Ian Eggleston

Worked with Sarah Guitart

1. The null hypothesis for the Chi squared test states that there is no preference for interior or edges of habitats by brown creepers, based on their presence/absence.
2. The brown creepers seem to show a preference for the interior. This is based on a chisq test of the residuals that showed a positive number for more observations seen in interior habitats.
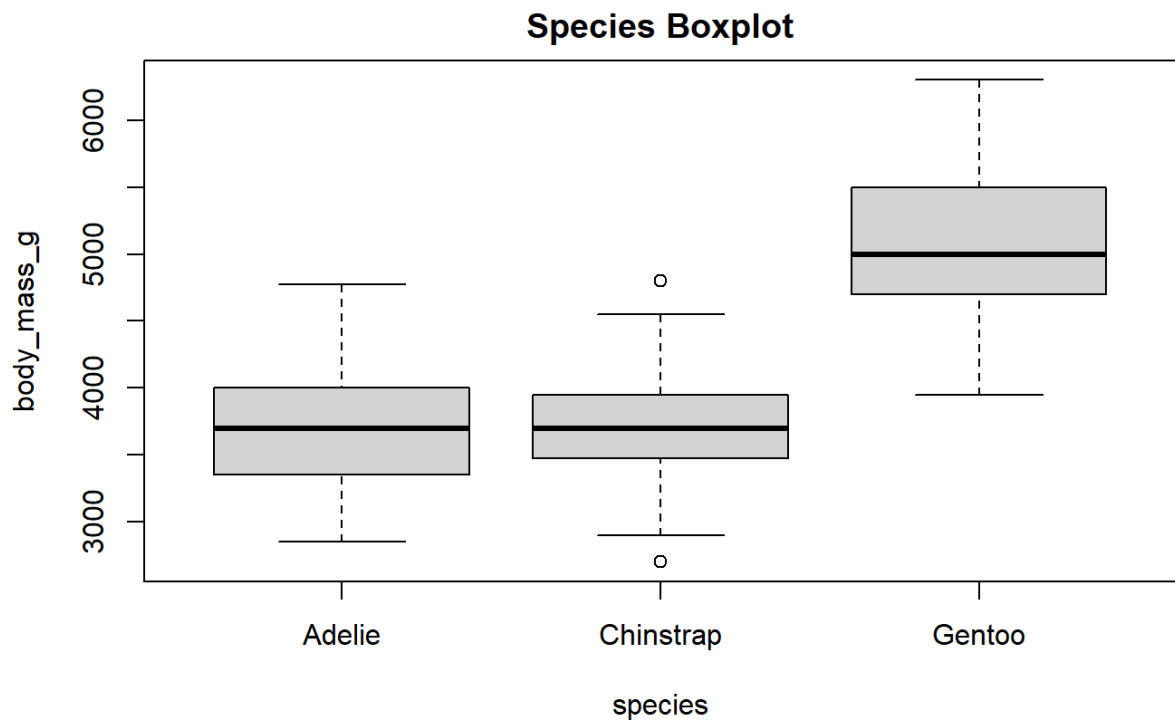3.

```
fit_species =
  lm(
    formula = body_mass_g ~ species,
    data = penguins)
```
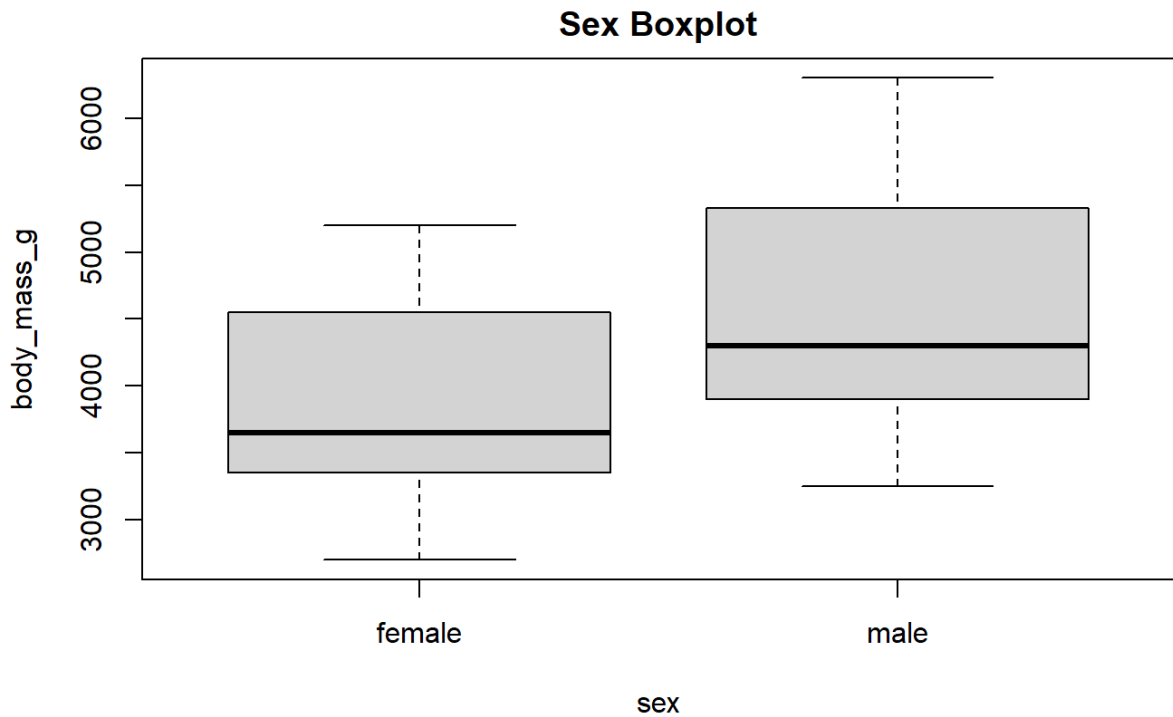
4.

```
fit_sex =
  lm(
    formula = body_mass_g ~ sex,
    data = penguins)
```
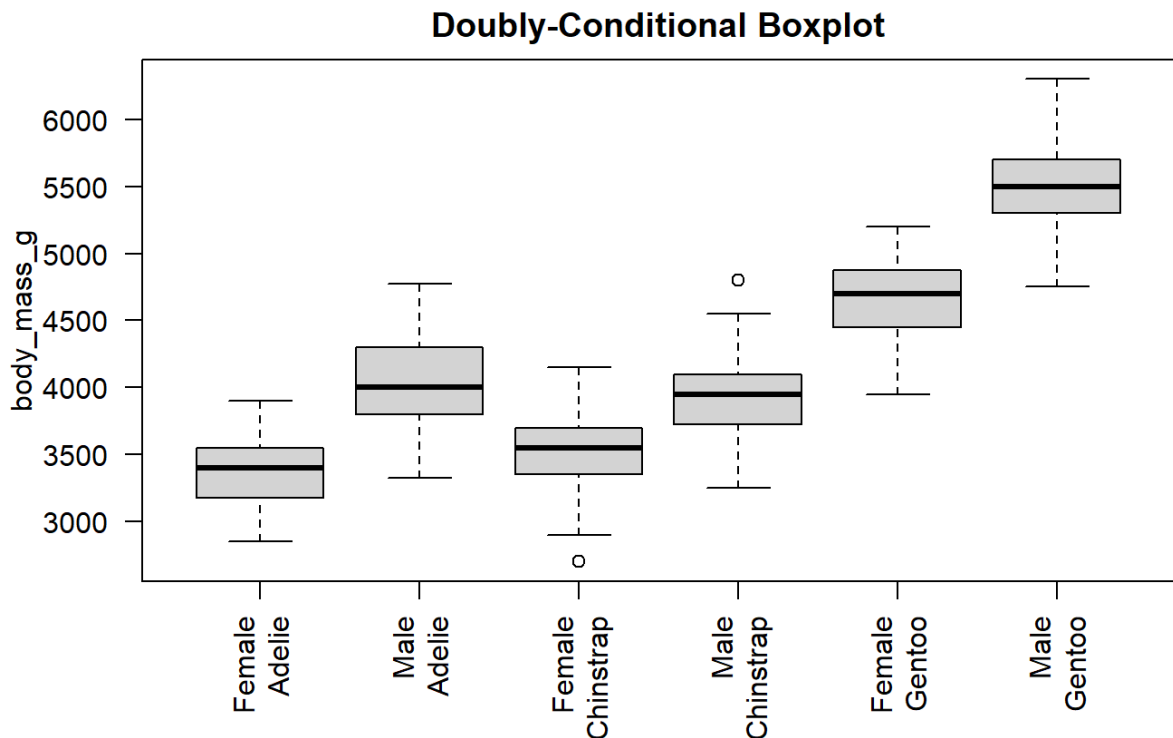
5.

```
fit_both =
  lm(
    formula = body_mass_g ~ (species * sex),
    data = penguins)
```

Species Boxplot

6.

