Assignment 3 (R)

Yonatan Ady, Nicole Keefner, Swan Tan

April 1, 2018

Working with data and writing a report using R

This report will summarize what we have done for this group assignment. There are 3 separate chapters that we use to group our work :

1. Exploratory Analysis.
2. Analysis based on Rodent.
3. Statistical Analysis.

Chapter 1. Exploratory Analysis

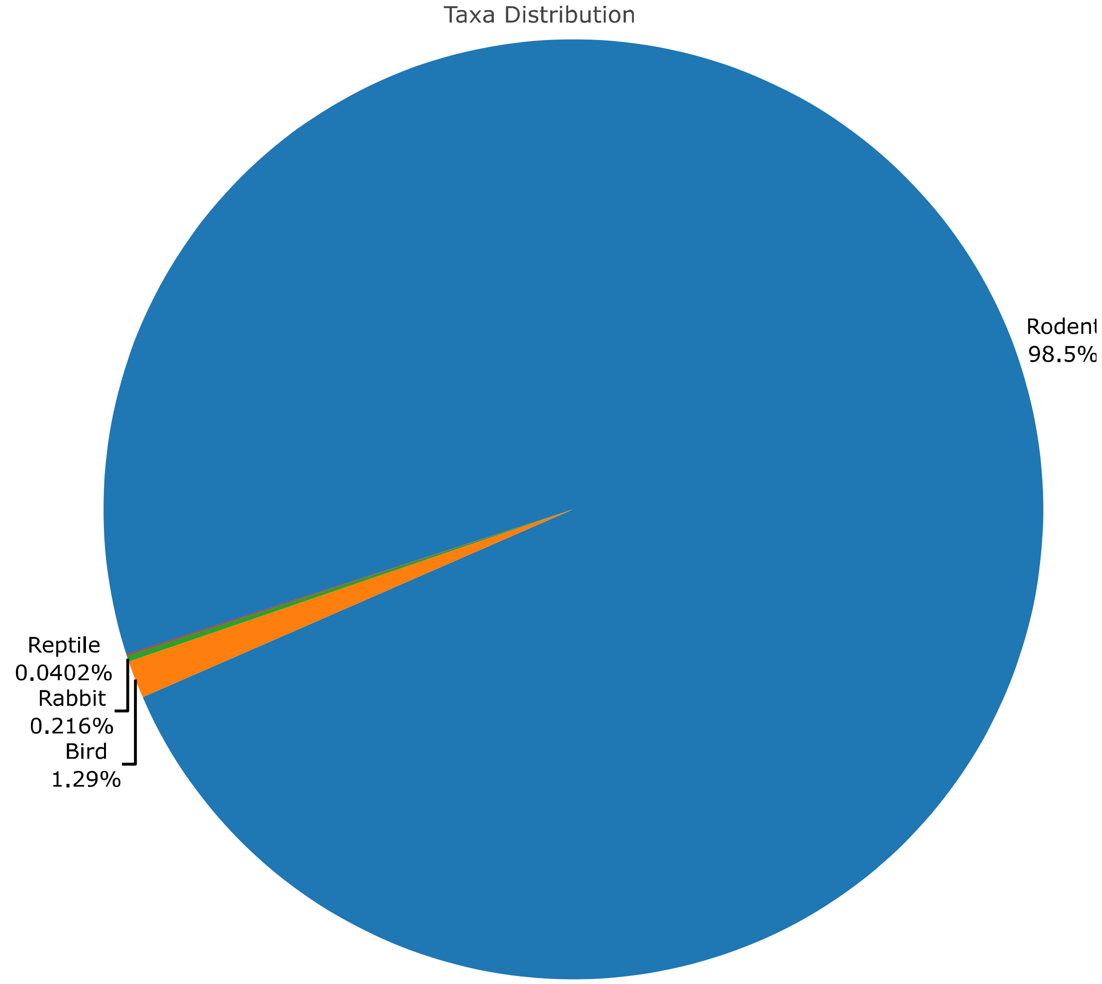
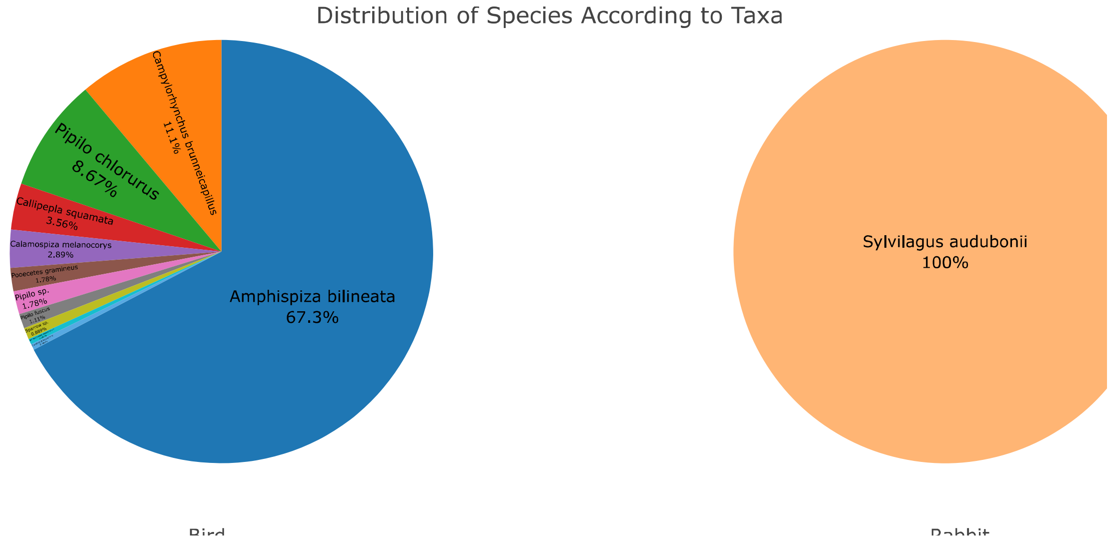


