

The Weight and Hindfoot Length Relationships of Rodents in Southern Arizona

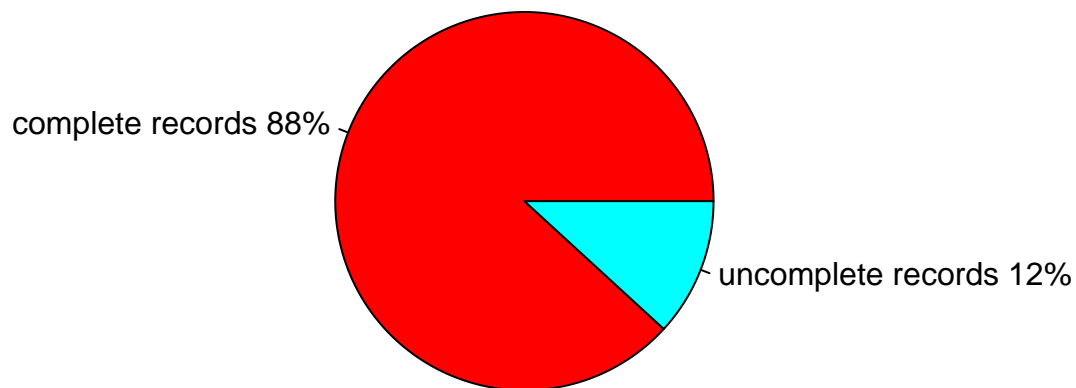
Deky and Tono

March 31, 2018

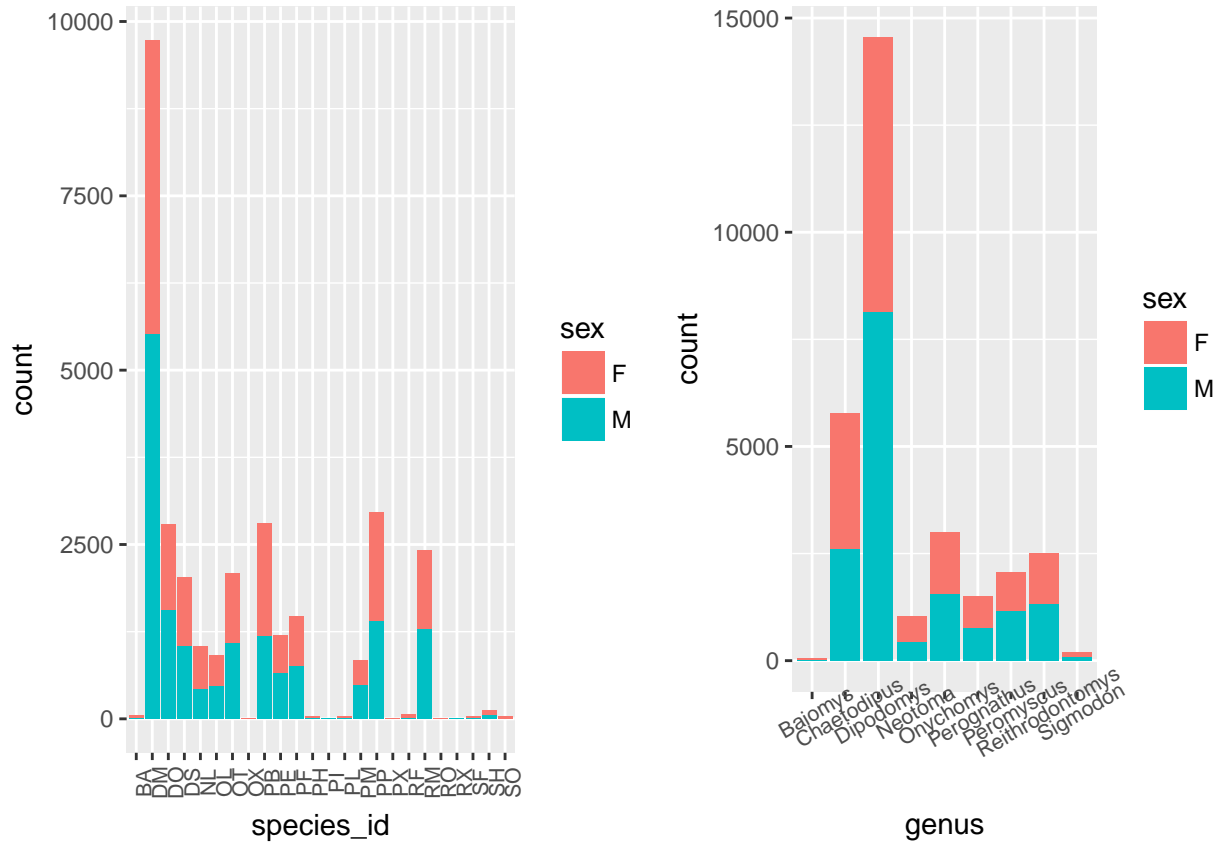
Summary of the Data

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. Out of the total data, there were 12% uncomplete records that we excluded from our analysis. The remaining 88% were served as the basis data of the analysis.

