length, and Flipper Length",
      side = 1, line = 4, font = 1, cex = 0.8)

```

constant variance and normality

```

par(mfrow = c(1,2), cex.lab = 0.8, cex.main = 1)
plot(model, 1:2)
mtext("Graph 7: Normality and Constant Variance assumptions checking by
      Residual Plot and Normal QQ Plot",
      side = 4, line = 1, font = 1, cex = 0.8)

```

```
summary(fit1)
```

Add interaction terms into the final model

```

Interaction_model <- lm(body_mass_g ~ species + bill_length_mm +
                        flipper_length_mm + sex + year +
                        species*bill_length_mm + species*flipper_length_mm +
                        species*sex + species*year +
                        bill_length_mm*flipper_length_mm + bill_length_mm*sex +
                        bill_length_mm*year + flipper_length_mm*sex+
                        flipper_length_mm*year + sex*year , penguins1)
summary(Interaction_model)

```

```

Interaction_model1 <- lm(body_mass_g ~ species + bill_length_mm +
                        flipper_length_mm + sex + year + species*sex, penguins1)
summary(Interaction_model1)

```

Compare Final Model without

```

## interaction term v.s. the final model with interaction of Species and Sex
Anova_Table1 <- anova(fit1, Interaction_model1)
pander(Anova_Table1, caption = "Final Model without interaction term v.s.
the final model with interaction of Species and Sex")

```

```

s <- summary(Interaction_model1)
pander(s, caption = "The measurements of all the predictors of the final model")

```

```

name = "Final Model with Interaction"
Radj_Squared = summary(Interaction_model1)$adj.r.squared
Rtable <- data.frame(name, Radj_Squared)
pander(Rtable, caption = "The Goodness Of Fit")

```

```
## assumption check for final model
```

```
# Linearity
```

```
avPlots(Interaction_model1)
```

```
# Residual Plot
```

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

```
residualPlot(model, cex = 0.5, main = "Full model")
```

```
residualPlot(Interaction_model1, cex = 0.5, main = "Final model")
```

```
# Normal QQplot
```

```
qqnorm(resid(model))
```

```
qqline(resid(model), col = "red", cex=0.5)
```

```
qqnorm(resid(Interaction_model1))
```

```
qqline(resid(Interaction_model1), col = "red", cex=0.5)
```

```
## ScatterPlotMatrix of the Final Model with interaction term
```

```
scatterplotMatrix(~ body_mass_g + species +  
                  species*sex + bill_length_mm +  
                  flipper_length_mm + sex + year, penguins1, cex = 0.3)  
mtext("Graph 10: Scatter Plots of the Final Model",  
      side = 4, line = 1, font = 1, cex = 0.8)
```

```
## Prediction
```

```
# Test for the random penguins:
```

```
set.seed(888)
```

```
random<-sample(1:nrow(penguins1),3)
```

```
x1<-data.frame(species= penguins1$species[random[1]],
```

```
bill_length_mm=penguins1$bill_length_mm[random[1]],
```

```
flipper_length_mm=penguins1$flipper_length_mm[random[1]],
```

```
sex=penguins1$sex[random[1]] ,
```

```
year=penguins1$year[random[1]])
```

```
x2<-data.frame(species= penguins1$species[random[2]],
```

```
bill_length_mm=penguins1$bill_length_mm[random[2]],
```

```
flipper_length_mm=penguins1$flipper_length_mm[random[2]],
```

```
sex=penguins1$sex[random[2]] ,
```

```
year=penguins1$year[random[2]])
```

```
x3<-data.frame(species= penguins1$species[random[3]],
```

```
bill_length_mm=penguins1$bill_length_mm[random[3]],
```

```
flipper_length_mm=penguins1$flipper_length_mm[random[3]],
```

```
sex=penguins1$sex[random[3]] ,
```

```
year=penguins1$year[random[3]])
```

```
pred1<-predict(Interaction_model1,newdata=x1,interval="prediction")
```

```
pred2<-predict(Interaction_model1,newdata=x2,interval="prediction")
```

```
pred3<-predict(Interaction_model1,newdata=x3,interval="prediction")
```

```
# table
```

```
Random_table<-data.frame(  
Penguin_number=random,  
Species = c(species[random[1]],species[random[2]],species[random[3]]),  
Bill_length= c(bill_length_mm[random[1]],bill_length_mm[random[2]],bill_length_mm[random[3]]),
```