Figure 1.1 | Taxa distribution of the survey data. Rodent is the major taxon, which contributed 98.50% of the overall taxa distribution. Bird, rabbit and reptile are the minority, which contributed approximately 1.29%, 0.22% and 0.04% respectively.



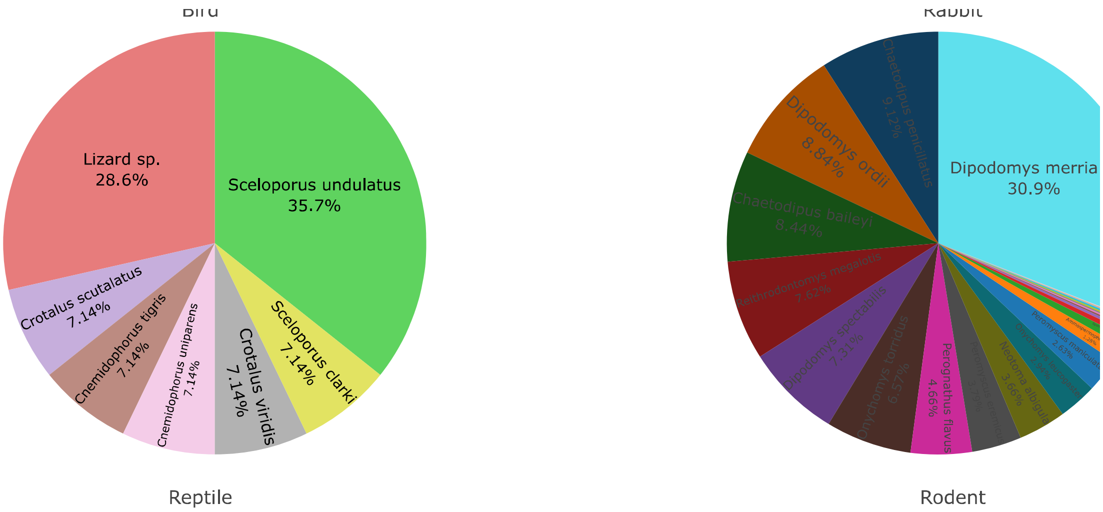


Figure 1.2 | Species distribution of individual taxon. Bird has a total of 11 species, with Amphispiza bilineata being the majority (67.3%), while rabbit only has one, Sylvilagus audubonii. Reptile has seven species in total, with Sceloporus undulatus (35.7%) and Lizard sp. (28.6%) being the two major species. Rodent has the most abundant species distribution, a total of 29 species, with Dipodomys merriami contributed 30.9% of them.

