BIO 539 – Big Data Analysis
Dr. Rachel Schwartz
Due: April 26th, 2021

Independent Research Project
Michael Andranovich

Introduction

Population and fitness trends of large carnivorous predators can play major roles in their ecosystems, having significant effects on the composition of entire communities by consuming lower trophic levels, and by altering the behavior or habitat selection of prey (Silliman and Angelini, 2012). Although labeled as marine mammals because their habitat and foraging behavior is reliant on ice floes in the Arctic Sea, polar bears (*Ursus maritimus*) are also classified as the largest land carnivore in the world, leading to the need to conduct extensive and important research. Polar bear populations face many threats including the impacts of climate change leading to the loss of sea ice habitat, ingestion of high levels of pollutants through their food, and human activities like oil exploration (World Wildlife Foundation, n.d.). As one of four marine mammal species managed by the U.S. Department of Interior, polar bears and their population dynamics require long-term research to inform local, state, national and international policy makers regarding conservation of the species and its habitat (Atwood et al., n.d.). Because assessing changes in population sizes for polar bears can be challenging, trends in weight and body condition can be a beneficial indicator to estimate population status and serve as an early warning sign of population change (Rode et al., 2020).

This independent research project set out to investigate two subpopulations of polar bears: one on the Chukchi Sea (ranges between the northwest coast of Alaska and eastern Russia) and the other on the Southern Beaufort Sea (ranges between the North Slope of Alaska and western Canada). For these populations, two topics were explored using the collected data (Rode, 2020) and statistical analysis through the R Project. First, the weight trends of these two populations over the course of the study (1981 – 2017) were calculated and examined through statistical and visual outputs. Understanding how polar bear weights are fluctuating (or not) over the course of the study is important because sustained success of a healthy population of large mammals usually depends on size, fitness, and metabolism, especially for females (Sergio et al., 2005). The list of threats to polar bears is growing, making this trend information critical for resource managers. Second, we set to identify which measured morphological attribute of polar bears plays the largest role in predicting total body mass. We analyzed the available recorded measurements for body length, axillary girth, skull width, tail lengths, etc., with linear regression modeling and selection based on AIC weights. Morphometric equations to estimate body mass are a powerful tool for monitoring trends in body condition over time, especially for a species like polar bears where it can be difficult to assess whole population trends (Gittleman, 1986; Thiemann et al., 2011).

Methods

The data used for this analysis was obtained from captured polar bears between 1971 and 2016 during studies conducted by U.S. Geological Survey's Alaska Science Center and U.S. Fish and Wildlife

Service. This dataset included over 3,300 captures of all sex and age classes, with about 2,300 captures being unique individuals. Although the populations studied occurred in the Southern Beaufort Sea (SB) and the Chukchi Sea (CS), the former area produced almost 75% of the data points used. For all years of the study, polar bears were located using a helicopter or fixed-wing aircraft and immobilized with a rapid-injection dart containing a mix of tranquilizing drugs (depending on the year). Locations of capture areas were primarily on the sea ice with the exception of some fall captures on land (Rode et al., 2020).

In order to produce a comprehensive list of data points to analyze, we set some filters on the data. To make sure that each model created was using the same amount of data, we focused on capture records that had complete data entries for our independent variables: population location (SB or CS), year, sex (M or F), approximate age, body length (cm), axillary girth (cm), and skull width (cm). We then created an additional independent variable called "Years Since", which calculated when the capture occurred in relation to the beginning of the study (1981). This type of variable added another temporal element to measure against instead of basic year data. For all aspects of this analysis, the dependent variable was the total body mass (kg) of the polar bear captured. After conducting these filters, we were left with a dataset that contained 1,227 complete entries from the years 1983 to 2011, each with the eight test variables and a total body mass measurement.

To understand how the weights of polar bears were trending, our team visually inspected generated statistical plots. The first of these plots grouped the recorded body mass of polar bears by year and averaged them. This allowed us to create a graph which showed the average body mass of captured polar bears by year over the length of our filtered dataset. We also created a plot which showed the average body mass of polar bears, broken up over standardized "epochs". For this analysis, we grouped the twenty-eight study years into three sections based on Early (9 years), Middle (9 years), and Most Recent (10 years) measurements. Lastly, we looked at the trend of polar bear weights broken up by sex (male and female) and epoch. This plot allowed us to examine if male and female weights were trending similarly, and if one sex contributed more to the overall trend of weights compared to the other.