```

Flipper_length=c(flipper_length_mm[random[1]],flipper_length_mm[random[2]],flipper_length_mm[random[3]])
Sex =c(sex[random[1]],sex[random[2]],sex[random[3]]),
Year =c(year[random[1]],year[random[2]],year[random[3]])
)
pander(Random_table, caption = "Random selected variables")

# table
Predict_table<-data.frame(Penguin_number=random,
Actual=c(penguins1$body_mass_g[random[1]],
penguins1$body_mass_g[random[2]],
penguins1$body_mass_g[random[3]]),
Fit= c(pred1[, "fit"],pred2[, "fit"],pred3[, "fit"]),
Lower=c(pred1[, "lwr"],pred2[, "lwr"],pred3[, "lwr"]),
Upper=c(pred1[, "upr"],pred2[, "upr"],pred3[, "upr"])
)
pander(Predict_table, caption = "The Predict table")

detach(penguins1)

```

Appendix 2: Figures & Tables & Test Results

Tables

Table 2: Sample size of different levels of character predictors

species	n	island	n.1	sex	n.2
Adelie	152	Biscoe	168	female	165
Chinstrap	68	Dream	124	male	168
Gentoo	124	Torgersen	52	NA	11

Table 3: Mean of Numerical predictors of different Specifics

species	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	year
Adelie	38.79	18.35	190	3701	2008
Chinstrap	48.83	18.42	195.8	3733	2008
Gentoo	47.5	14.98	217.2	5076	2008

Table 4: Mean of Numerical predictors of different Islands

island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	year
Biscoe	45.26	15.87	209.7	4716	2008
Dream	44.17	18.34	193.1	3713	2008
Torgersen	38.95	18.43	191.2	3706	2008

Table 5: Mean of Numerical predictors of different Sex

sex	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	year
female	42.1	16.43	197.4	3862	2008
male	45.85	17.89	204.5	4546	2008
NA	41.3	16.64	199	4006	2008

Table 6: Correlation Matrix of Predictors

	bill_length_mm	bill_depth_mm	flipper_length_mm
bill_length_mm	1	-0.23	0.65
bill_depth_mm	-0.23	1	-0.58
flipper_length_mm	0.65	-0.58	1

Table 7: Random selected variables

Penguin_number	Species	Bill_length	Flipper_length	Sex	Year
136	Adelie	40.6	187	male	2009
193	Gentoo	50.5	225	male	2008
267	Chinstrap	50	196	male	2007

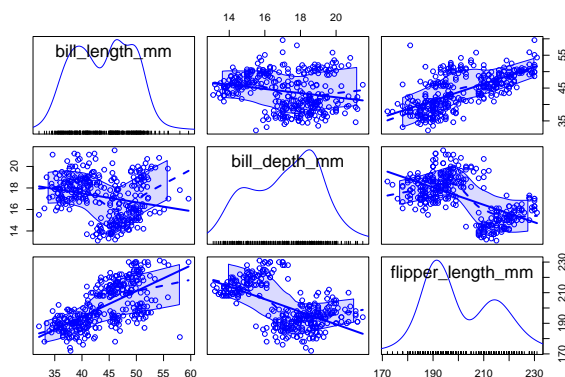
Table 8: Goodness of Fit of different models

Name	Radj_squared
Model remove Island	0.8738
Model remove Island and Bill Depth	0.8706
Model remove Island, Bill Depth, and Year	0.8685
Full model	0.8734

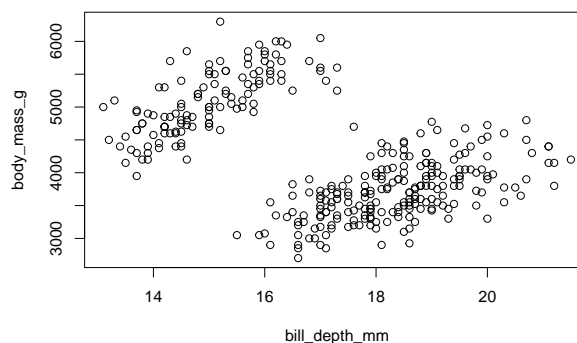
Table 9: The Goodness Of Fit

name	Radj_Squared
Final Model with Interaction	0.8791

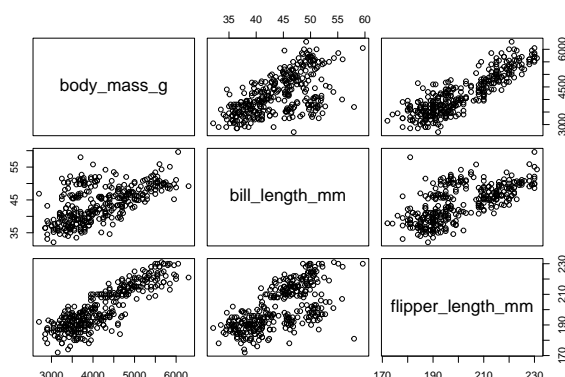
Figures



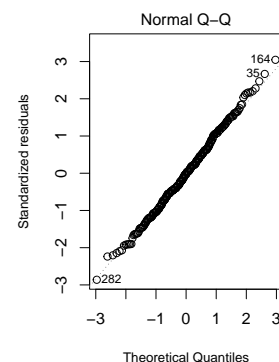
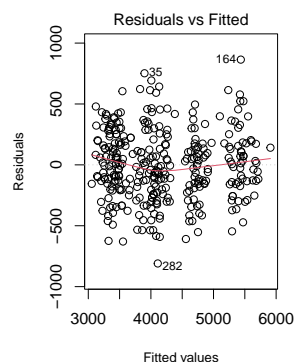
Graph 2: Scatter plots of different numerical predictors



Graph 5: Scatter plots of body mass on predictor Bill Depth



Graph 6: Linearity assumptions checking by Scatter plots between Body Mass, Bill Length, and Flipper Length



Graph 7: Normality and Constant Variance assumptions checking by Residual Plot and Normal QQ Plot

Test Results

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	84088	41912	2.006	0.04566