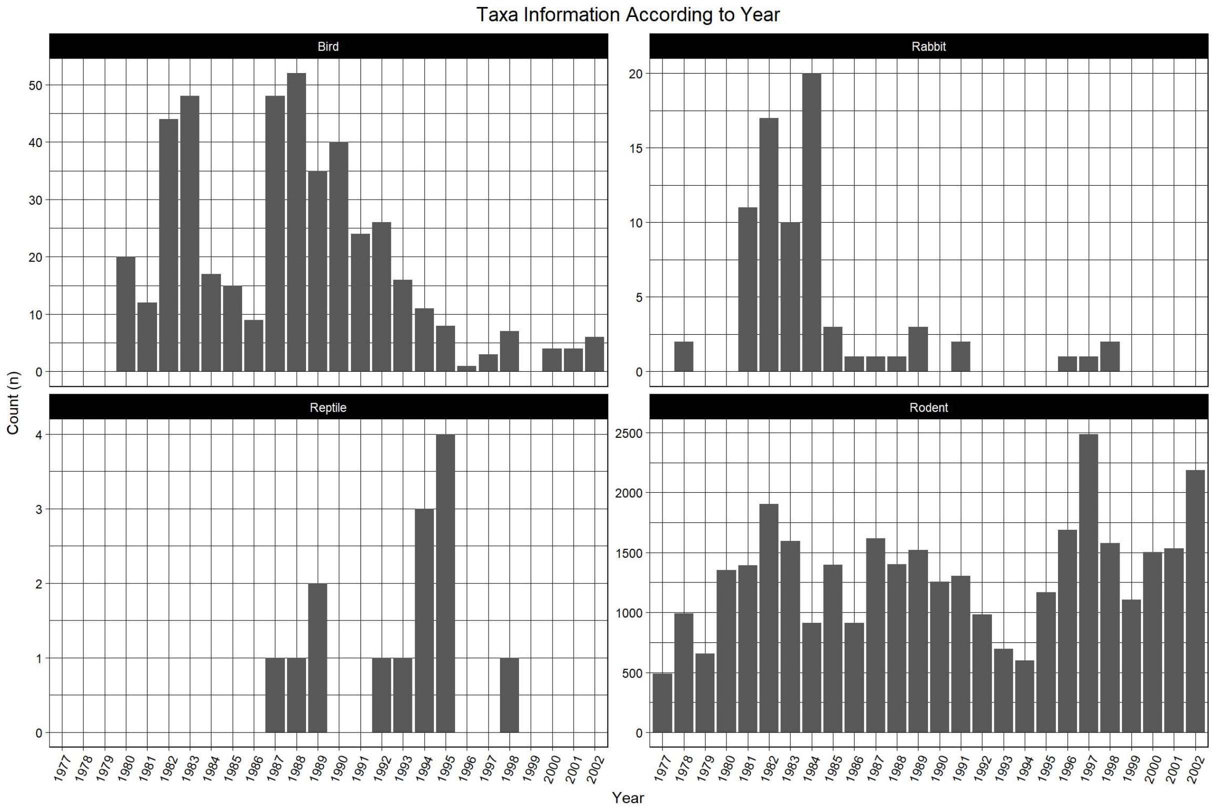


Figure 1.3 | Taxa observations spanning 26 years (1977 to 2002). Rodent showed consistent observations every year, 1977 being the peak of observation. Bird and rabbit have the similar pattern but missing information for a couple of years. Reptile was only observed for eight intermittent years.