In order to identify which morphological trait contributes the most towards predicting total body mass, we constructed numerous models and created an Akaike's Information Criterion (AIC) table to compare and evaluate the usefulness of each model. An AIC table uses the number of independent variables and maximum likelihood estimate of the model to produce an AIC_c information score: the lower the value, the better the model fits the data. The table also provides useful statistics on the number of parameters in the model (K), the difference in AIC score between the best model and the compared model (Δ AIC_c), and the weight (AIC_c Weight), which represents the proportion of the total amount of predictive power contained in the assessed model. For this analysis, we created twelve models in total to examine the individual and additive effects of the factors.

<u>Model 0</u>: $\boldsymbol{y} = \beta_0$
<u>Model 1</u>: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Age}$
<u>Model 2</u>: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Total\ Length}$
<u>Model 3</u>: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Auxiliary\ Girth}$
<u>Model 4</u>: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Skull\ Width}$
<u>Model A</u>: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Population}$

Model B: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Years\ Since}$
Model C: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Sex}$
Model 3C: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Auxiliary\ Girth} + \beta_2 \boldsymbol{Sex}$
Model AC: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Population} + \beta_2 \boldsymbol{Sex}$
Model 3AC: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Auxiliary\ Girth} + \beta_2 \boldsymbol{Population} + \beta_3 \boldsymbol{Sex}$
Model 13AC: $\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{Age} + \beta_2 \boldsymbol{Auxiliry\ Girth} + \beta_3 \boldsymbol{Population} + \beta_4 \boldsymbol{Sex}$

We fit and analyzed these models by producing a set of predictive coefficients, p-values to determine significance, and a coefficient of determination (R^2), which shows what percentage of data observed can be explained by the produced model. After completing this initial step for each model, we grouped all models together and produced an AIC table.

Results

Visually examining data in plots is a good way to understand how populations may be trending without having to analyze critical datasets. Because our sample size for this analysis is sufficiently large ($n > 1000$), it is possible to glean information from these plots. For Figure 1, we looked at the average mass of a polar bear captured by year, for years of the study where they're complete data records. Immediately, it is clear that there are three outlier years (1996, 2004, and 2005), which are a result of only a few captures in those given years being filtered for this analysis. When looking at the data, we can see that the average weight of a polar bear is fluctuating very slightly; however, the final five years of the plot show a consistently higher average weight than the first five years. In Figure 2, we use three boxplots to represent the three "epochs" of years for this study. All three plots show a very small area between Q1 and Q3 (interquartile range), which means there isn't a large amount of variation in the data for each epoch, an important factor when determining if the information is relevant and can be trusted. Lastly, we visually compared the three epochs, but with the added factor of grouping the results by polar bear sex (Figure 3). This allowed us to see if the change over the course of an epoch was similar for both sexes of polar bears, and if one had a greater influence than the other. This plot showed that the weights of male polar bears increased as the study went on but showed a larger amount of variability than female polar bear weights. Because of this, further testing would need to be completed to see if the change in male polar bear weights was statistically significant, or if the change could be explained by random variation in the sampling pattern.

The model selection of this analysis was conducted to better understand which morphological factor in polar bears contributed as the best predictor for body mass. Although twelve models were created (Tables 1-12), some results contributed much more significant results. Model 0 (Table 1), which represented a null model of no effect, shows us that the average weight of a polar bear capture during this study would be about 187 kilograms if there were no influences from other factors. In all models which fit for one factor (Tables 1-8), only one factor was proven to not be statistically significant based on its p-value if we were estimating to 95% confidence: Population. Reasons why this may be true will be discussed in the next section, but it should be noted that the results of that model showed that on average, polar bears from the Southern Beaufort Sea were about 22.2 kilograms heavier (Table 6). Based on coefficients of determinations, we can conclude that morphological factors rather than location or time-based factors play the greatest influence on predicting polar bear weights, as the three highest R^2

values were results of axillary girth, skull width, and total length, respectively. Axillary girth produced the highest value at 0.875, which shows that about 87.5% of the variation can be explained by the data. Because of this, we included axillary girth in most of the other models with two or more factors in an attempt to represent even more of the data (Table 9, 11, 12). In Model 13AC, the most complicated of the models fitted, we received the highest multiple R^2 value (Table 12); however, it should be noted that with a finite amount of observed data, every time we add another independent variable, we need to estimate another parameter. Because of this, a trade-off occurs where the uncertainty of our estimates will get bigger, but our bias, the difference between the model and truth, will go down.