**Sex Boxplot**

7.

Doubly-Conditional Boxplot

8.
9. From the above boxplots, we can see the model that grouped based on species has widely varying boxes. This suggests variance is not constant among the samples. This also occurs with the doubly conditional boxplot because the box sizes are all variable, without a constant size which suggests constant variance and homogeneity among samples.

10. The null hypothesis of the Bartlett test is that variance is constant across samples tested. It tests the homogeneity of our data across different categories to make sure variance is constant.
11. The p-value for species is 0.0501
12. The p-value for sex is 0.0319
13. The p-value for the combined Bartlett test is 0.1741
14. Based on the Bartlett tests, we can reject the null hypothesis for the model grouped by sex. This is because the p-value is less than 0.05, which allows us to reject the null hypothesis. For the Bartlett test, this then means the data does not have constant variance and there is a heterogeneity issue. This contrasts with the visual inspection of the boxplots, which suggested sex had the best grouping with constant variance. The other two values are above 0.05, even if only by a little, which means we cannot reject the null hypothesis.
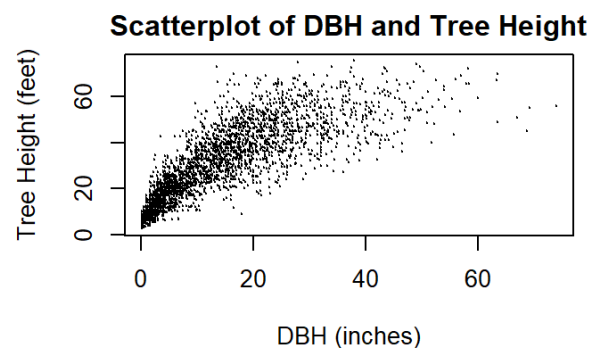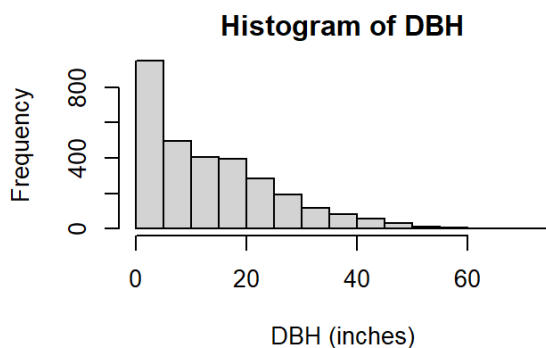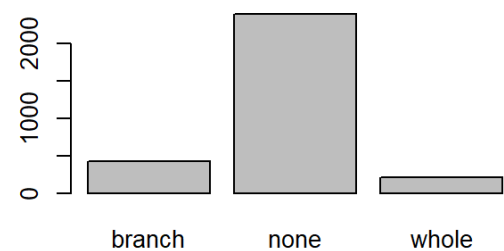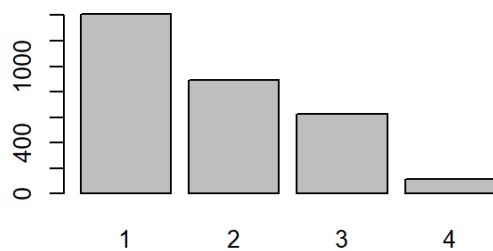
15.
16. The null hypothesis for the Kolmogorov-Smirnov test, tests if the objects come from the same distribution. In this case, that means does DBH for whole or broken branches come from the same distribution.
17. The p-value is 0.03254, which allows us to reject the null hypothesis that the distribution is not the same for these two groups.
18. The shape of this curve shows a logarithmic and monotonic relationship, where the data is constantly increasing but is not linear.
19. The Spearman correlation coefficient is most appropriate here.
20. The p-value is 2.2e-16, therefore the two variables are significantly correlated.
21. The X-squared value is 202.65 and p-value is 2.2e-16.
22. For category 1, the residual was -136.
23. There were fewer tree failures than expected in category 1.
24. There were more failures than expected by chance in category 4.

25. The probability failure rating system seems to work well because you can see that any probability rating higher than 1, had a greater chance of failing. This could mean category 1 is the safest and least likely to fail while the others show significant damage to the tree branches causing failures.