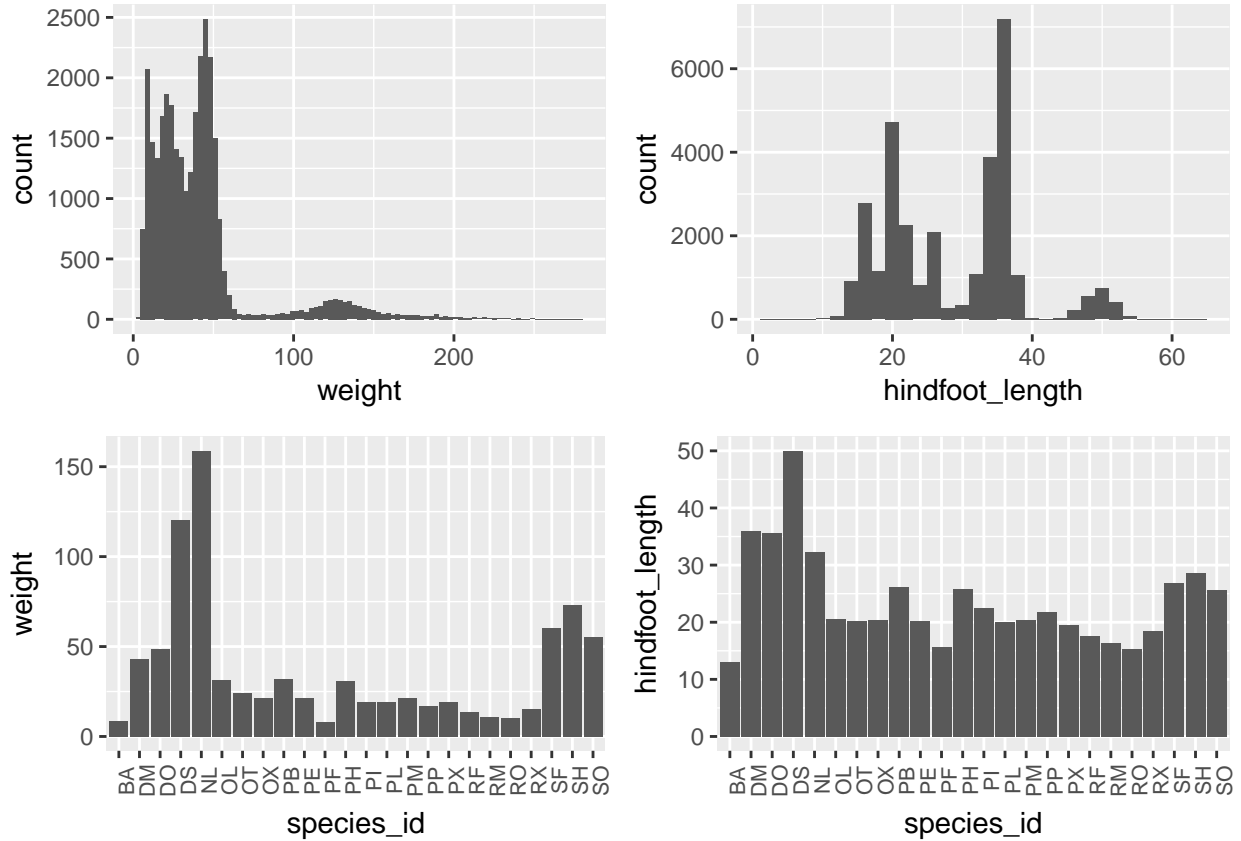
Pie Chart of Data Completeness



From the total 88% of complete data, we further delineated the data profile based on the number of species and genus that had been recorded. 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. From the figure below, we can see that species_id DM (merriami species) dominated the species entered in this data. With this condition we can assume that this species is the most abundance species in the study area. As merriami species is belong to the Dipodomys, the abundance of merriami species controled the amount of Dipodomys genus entered in this data. This figure also delineates roughly that the males are more dominant than the females. Here we can see that the blue bars are taller than the orange bars almost at every bars in both plots (species_id and genus).



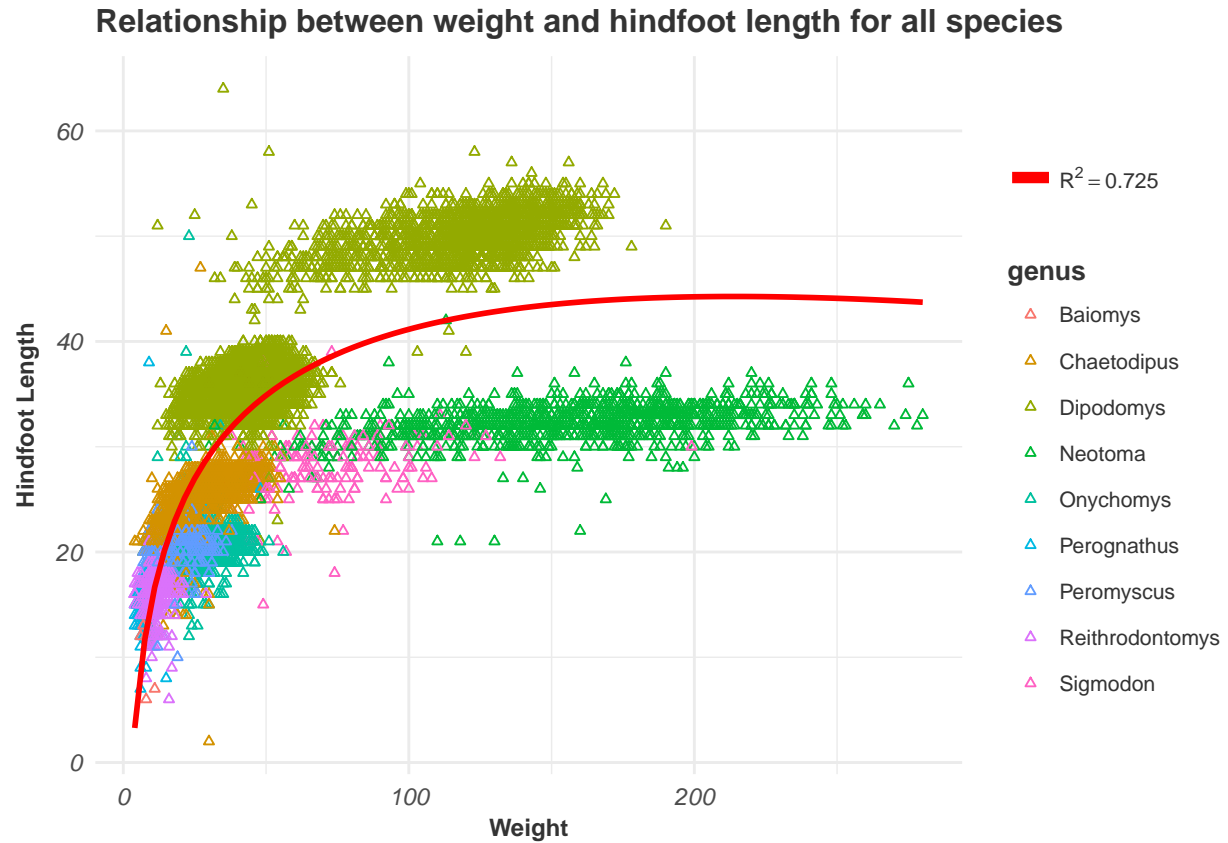
In addition to the general pattern of the data in the above, we also plotted the distribution of weight and hindfoot length for the overall sample and the distribution of the average (mean) weight and hindfoot length for each species. From these plots, we can see that most of the weight are fall in between 10 to 60 grams. Meanwhile, from the hindfoot length, we can see that there are three clusters such as 15 mm to 25 mm, 32 mm to 38 mm, and small amount of ± 50 mm. From the mean weight for each species id, the species id NL (albigula species) have the highest weight which is above 150 grams, followed by the species id DS (spectabilis species) in the second place just below 125 grams. However, the weight of species id NL does not correspond to the its hindfoot length. In this category, the longest hindfoot length is found on species is DS where the species id NL just in fourth place (around 32 mm).