Selecting for a model of best fit produced much more concrete results. Based on these results (Table 13), the model selected for best representation would be Model 13AC, which looked at the effect of age, axillary girth, population, and sex on total body mass of polar bears in this study. Compared to the other models, Model 13AC had the lowest AIC score (albeit still very, very high), and it carries 100% of the total amount of predictive power contained in the assessed model table, based on AIC weight. Despite researchers generally being ecstatic about receiving a model that produces 100% prediction power, there is some room for reservation, which will be discussed further. However, we can also see that the top four models produced through this system all had one factor in common: axillary girth. This result, paired with its largest coefficient of determination, leads us to believe that axillary girth, which is measured as the circumference of the thorax immediately behind the forelimbs (Thiemann et al., 2011), is the most important factor.

Discussion

For this analysis, we set out to use this dataset to examine two main questions: are polar bear weights trending downward and which morphological factor contributes the most towards predicting polar bear body mass. After visually examining the produced plots, especially Figure 2, we can see that polar bear weights may be trending upwards, or at the very least, not significantly lower than when the study started. Some cause for concern and further research may be needed to isolate the trend of female weights, as they did not show the same gradual rise that male weights did (Figure 3). Female weights are often more important for populations as they can be useful indicators of population responses to environmental change, linking nutritional intake to effects on reproduction and survival (Sergio et al., 2005).

As stated early, understanding the relationship between morphological factors to estimate body mass are a significant means for monitoring trends in body condition over time. This analysis shows us that axillary girth played the largest role in predicting body weight, which is important to understand for monitoring and research purposes. Firstly, axillary weight is easy to measure and highly replicable among individual researchers, without the need for specialized equipment (Thiemann et al., 2011). Secondly, we can use this information to inform future research about large carnivores that play essential roles in ecosystems. We can also monitor for threats to this factor, whether it be loss of dietary consumption due to human activity or sea ice loss leading to higher periods of stress and movement.

References

Atwood, T., Durner, G., Rode, K., & Simac, K. (n.d.). Polar bear research. United States Geological Survey. Retrieved April 1, 2021, from https://www.usgs.gov/centers/asc/science/polar-bear-research

Gittleman, J. L. (1986). Carnivore brain size, behavioral ecology, and phylogeny. Journal of Mammalogy, 67(1), 23–36. https://doi.org/10.2307/1380998

Rode, K. D., Atwood, T. C., Thiemann, G. W., St. Martin, M., Wilson, R. R., Durner, G. M., Regehr, E. V., Talbot, S. L., Sage, G. K., Pagano, A. M., & Simac, K. S. (2020). Identifying reliable indicators of fitness in polar bears. PLOS ONE, 15(8), e0237444. https://doi.org/10.1371/journal.pone.0237444

Rode, K. D. (2020). Measurement data of Polar Bears captured in the Chukchi and southern Beaufort Seas, 1981-2017. U.S. Geological Survey data release. https://doi.org/10.5066/P9TVK3PX.

Sergio, F., Newton, I., & Marchesi, L. (2005). Top predators and biodiversity. Nature, 436(7048), 192. https://doi.org/10.1038/436192a

Silliman, B. R. & Angelini, C. (2012) Trophic Cascades Across Diverse Plant Ecosystems. Nature Education Knowledge 3(10):44

Thiemann, G. W., Lunn, N. J., Richardson, E. S., & Andriashek, D. S. (2011). Temporal change in the morphometry–body mass relationship of polar bears. The Journal of Wildlife Management, 75(3), 580–587. https://doi.org/10.1002/jwmg.112

