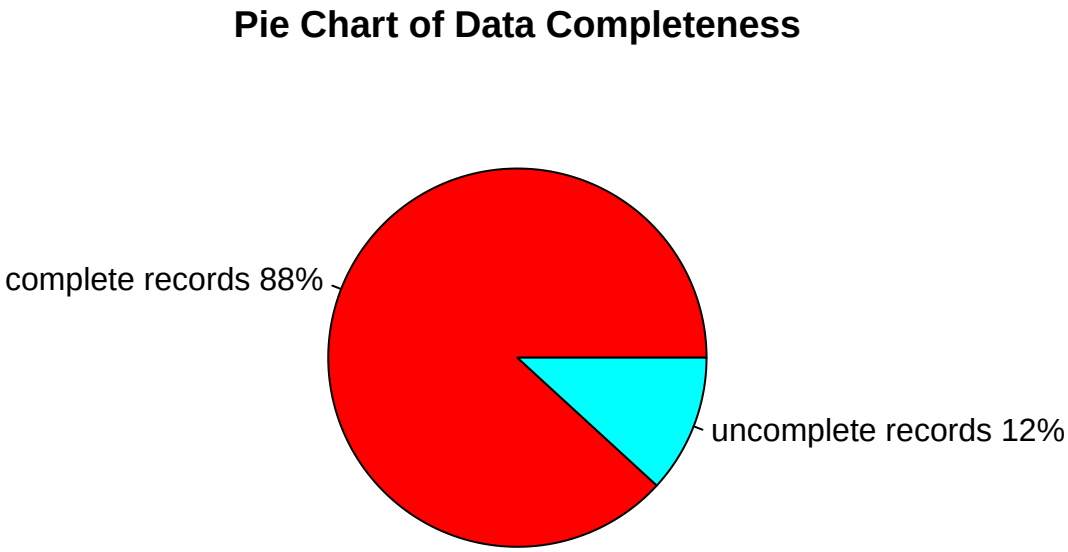
The Weight and Hindfoot Length Relationships of Rodents in Southern Arizona