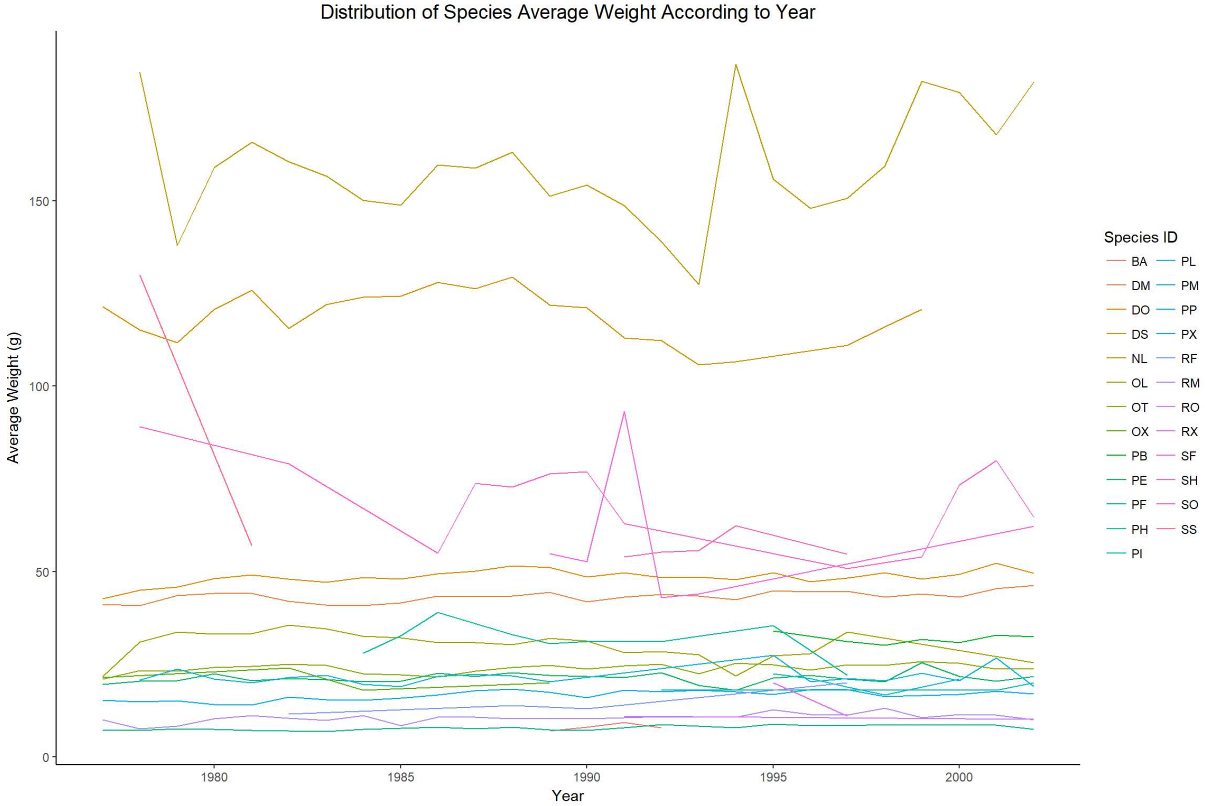


Figure 1.4 | Species average weight distribution from the year 1977 to 2002. Only rodent has the complete information on average weight, hence the species IDs herein refer to the species from rodent taxon. Most species have low average weight, i.e. below 50g and not every species has consistent distribution from the year 1977 to 2002. NL has the maximum of average weight among others and shows consistent distribution throughout the years.

Chapter 2. Analysis based on Rodent Data

Based on preliminary investigations, we determined that Rodent was the only taxon for which hindfoot length or sex were recorded. Because of this, any graphs including information about hindfoot length or sex will be strictly for rodent species in the study. The number of rodents in the study is below.