World Wildlife Foundation. (n.d.). Threats to polar bears. WWF Arctic. Retrieved April 30, 2021, from https://arcticwwf.org/species/polar-bear/threats/

Tables

Linear Regression Model 0 Null Model of No Effect				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	187.17	3.28	57.1	< 2e-16
Residual Standard Error		115 on 1226 degrees of freedom		

Table 1

Linear Regression Model 1 (Independent variable = Age // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	114.953	3.832	30	< 2e-16
Age (β_1)	11.333	0.437	26	< 2e-16
Residual Standard Error		92 on 1225 degrees of freedom		
Multiple R^2		0.355 (≈35.5% of the variation can be explained)		
F-statistic		674 on 1 and 1225 degrees of freedom		
Model		y = 114.953 + 11.333 * (β_1)		

Table 2

Linear Regression Model 2 (Independent variable = Total Length // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-213.7201	6.9273	-30.9	< 2e-16
Total Length (β_1)	1.9965	0.0335	59.6	< 2e-16
Residual Standard Error		58 on 1225 degrees of freedom		
Multiple R^2		0.744 (≈74.4% of the variation can be explained)		
F-statistic		3550 on 1 and 1225 degrees of freedom		
Model		y = -213.7201 + 1.9965 * (β_1)		

Table 3

Linear Regression Model 3 (Independent variable = Axillary Girth // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-193.5289	4.2742	-45.3	< 2e-16
Axil Girth (β_1)	3.4502	0.0373	92.5	< 2e-16
Residual Standard Error		41 on 1225 degrees of freedom		
Multiple R^2		0.875 (≈87.5% of the variation can be explained)		
F-statistic		8560 on 1 and 1225 degrees of freedom		
Model		y = -193.5289 + 3.4502 * (β_1)		

Table 4

Linear Regression Model 4 (Independent variable = Skull Width // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-264.980	5.481	-48.3	< 2e-16
Skull Width (β_1)	24.011	0.283	84.7	< 2e-16
Residual Standard Error		44 on 1225 degrees of freedom		
Multiple R^2		0.854 (≈85.4% of the variation can be explained)		
F-statistic		7180 on 1 and 1225 degrees of freedom		
Model		y = -264.98 + 24.011 * (β_1)		

Table 5

Linear Regression Model A (Independent variable = Population // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	166.9	11.1	15.0	< 2e-16
Population [SB] (β_1)	22.2	11.7	1.9	0.057
Residual Standard Error		115 on 1225 degrees of freedom		
Multiple R^2		0.00295 (≈0.3% of the variation can be explained)		
F-statistic		3.63 on 1 and 1225 degrees of freedom		

Table 6

Linear Regression Model B (Independent variable = Years Since // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	166.058	6.961	23.85	< 2e-16
Years Since (β_1)	1.217	0.354	3.43	0.00062
Residual Standard Error		114 on 1225 degrees of freedom		
Multiple R^2		0.00953 (≈0.95% of the variation can be explained)		
F-statistic		11.8 on 1 and 1225 degrees of freedom		
Model		y = 166.058 + 1.217 * (β_1)		

Table 7

Linear Regression Model C (Independent variable = Sex // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	151.4	4.03	37.6	< 2e-16
Sex [M] (β_1)	84.39	6.19	13.6	< 2e-16
Residual Standard Error		107 on 1225 degrees of freedom		
Multiple R^2		0.132 (≈13.2% of the variation can be explained)		
F-statistic		186 on 1 and 1225 degrees of freedom		

Table 8

Linear Regression Model 3C (Independent variable = Axillary Girth + Sex // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-194.1509	3.8717	-50.1	< 2e-16
Axil Girth (β_1)	3.318	0.0347	95.6	< 2e-16
Sex [M] (β_2)	35.8765	2.1871	16.4	< 2e-16
Multiple R^2		0.897 (≈89.7% of the variation can be explained)		
Model		y = -194.1509 + 3.318 * (β_1) + 35.8765 * (β_2)		

Table 9

Linear Regression Model AC (Independent variable = Population + Sex // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	135.15	10.65	12.69	< 2e-16
Population [SB] (β_1)	17.93	10.88	1.65	0.1
Sex [M] (β_2)	84.09	6.18	13.6	< 2e-16
Multiple R^2		0.134 (≈13.4% of the variation can be explained)		
Model		y = 135.15 + 17.93 * (β_1) + 84.09 * (β_2)		