speciesChinstrap	-282.5	88.79	-3.182	0.001604
speciesGentoo	891	144.6	6.163	2.124e-09
islandDream	-21.18	58.39	-0.3627	0.717
islandTorgersen	-58.78	60.85	-0.9659	0.3348
bill_depth_mm	60.8	20	3.04	0.002563
bill_length_mm	18.96	7.112	2.667	0.00805
flipper_length_mm	18.5	3.128	5.915	8.465e-09
sexmale	379	48.07	7.883	4.949e-14
year	-42.78	20.95	-2.042	0.04194

Table 11: Summary Analysis of important of all the predictors

Observations	Residual Std. Error	R^2	Adjusted R^2
333	286.5	0.8768	0.8734

Table 12: ANOVA Test on Model **with Bill Depth** and **without Bill Depth**

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
326	27357143	NA	NA	NA	NA
325	26596486	1	760657	9.295	0.002487

Table 13: ANOVA Test on Model **with Island** and **without Island**

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
325	26596486	NA	NA	NA	NA
323	26517018	2	79468	0.484	0.6168

Table 14: Goodness of Fit of different models

Name	Radj_squared
Model remove Island	0.8738
Model remove Island and Bill Depth	0.8706
Model remove Island, Bill Depth, and Year	0.8685
Full model	0.8734

Table 15: Table continues below

	Estimate	Std. Error	t value
(Intercept)	-644633	1371033	-0.4702
speciesChinstrap	-6542	2e+05	-0.03271
speciesGentoo	-208347	242047	-0.8608
bill_length_mm	-3640	17655	-0.2062
flipper_length_mm	4788	7160	0.6688
sexmale	93093	109033	0.8538
year	316.6	683.7	0.463
speciesChinstrap:bill_length_mm	7.486	19.14	0.3912
speciesGentoo:bill_length_mm	44.19	39.21	1.127
speciesChinstrap:flipper_length_mm	24.5	14.85	1.65
speciesGentoo:flipper_length_mm	11.34	13.41	0.8459
speciesChinstrap:sexmale	-641	254	-2.524
speciesGentoo:sexmale	-388	306.2	-1.267
speciesChinstrap:year	0.7365	100	0.007363
speciesGentoo:year	102.1	121	0.8436
bill_length_mm:flipper_length_mm	-1.269	1.247	-1.017
bill_length_mm:sexmale	11.36	19.87	0.5716
bill_length_mm:year	1.94	8.823	0.2199
flipper_length_mm:sexmale	10.62	7.27	1.461
flipper_length_mm:year	-2.354	3.568	-0.6597
sexmale:year	-47.32	54.61	-0.8665

	Pr(> t)
(Intercept)	0.6386
speciesChinstrap	0.9739
speciesGentoo	0.39
bill_length_mm	0.8368
flipper_length_mm	0.5041
sexmale	0.3939
year	0.6437
speciesChinstrap:bill_length_mm	0.696
speciesGentoo:bill_length_mm	0.2606
speciesChinstrap:flipper_length_mm	0.09994
speciesGentoo:flipper_length_mm	0.3983
speciesChinstrap:sexmale	0.0121
speciesGentoo:sexmale	0.2061
speciesChinstrap:year	0.9941
speciesGentoo:year	0.3996
bill_length_mm:flipper_length_mm	0.3098
bill_length_mm:sexmale	0.568
bill_length_mm:year	0.8261
flipper_length_mm:sexmale	0.1449
flipper_length_mm:year	0.5099
sexmale:year	0.3869

Table 17: Summary Analysis of the important of Interaction Terms

Observations	Residual Std. Error	R^2	Adjusted R^2
333	281.6	0.8851	0.8777

Table 18: Final Model without interaction term v.s. the final model with interaction of Species and Sex

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
326	27357143	NA	NA	NA	NA