Deky and Tono

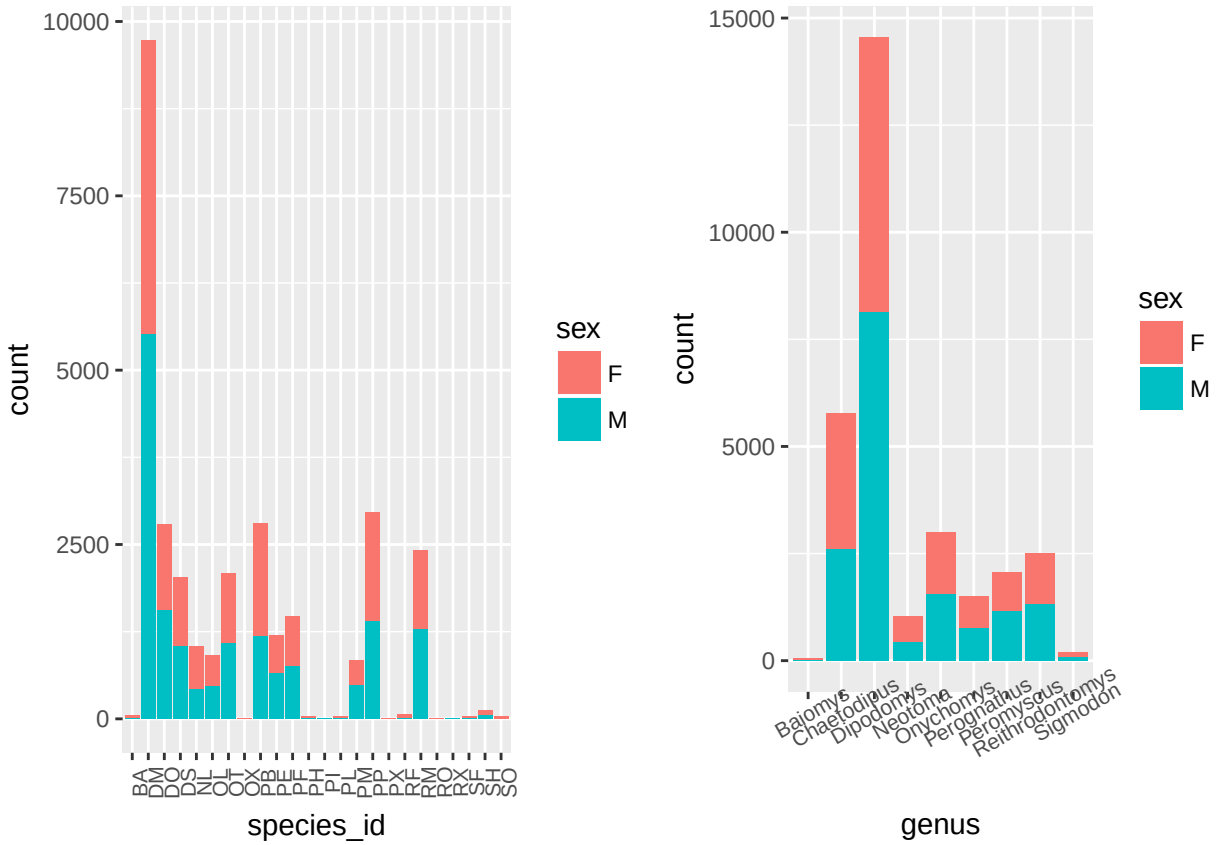
March 31, 2018

Summary of the Data

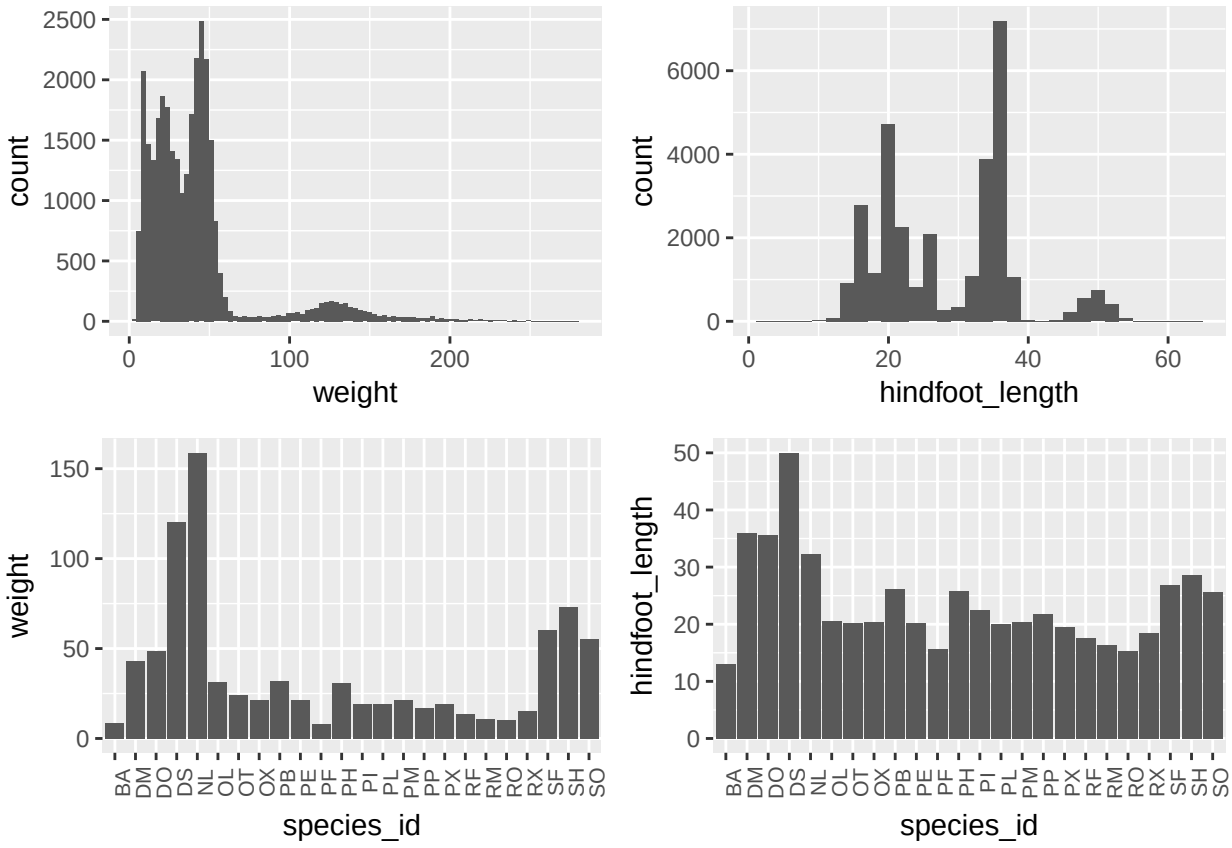
Data used in this assignment was derived from the time series data which were collected from 1977 to 2002 with the total numbers of recorded data is 34,786. However, we found that some of the records are incomplete which could lead to bias in our analysis. Therefore, we excluded the incomplete data in our further analysis. The comparison of the complete and incomplete data entries can be seen in Figure below.



The figure below delineates the general patterns of the data based on the number of samples for each species and genus, also the proportion of sex (female and male) for each category.