The Weight and Hindfoot Length Relationship Analysis

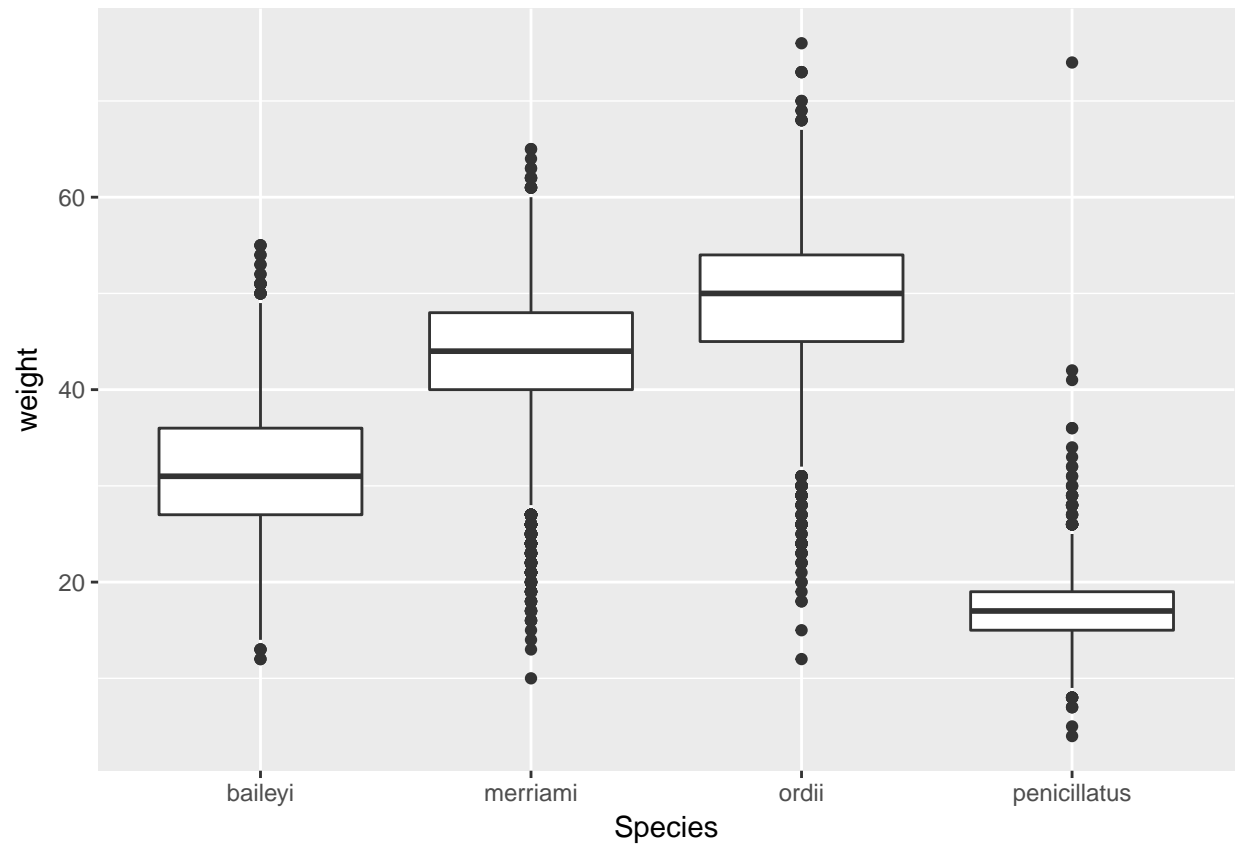
All species

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 0.725 of the population are matched with this trendline. From this trendline, we can assume that, overall, the growth in hindfoot length are likely stopped at around 40 to 50 mm while the weight could continue growing more than 200 grams. In another word, we can assume that the rodent could gain more weight without gaining more length in hindfoot length.

