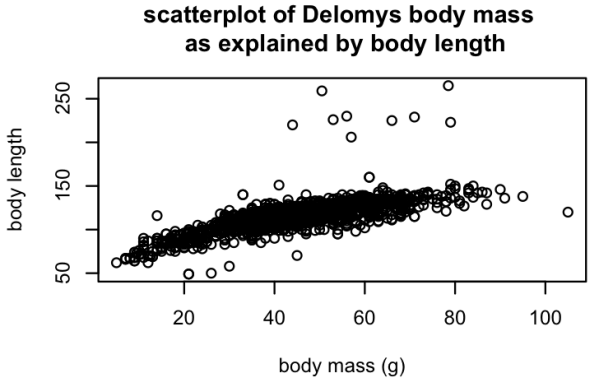
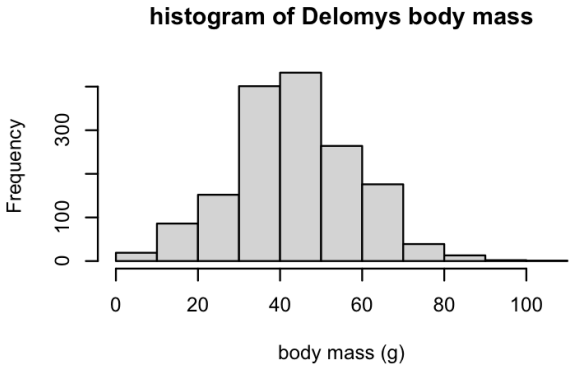
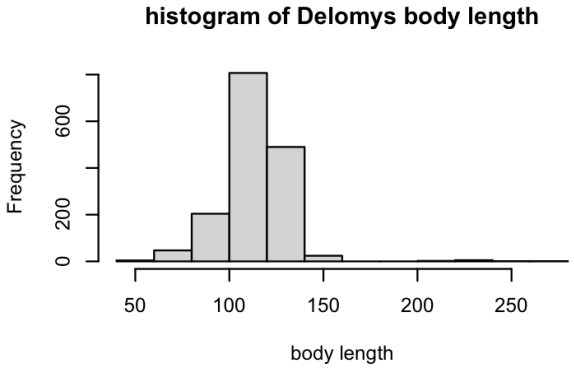
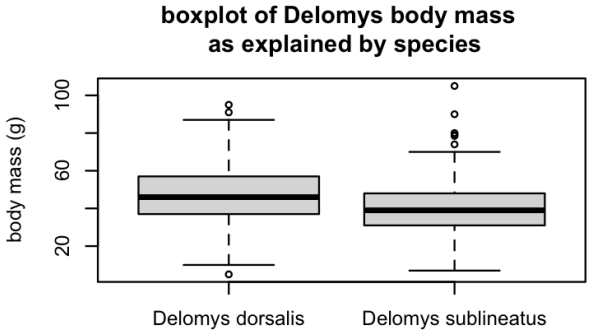
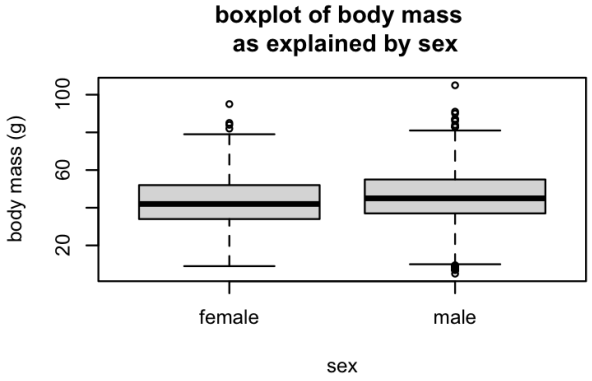
Matt Fertakos
Final Assignment Part 2
Worked with: Bonnie, John, and Mandy