Table 10

Linear Regression Model 3AC (Independent variable = Axillary Girth + Pop + Sex // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-214.1712	5.1046	-41.96	< 2e-16
Axil Girth (β_1)	3.3202	0.0342	96.95	< 2e-16
Population [SB] (β_2)	21.8284	3.6929	5.91	4.4e-09
Sex [M] (β_3)	35.4856	2.1584	16.44	< 2e-16
Multiple R^2		0.900 (≈90% of the variation can be explained)		
Model		y = -214.1712 + 3.3202 * (β_1) + 21.83 * (β_2) + 35.4856 * (β_3)		

Table 11

Linear Regression Model 13AC (Independent variable = Age + Axil Girth + Pop + Sex // Dependent variable = Body Mass)				
Coefficients	Estimate	Standard Error	t-value	Pr (>|t|)
Intercept (β_0)	-198.8655	5.1426	-38.67	< 2e-16
Age (β_1)	2.1633	0.2163	10.00	< 2e-16
Axil Girth (β_2)	3.0447	0.0429	70.91	< 2e-16
Population [SB] (β_3)	20.3298	3.5550	5.72	1.3e-08
Sex [M] (β_4)	41.8072	2.1700	19.27	< 2e-16

Multiple R^2	0.908 (≈90.8% of the variation can be explained)

Table 12

AIC$_c$ Table for Model Selection					
Model Name	K	AIC$_c$	Δ AIC$_c$	AIC$_c$ Weight	Cumulative Weight
Age + Axil Girth + Population + Sex (Model 13AC)	6	12210	0	1.00	1.00
Axil Girth + Population + Sex (Model 3AC)	5	12304	95	0.00	1.00
Axil Girth + Sex (Model 3C)	4	12337	127	0.00	1.00
Axillary Girth (Model 3)	3	12578	369	0.00	1.00
Skull Width (Model 4)	3	12766	556	0.00	1.00
Total Length (Model 2)	3	13458	1249	0.00	1.00
Age (Model 1)	3	14591	2381	0.00	1.00
Population + Sex (Model AC)	4	14954	2745	0.00	1.00
Sex (Model C)	3	14955	2746	0.00	1.00
Years Since (Model B)	3	15117	2907	0.00	1.00
Population (Model A)	3	15125	2915	0.00	1.00
Null Model (Model 0)	2	15127	2917	0.00	1.00