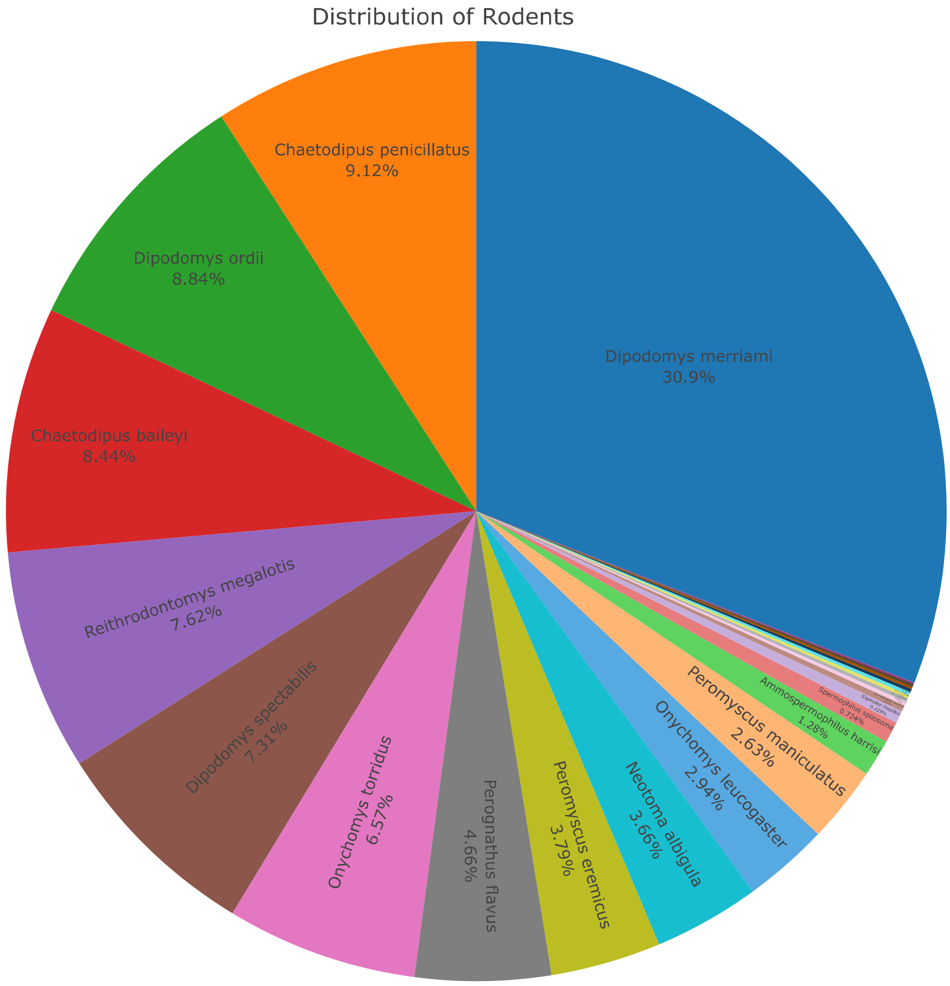


Figure 2.1 | The pie chart showed the distribution of 29 rodent species. Dipodomys merriami is the most abundant species, contributing 30.90% of the total. The second abundant species is Chaetodipus penicillatus, 9.12%, followed by Dipodomys ordii, 8.84%. The three least abundant species are Chaetodipus sp., Reithrodontomys sp. and Spermophilus tereticaudus.