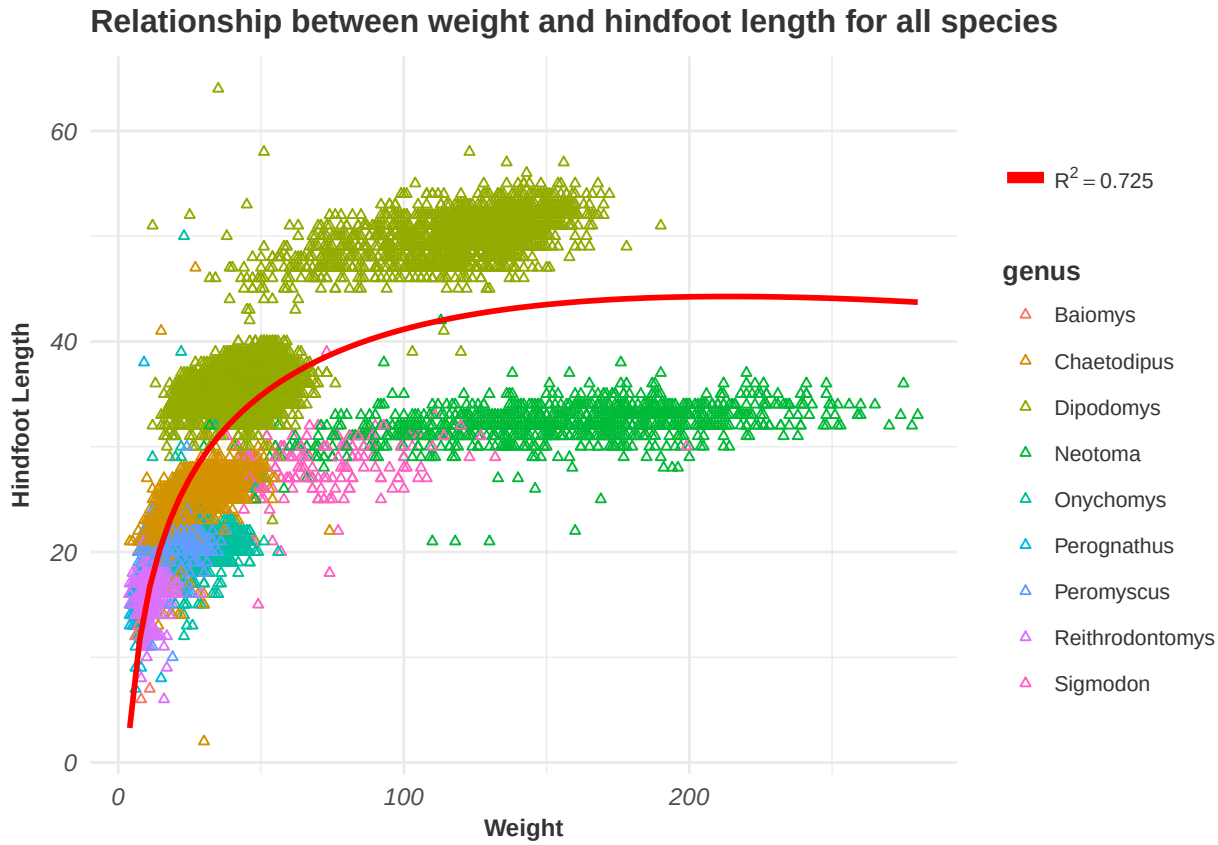
In addition to the general pattern of the data in the above, we also plotted the distribution of weight and hindfoot length for the overall sample and the distribution of the average (mean) weight and hindfoot length for each species, as follows:



The Weight and Hindfoot Length Relationship Analysis

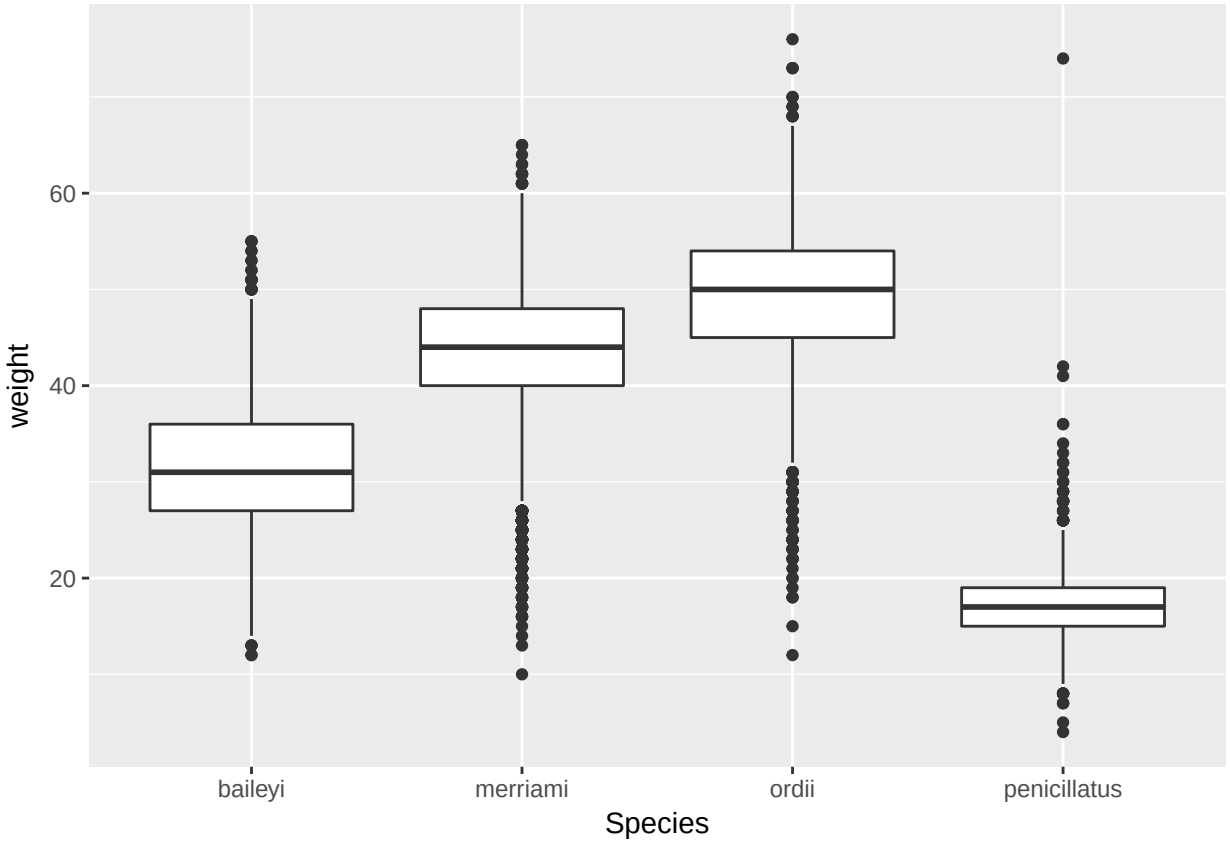
All species

The general pattern of the weight and hindfoot length relationship of all species combined follows the bounded exponential trendline. Based on the R^2 result, roughly around 72.5% of the population are matched with this trendline.

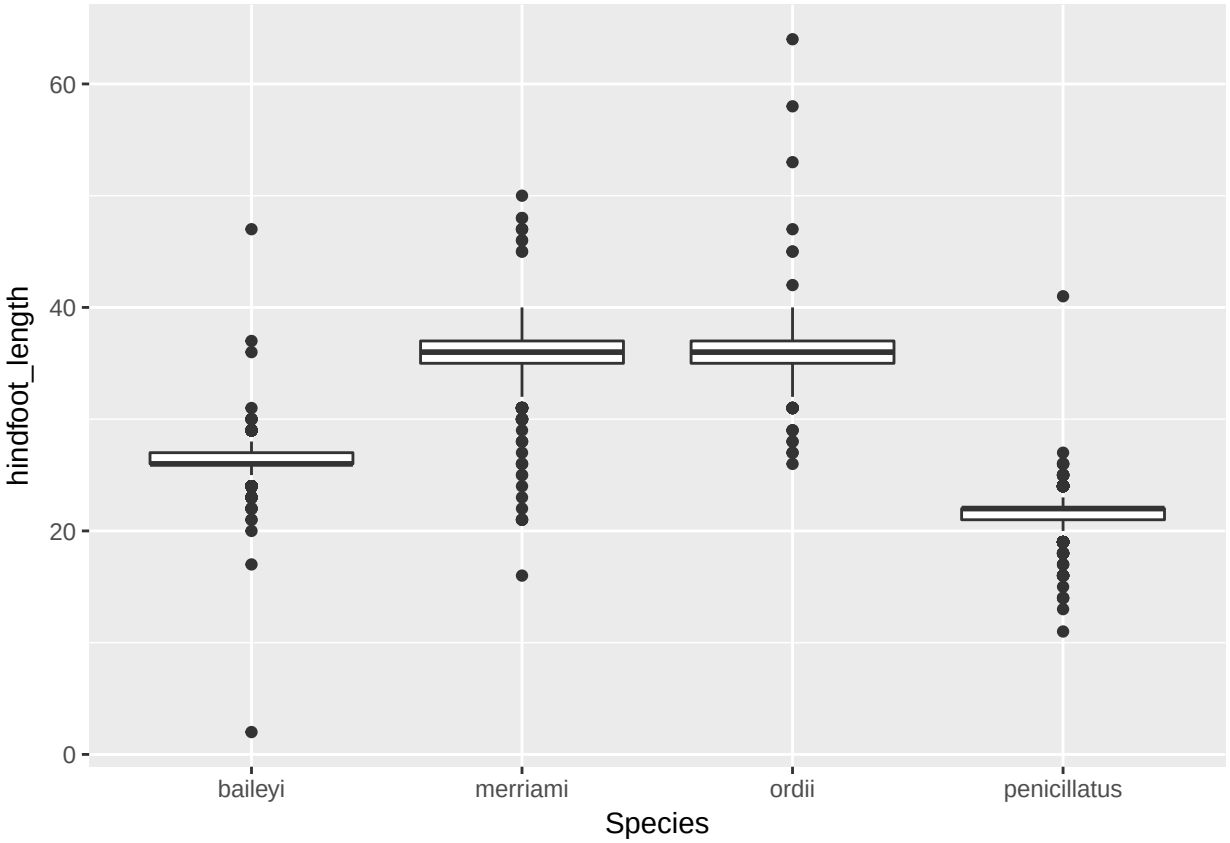


The most recorded species

From the distribution of the number of samples per species in the above, it shows that only four species that have recoded sample above 2500 entries from the period of 1977 to 2002. These species included, baileyi, merriami, ordii and penicillatus. Figures below are the summary of weight and length of hindfoot of the four species.