Table 13

Figures

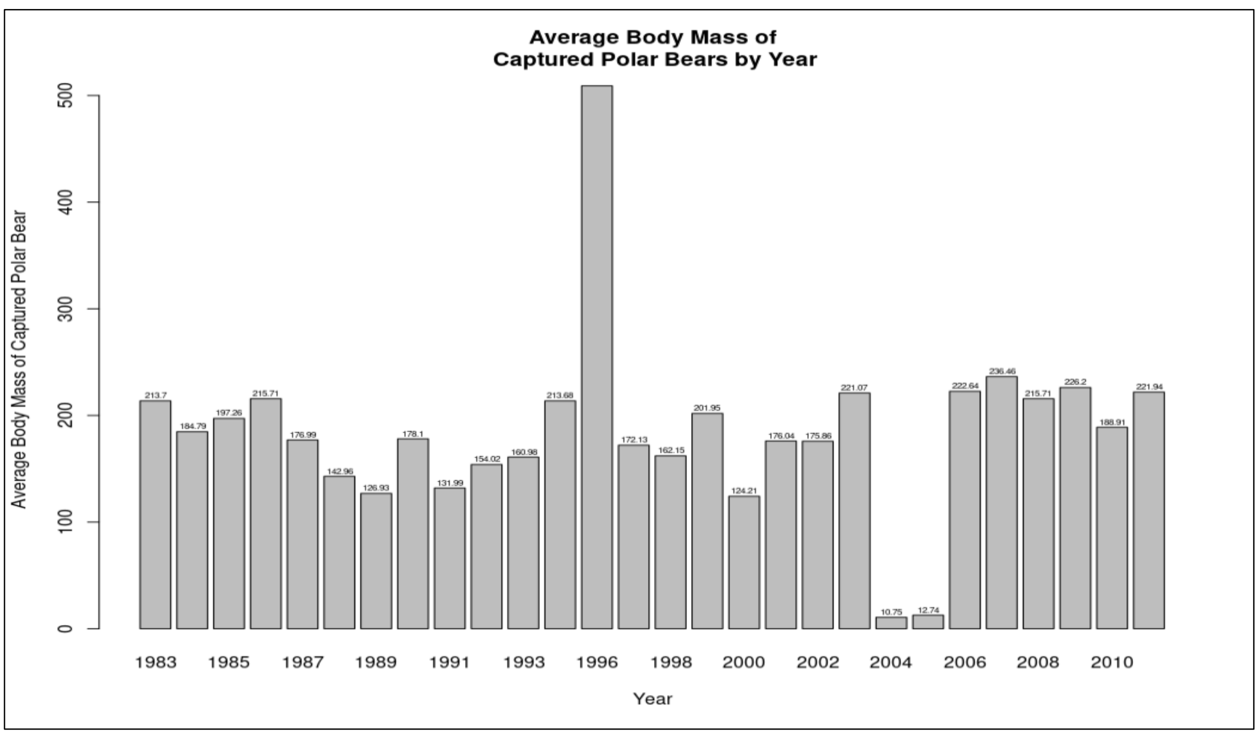

Figure 1