324	25393312	2	1963831	12.53	5.742e-06

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	106016	40281	2.632	0.008896
speciesChinstrap	-159.4	85.66	-1.861	0.06369
speciesGentoo	588.5	96.99	6.067	3.626e-09
bill_length_mm	24.85	6.884	3.61	0.0003551
flipper_length_mm	20.74	2.981	6.959	1.91e-11
sexmale	501	51.74	9.684	1.203e-19
year	-53.52	20.15	-2.656	0.008309
speciesChinstrap:sexmale	-371.2	83.18	-4.463	1.117e-05
speciesGentoo:sexmale	23.46	70.3	0.3337	0.7389

Table 20: The measurements of all the predictors of the final model

Observations	Residual Std. Error	R^2	Adjusted R^2
333	280	0.882	0.8791

8. Reference & Source:

Jenouvrier, S., Caswell, H., Barbraud, C., Holland, M., Stroeve, J., & Weimerskirch, H. (2019). Demographic models and IPCC climate projections predict the decline of an emperor penguin population. *Proceedings of the National Academy of Sciences*, 116(49), 24526-24531.

Williams, T. D., & Naveen, R. (2019). Body mass in penguins: a reflection of species-specific adaptations, ecological and physiological constraints, and niche differentiation. *The Auk*, 136(2), 1-23.

Originally published in: Gorman KB, Williams TD, Fraser WR (2014) Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus *Pygoscelis*). *PLoS ONE* 9(3): e90081. doi:10.1371/journal.pone.0090081

Adélie penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Adélie penguins (*Pygoscelis adeliae*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 5. Environmental Data Initiative. doi: 10.6073/pasta/98b16d7d563f265cb52372c8ca99e60f

Gentoo penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Gentoo penguin (*Pygoscelis papua*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 5. Environmental Data Initiative. doi: 10.6073/pasta/7fca67fb28d56ee2ffa3d9370ebda689

Chinstrap penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Chinstrap penguin (*Pygoscelis antarcticus*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 6. Environmental Data Initiative. doi: 10.6073/pasta/c14dfcfa8ea13a17536e73eb6fbe9e

Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0. <https://allisonhorst.github.io/palmerpenguins/>. doi: 10.5281/zenodo.3960218.

Lescroël, A., Schmidt, A., Elrod, M. et al. Foraging dive frequency predicts body mass gain in the Adélie penguin. *Sci Rep* 11, 22883 (2021). <https://doi.org/10.1038/s41598-021-02451-4>

How do Penguins Keep Warm? *Science of the Cold*. https://www.coolantarctica.com/Antarctica%20fact%20file/science/cold_penguins.php

René Groscolas, The use of body mass loss to estimate metabolic rate in fasting sea birds: A critical examination based on emperor penguins (*Aptenodytes forsteri*), *Comparative Biochemistry and Physiology Part A: Physiology*, Volume 90, Issue 3, 1988, Pages 361-366, ISSN 0300-9629, [https://doi.org/10.1016/0300-9629\(88\)90203-4](https://doi.org/10.1016/0300-9629(88)90203-4). (<https://www.sciencedirect.com/science/article/pii/0300962988902034>)