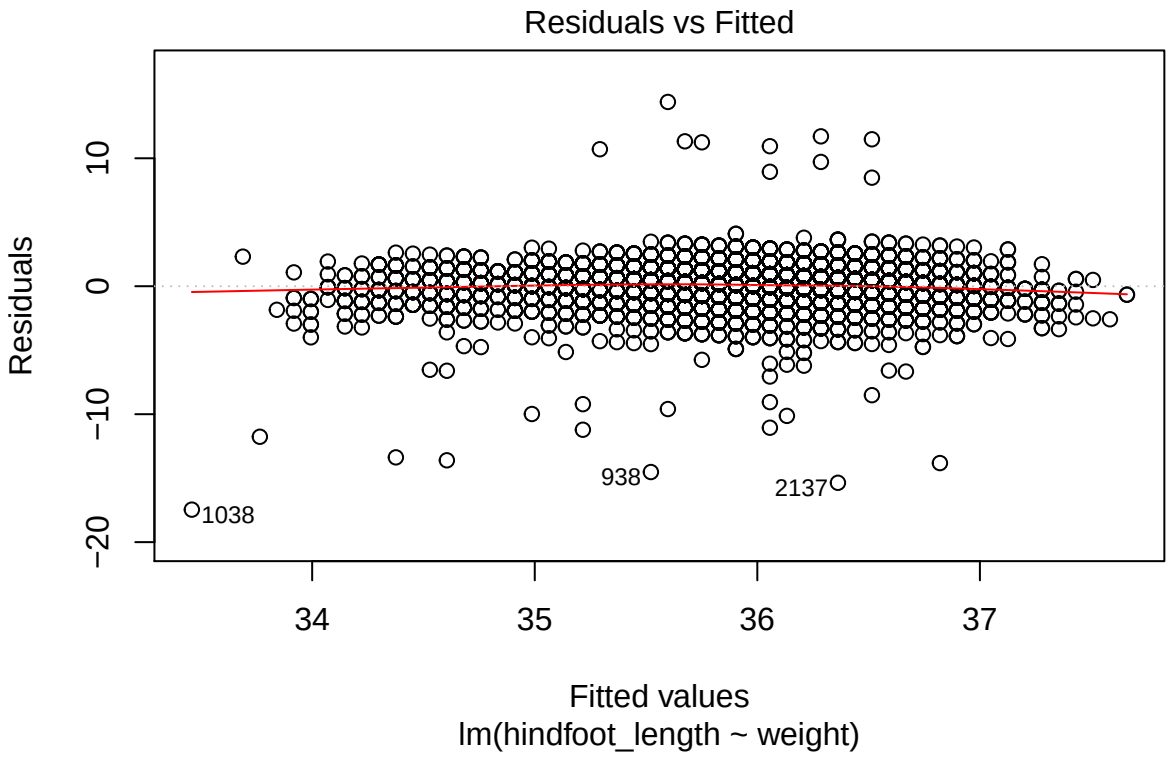


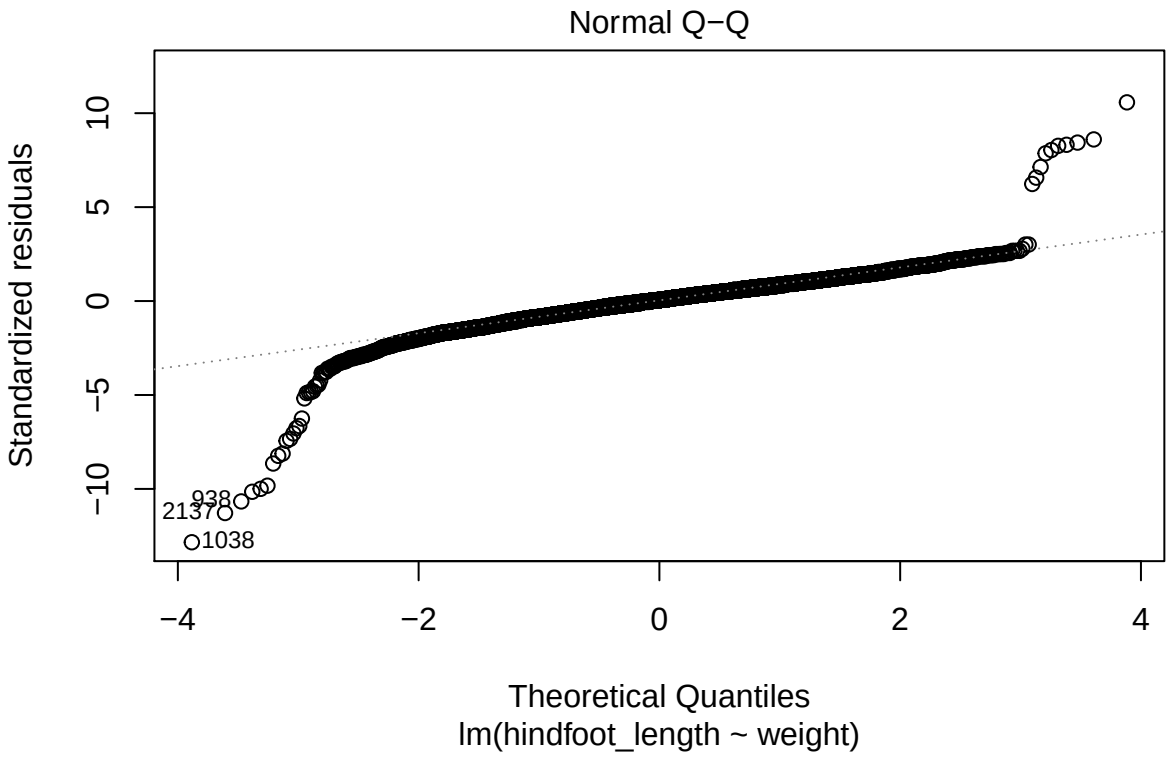
From the results, the ordii species has the highest mean weight among the other, whereas, the merriami species has the highest hindfoot length followed by ordii species in the second place.