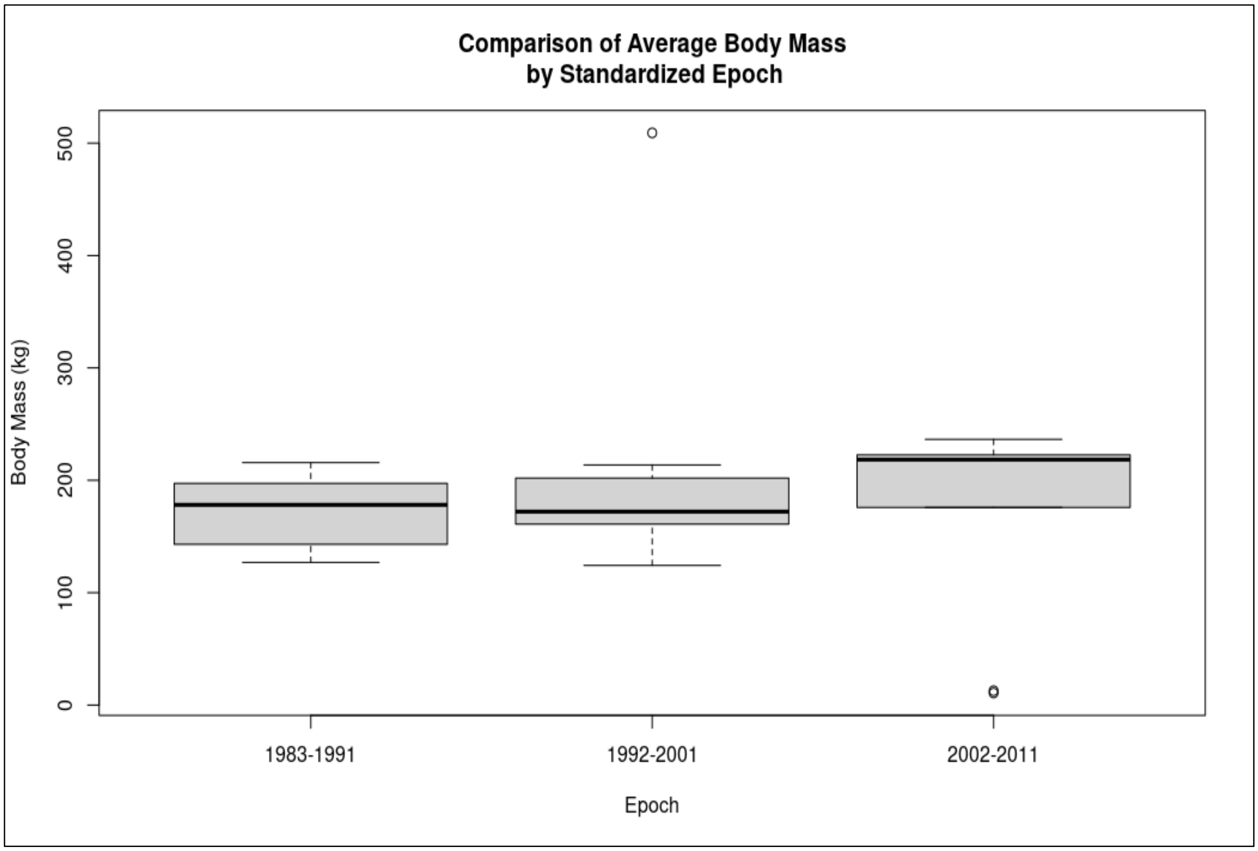

Figure 2

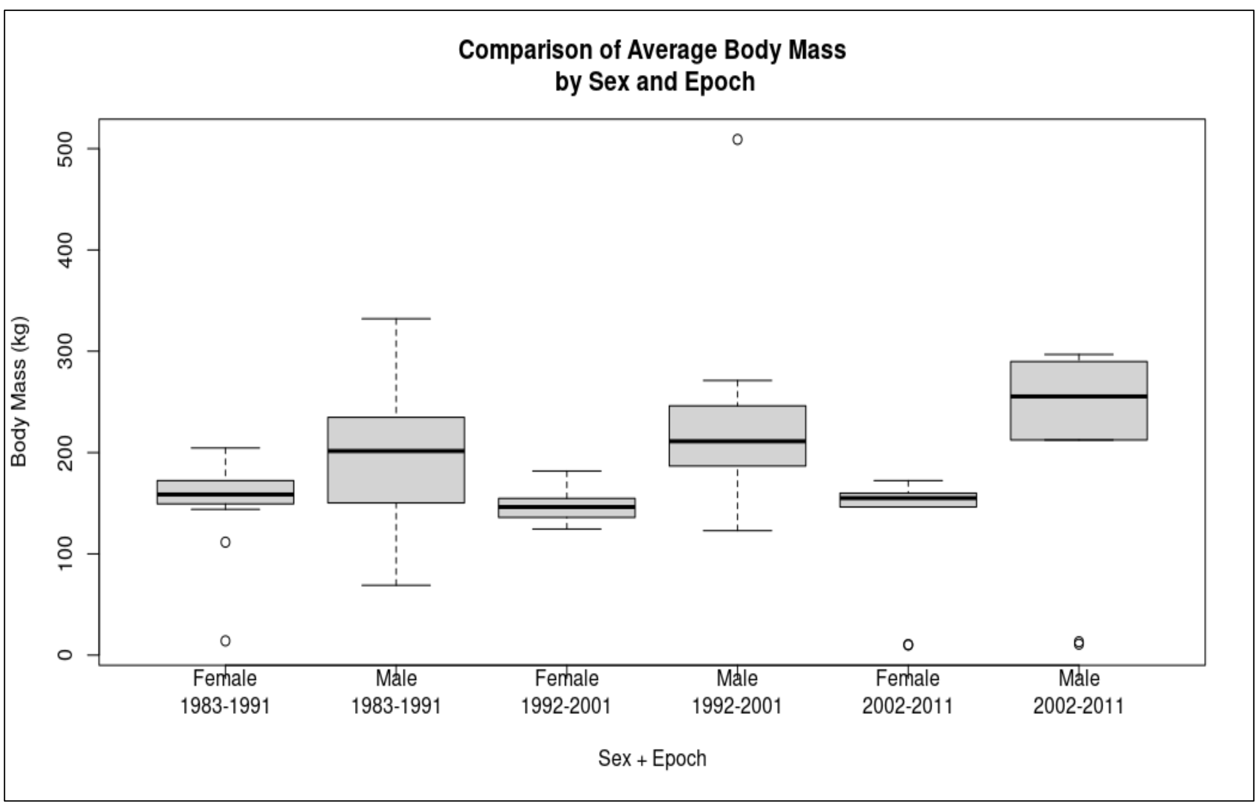

Figure 3