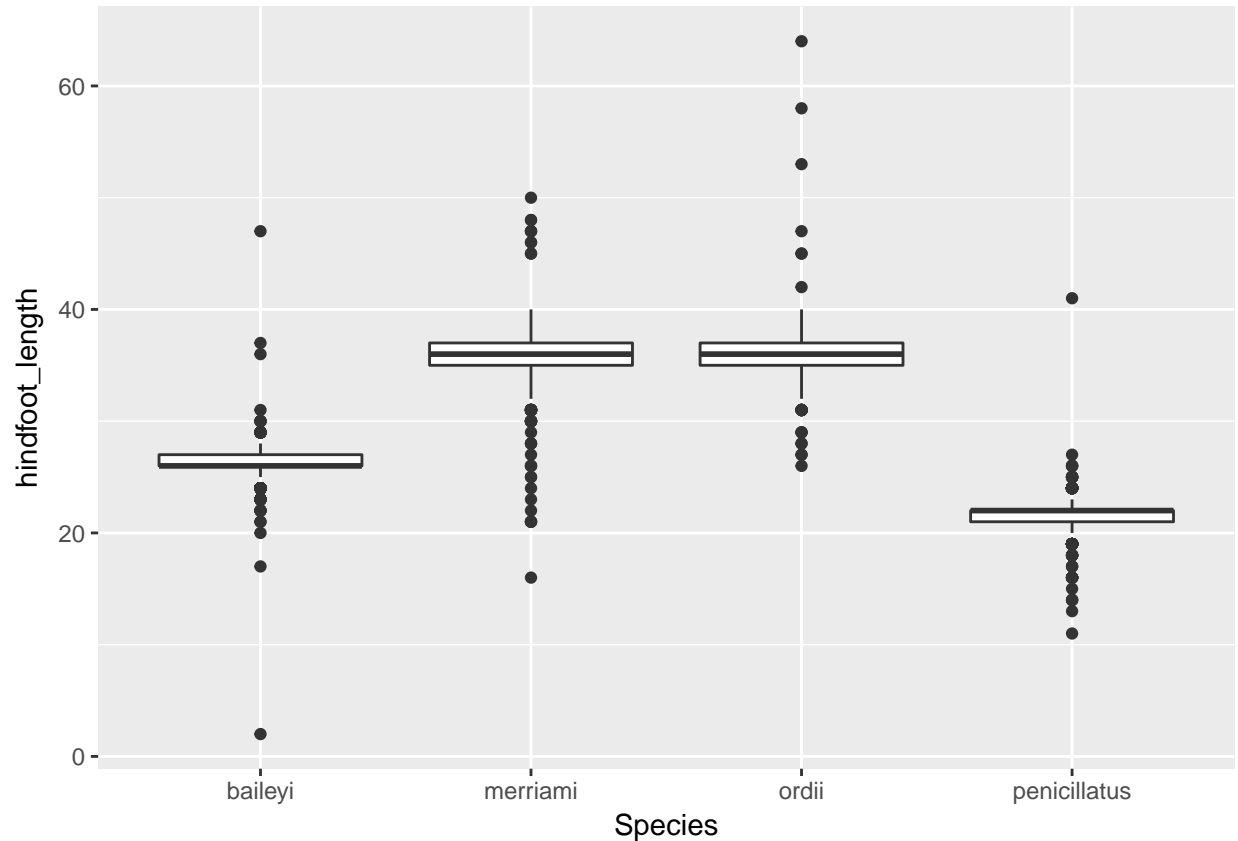


The most recorded species

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, with summary as follows:

```
##
## 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
```

From the intercept, it can be estimated that the mean value of hindfoot length considering the weight is 32.69 mm. From the slope term in the model, it can be said that for every 1 gram increase in weight, the hindfoot length will rise by 0.076386 mm. The Standard Error measures the average number that the coefficient estimates vary from the true average value of variables, which can be said that the length of hindfoot corresponding to the weight stated earlier may vary by 0.002024 mm.