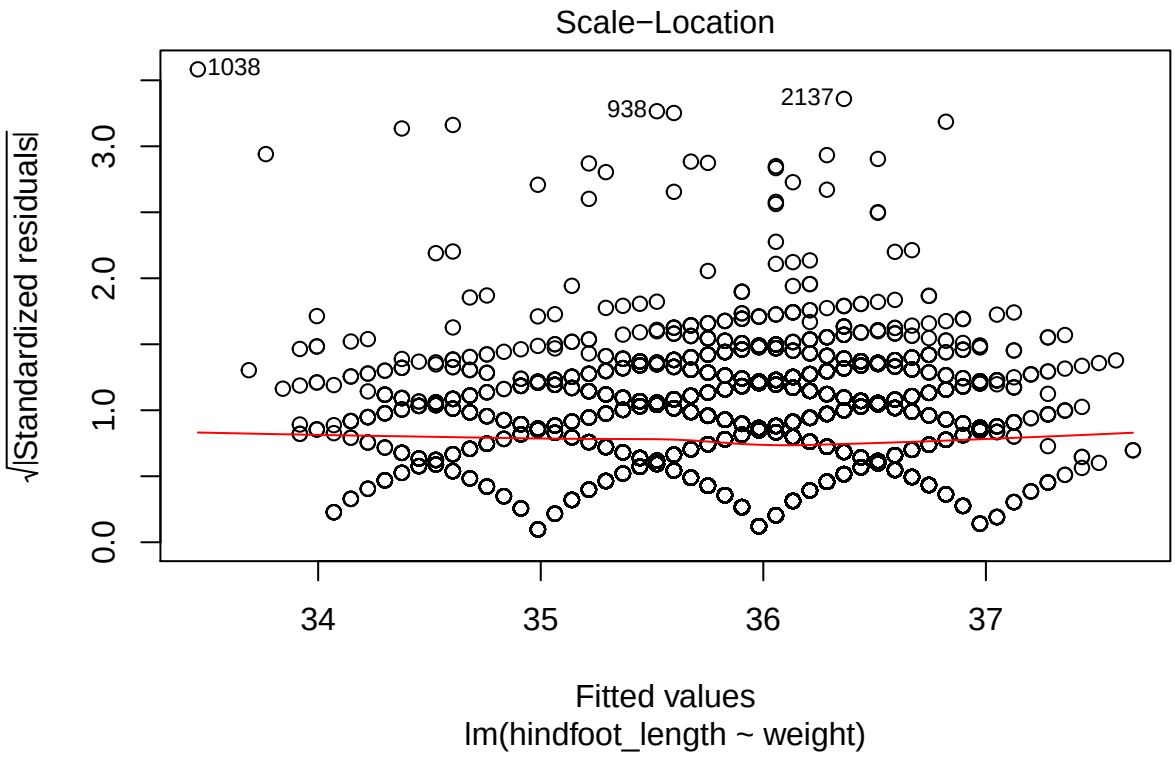


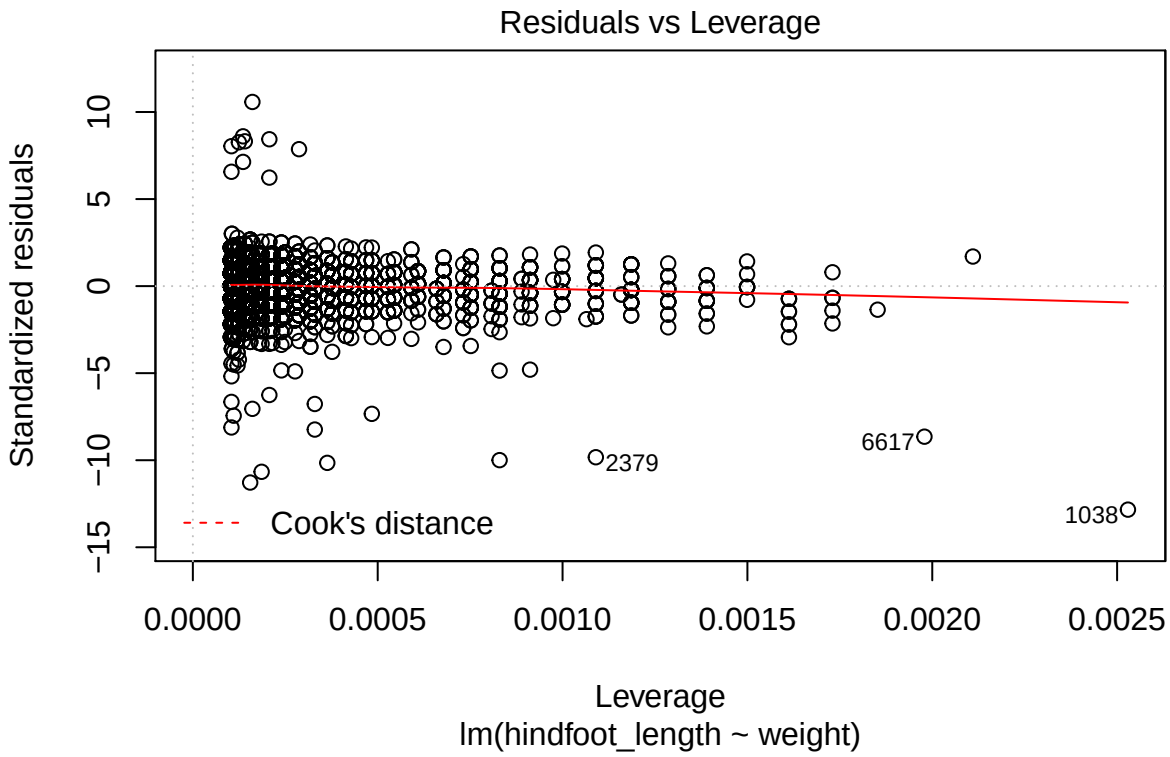
The linear model of weight and hindfoot length of the merriami species