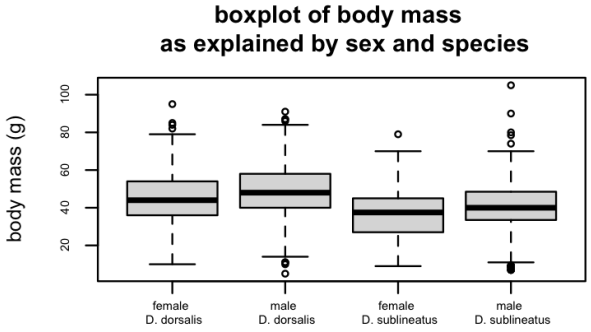












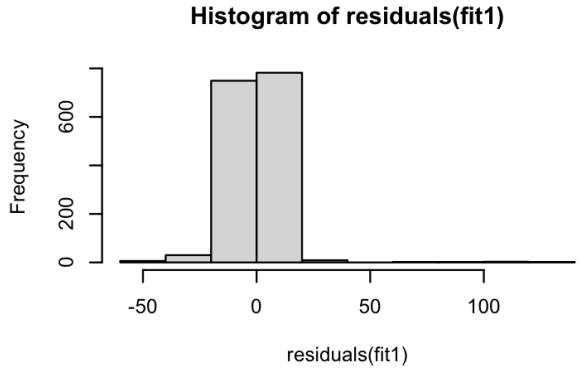
Question 1:
The relationship between body mass and length appears linear. In the histogram, it appears as body mass increases body length also increases. There are a few outliers though.

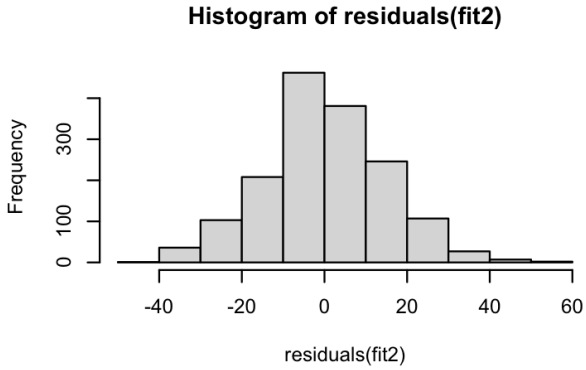
Question 2:
The histogram of body mass appears normally distributed, and the histogram of body length does as well although it is much narrower in its range and lacks data to the right of what appears to be the mean.

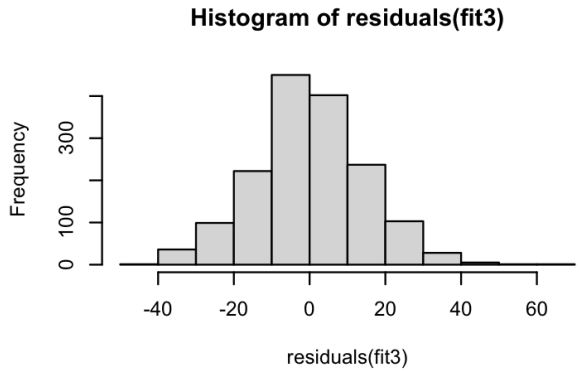
Question 3:
It turns out they are not normally distributed based on the results of a Shapiro test being significant. The null hypothesis of a Shapiro test is that they are normally distributed, so our p value less than 0.05 rejects this null meaning they are not normally distributed. This emphasizes the importance of statistical tests because to me both histograms appeared normal or nearly normal, when in fact neither are.