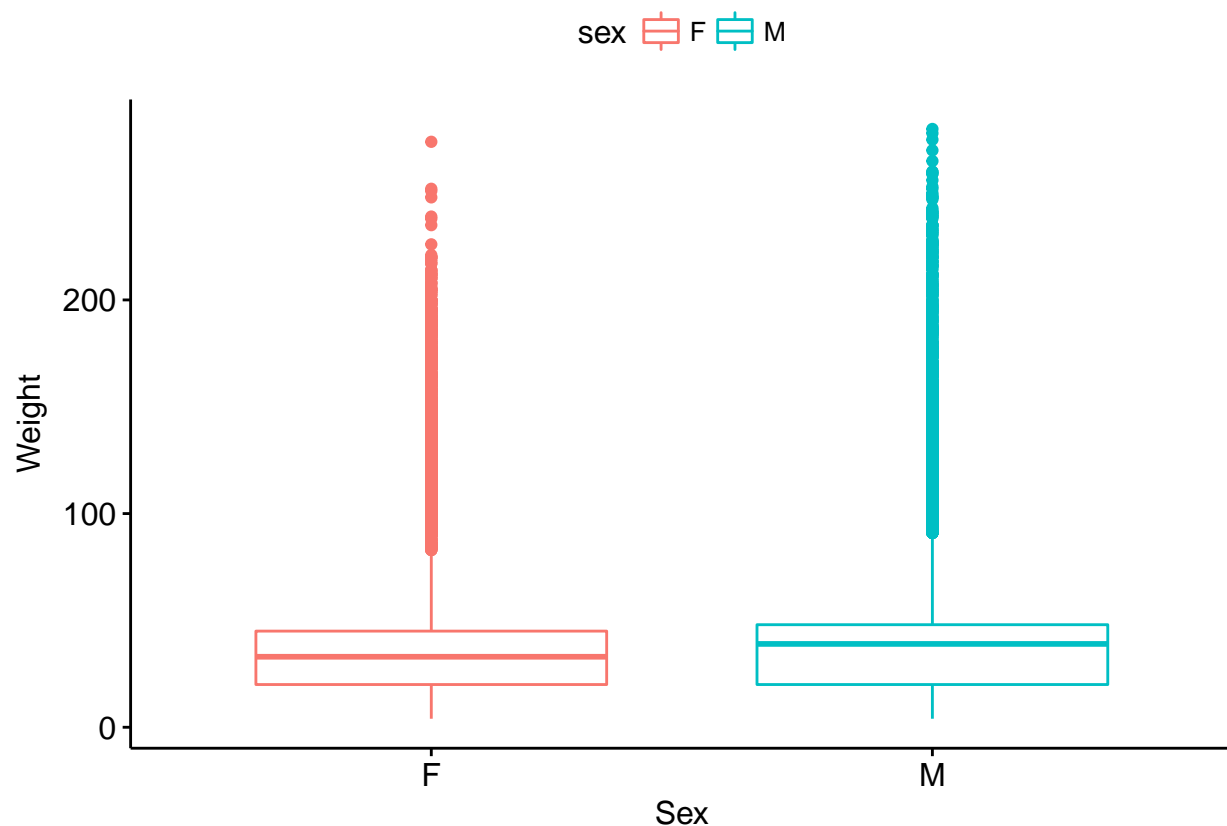
From the value of t and The $\text{Pr}(> |t|)$ generated from the model, the relationship between the hindfoot length and weight can be predicted. The t -value coefficient is a measure of how much the standard deviation estimates the coefficient away from 0. The t -statistical value is relatively far from zero and relatively large against the standard error, which can indicate the relationship exists. $\text{Pr}(> |t|)$ found in the model output corresponds to the probability of observing any value equal to or greater than t . From the model, the p -value is very close to zero. Three “stars” represent a very significant p value. Consequently, the small p -values for intercept and slope indicate that we can conclude that **there is a relationship between the length of hindfoot and the weight of species merriami**.