In order to analyze the weight and hindfoot length relationship thoroughly, we analyze the data from the species which has the most significant sample number which is *merriami* species. The analysis was done using the linear model, as follows:


Residuals vs Fitted
Residuals
10
0
−10
−20
34
35
36
37
1038
938
2137
Fitted values
lm(hindfoot_length ~ weight)


Normal Q−Q
Standardized residuals
10
5
0
−5
−10
938
2137
1038
−4
−2
0
2
4
Theoretical Quantiles
lm(hindfoot_length ~ weight)


Scale−Location
1038
938
2137
√|Standardized residuals|
0.0
1.0
2.0
3.0
34
35
36
37
Fitted values
lm(hindfoot_length ~ weight)



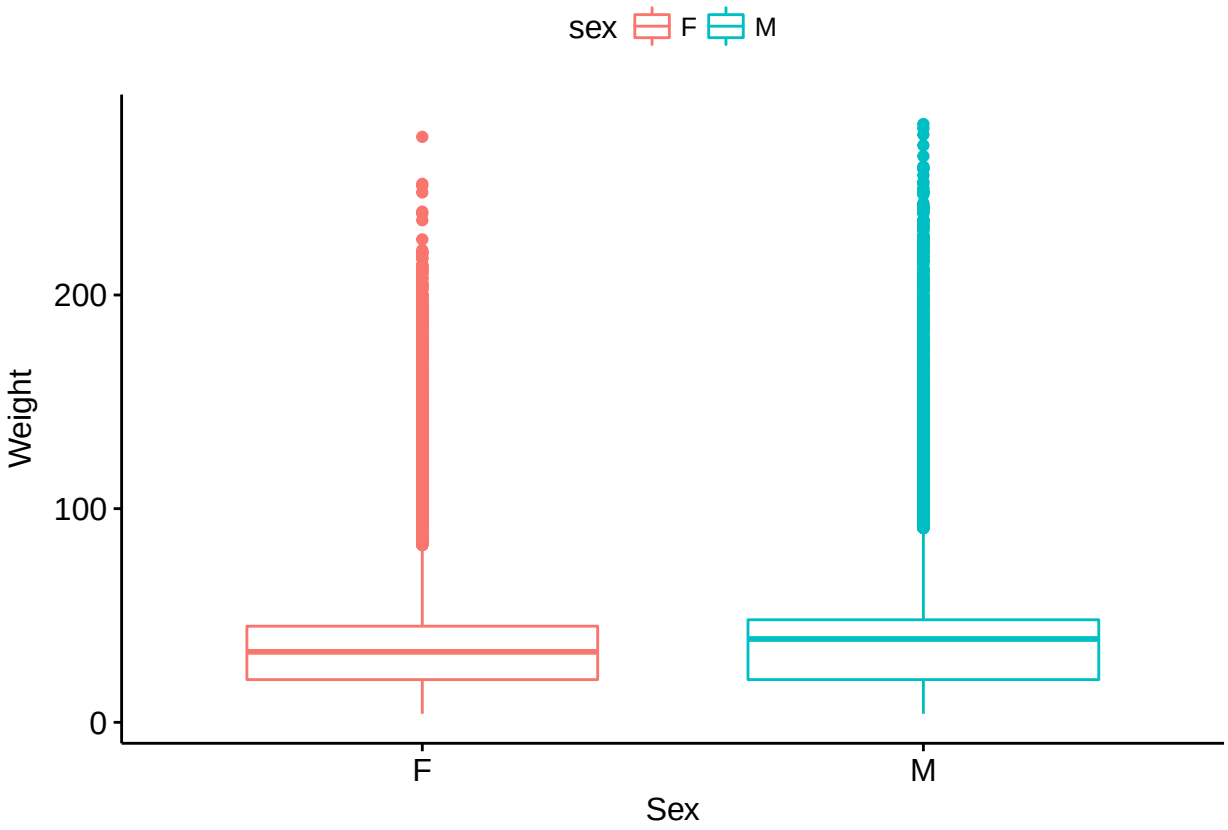
The summary of the linear model is listed below:

```
##
## Call:
## lm(formula = hindfoot_length ~ weight, data = most_common_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.4594  -0.7510   0.0962   0.8566  14.4018
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.695533   0.088406  369.83   <2e-16 ***
## weight       0.076386   0.002024   37.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.362 on 9725 degrees of freedom
## Multiple R-squared:  0.1277, Adjusted R-squared:  0.1276
## F-statistic:  1424 on 1 and 9725 DF,  p-value: < 2.2e-16
```

As we can from the linear model result in the above, the R^2 shows that roughly around 12% of the population fitted to the linear model. The p-value is $<$ 2.2e-16 which is a lot smaller than the significance level alpha $=$ 0.05 implying that there is no relationship between the mean weight and the hindfoot length of the merriami species. In other words, the size of the hindfoot length is not associated with the weight of this species and vice versa.