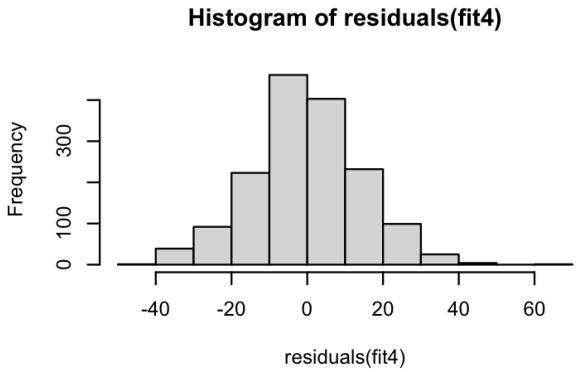
p-value of body length: < 2.2e-16
p-value of body mass: 4.33e-05

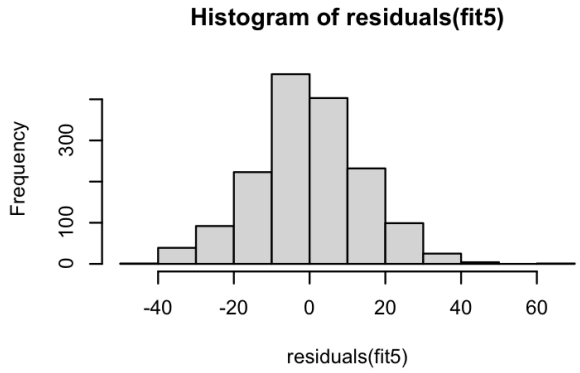
Question 4:
Based on the conditional boxplots, I do not see any graphical evidence for body mass differences based on species and/or sex. In all three boxplots the individual plots overlap with each other.











Question 5:

The histograms of the residuals for each model appear normally distributed. Results of the Shapiro test of each model do not confirm this though, as all of them have p values less than 0.05.

Question 6:

No, these violations are not as equally severe for all the models. For example, the p value for fit2 is larger (not as strong evidence for normality ; 0.000154) than the others (~2.2e-16, 0.00008, 0.00007, 0.00007). This means that there is a greater chance to get a false positive where we accept non-normality and it is actually normal in fit2. This corresponds to it being closer to normal (and therefore a less severe violation) than the other models fit.

	Estimate	Std. Error	t value	Pr(>	t	)
(Intercept)	76.1246565	0.9158120	83.12258	0
body_mass	0.8754988	0.0196905	44.46298	0

Question 7:
The magnitude of the body mass/length relationship is 0.875. For every 1 unit increase in body mass (g) you add 0.875 length units.

Question 8:
The expected body length of an animal that weighs 100g is 76.12 + (0.875*100) = 163.6 length units.

Question 9:
The expected body length of an animal that weighs 0g is 76.12+ (0.875*0)= 76.12 length units which in reality wouldn't exist because there are no animals that weigh 0g.

Question 10:
The base level for sex is female.

Question 11:
The base level for binomial is *Delomys dorsalis*.

Question 12:
Male is heavier.

Question 13:
Delomys dorsalis is heavier.

Question 14:
Yes they are both significant based on the p values of fit2 and fit3 being less than 0.05.

Question 15:
There is not a significant interaction because the ANOVA table of fit5 (body_mass ~ sex * binomial) show of p value of 0.95.

Question 16:

From the single predictor model to the additive model the p value (significance) of sex decreases (from p=0.000195 to p=0.000114), while the p value (significance) of species stays the same ($p < 2.2e-16$). The significance level of sex increases slightly in the interactive model (p=0.000114 to p=0.000115) and the significance of species stays the same from the additive to the interactive model ($p < 2.2e-16$). The differences between the p values of sex are not THAT

different from one another though, so this might not be a necessary observation. The p value does not decrease with additional predictors.

Question 17:
fit4 has the lowest AIC value (12896.71) and fit5 has the second lowest AIC value (12898.72).

Question 18:
The AIC values are also almost the same between fit4 and fit5 meaning they both fit the data well. Either could be logically chosen. **BUT,** because the p value of the ANOVA test of the interactive model (fit5) was 0.95 there is not a significant interaction between sex and body mass, so I would pick the **additive model (fit4)**. The interactive component did not add any significant interactions based on the ANOVA p value, which prompts us to minimize the complexity using the additive model as opposed to the interactive model. There is a tradeoff between model complexity and model fit in this matter.