The Residual Standard Error is 1.362, which it is can be said that the percentage error is 4.17%. Also note that the Residual Standard Error is calculated with 9725 degrees of freedom. Simply put, the degree of freedom is the number of data points that go into the estimated parameters used after taking into account this parameter (restriction). In this case, we have more than 9,000 data points and two parameters (intercept and slope).

R-squared statistics (R^2) give a measure of how well this model fits with the actual data. The R^2 we get is 0.1277. Or about 13% of the variants found in the response variable (hindfoot length) can be explained by the predictor variable (by weight)

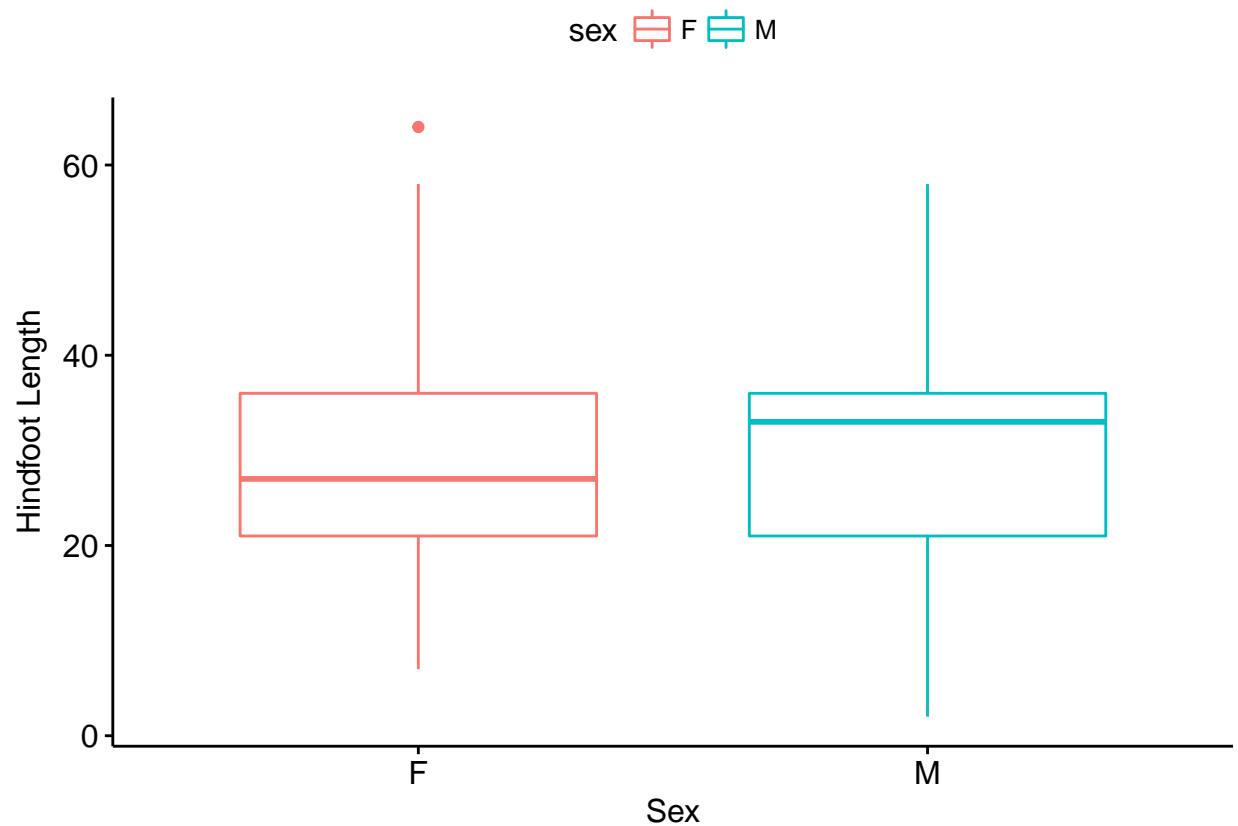
Additional analysis

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 $< 2.2e-16$, 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.