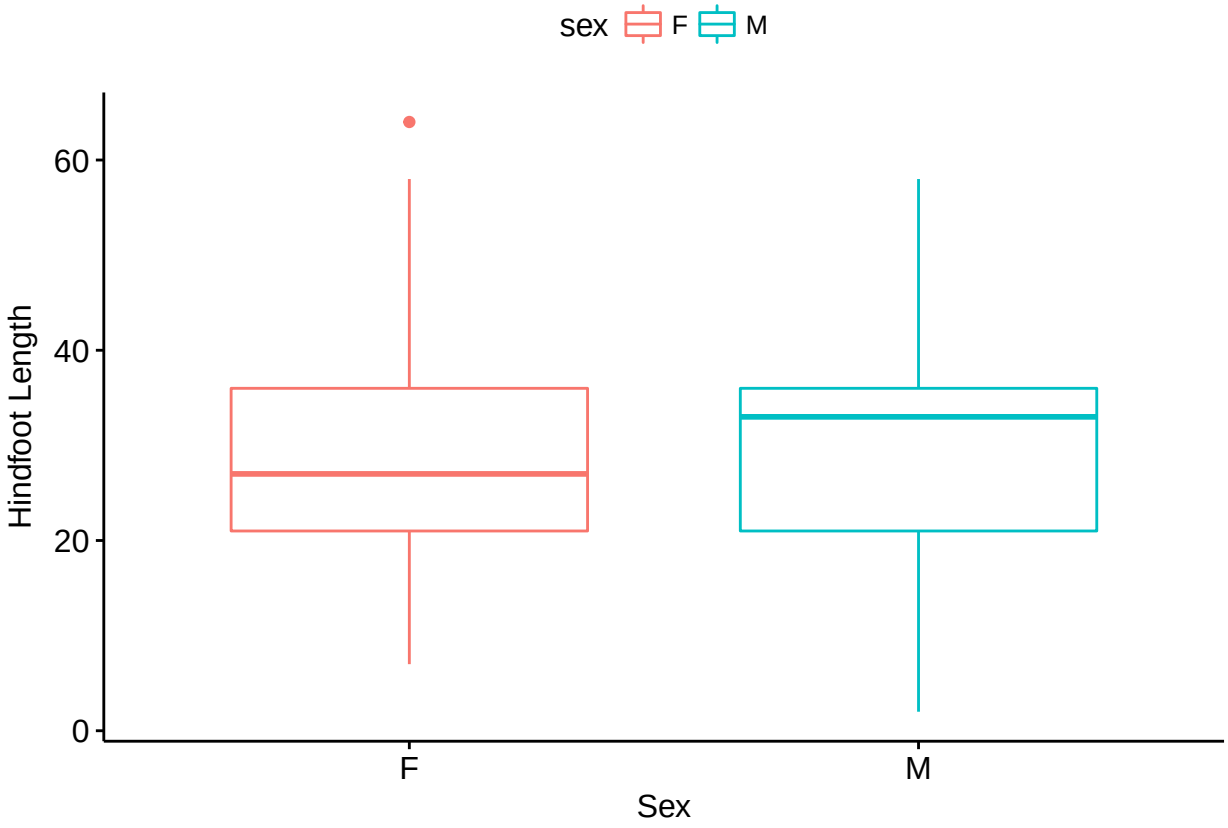
Additional analysis

In addition to all analysis that been committed, we add the t-test statistical analysis to see if the mean weight and the hindfoot length of female species are significantly different from the male species. We also use the box plots to visualize the data for both variables. The result of this analysis can be seen below.




```
##
##  Welch Two Sample t-test
##
## data:  weight by sex
## t = -1.4996, df = 30195, p-value = 0.1337
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.4128861  0.1880436
## sample estimates:
## mean in group F mean in group M
##        41.47010        42.08253
```

In the mean weight analysis, the p-value of the test is 0.1337, which is higher than the significance level alpha = 0.05. With this result, we conclude that female's average weight is *not significantly different* from the male's average weight.




```
##
##  Welch Two Sample t-test
##
## data:  hindfoot_length by sex
## t = -8.3733, df = 30502, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.1246565 -0.6980042
## sample estimates:
## mean in group F mean in group M
##        28.73533        29.64666
```

In the mean hindfoot length analysis, the p-value of the test is $< 2.2e\text{-}16$, which is a lot lesser than the significance level alpha $= 0.05$. It concludes that female's average hindfoot length is *significantly different* from the male's average hindfoot length.