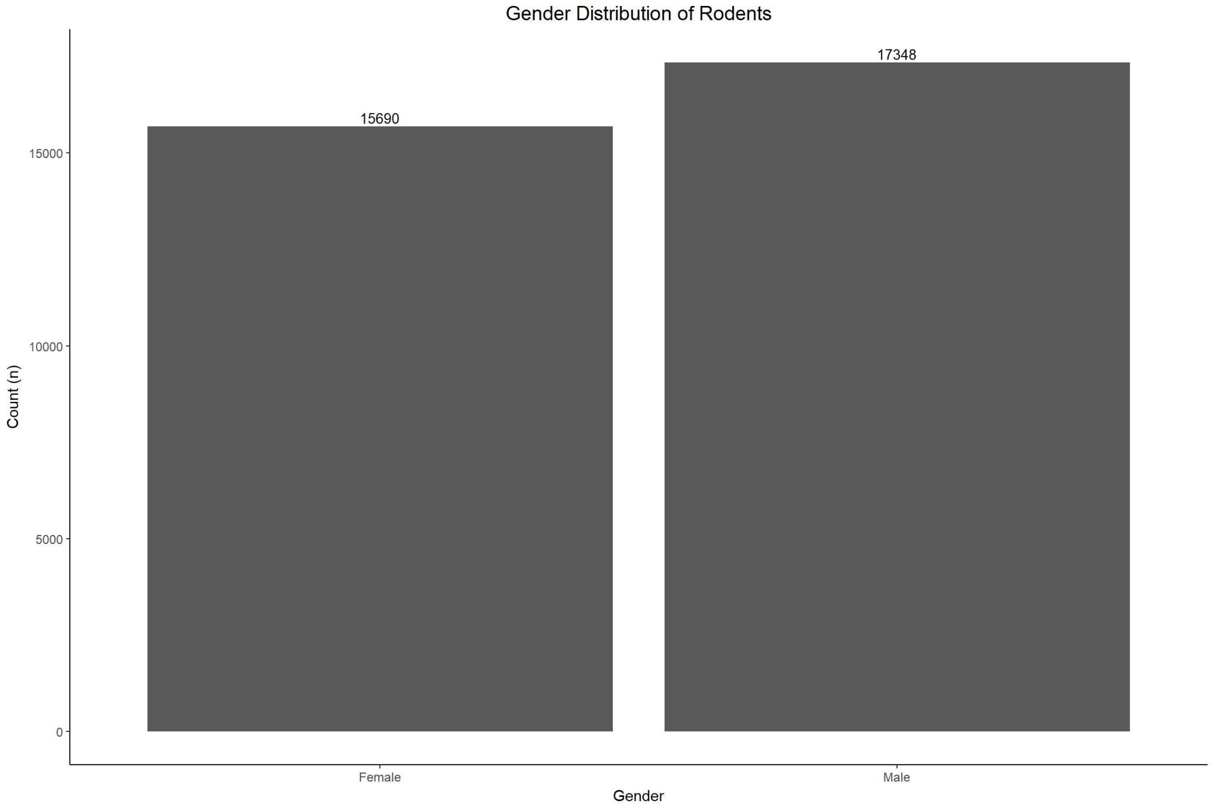


Figure 2.2 | The bar plot showed gender distribution of rodents. There are 17,348 of male and 15,690 of female. Male has a higher count of 1,658 than female.

```
## Rodent
##  31438
```

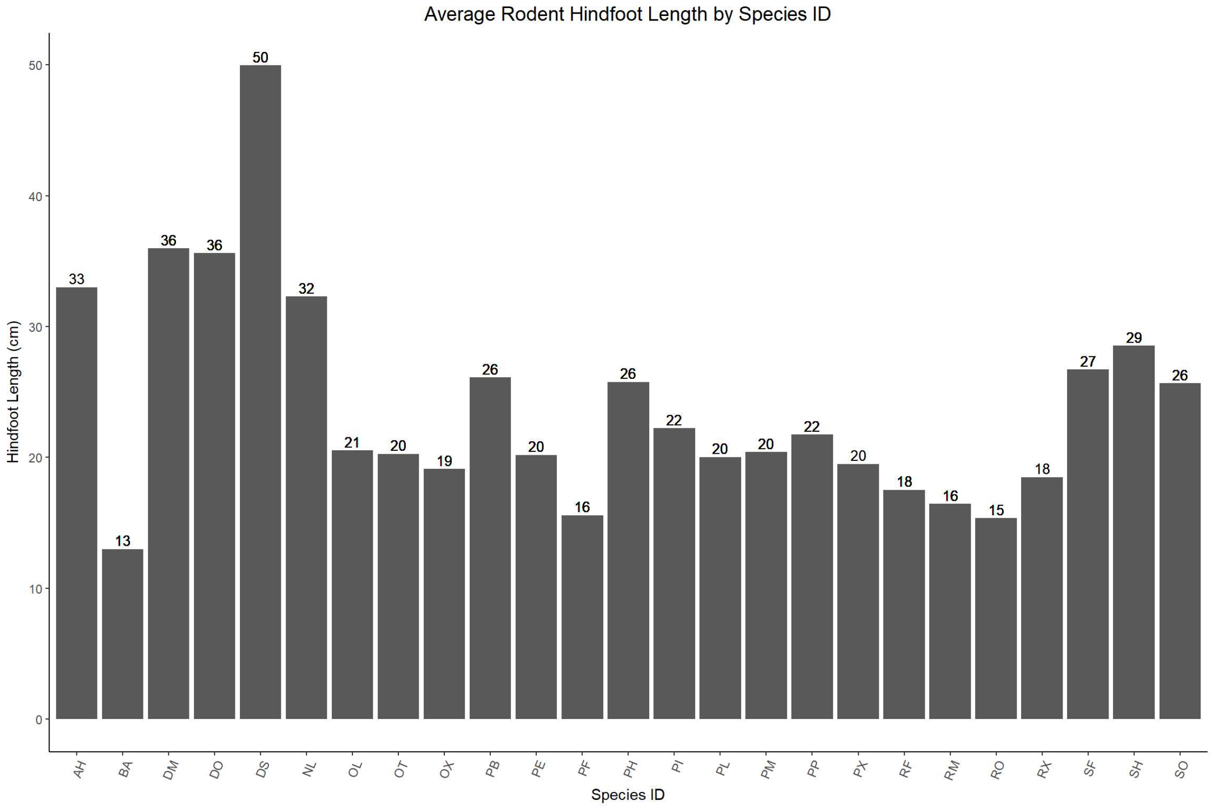


Figure 2.3 | The bar plot showed average hindfoot length by rodent species ID. DS has the longest average hindfoot length, i.e. 50cm, whereas BA has the shortest, i.e. 13cm. Most of the species are within the range of 18-22cm.

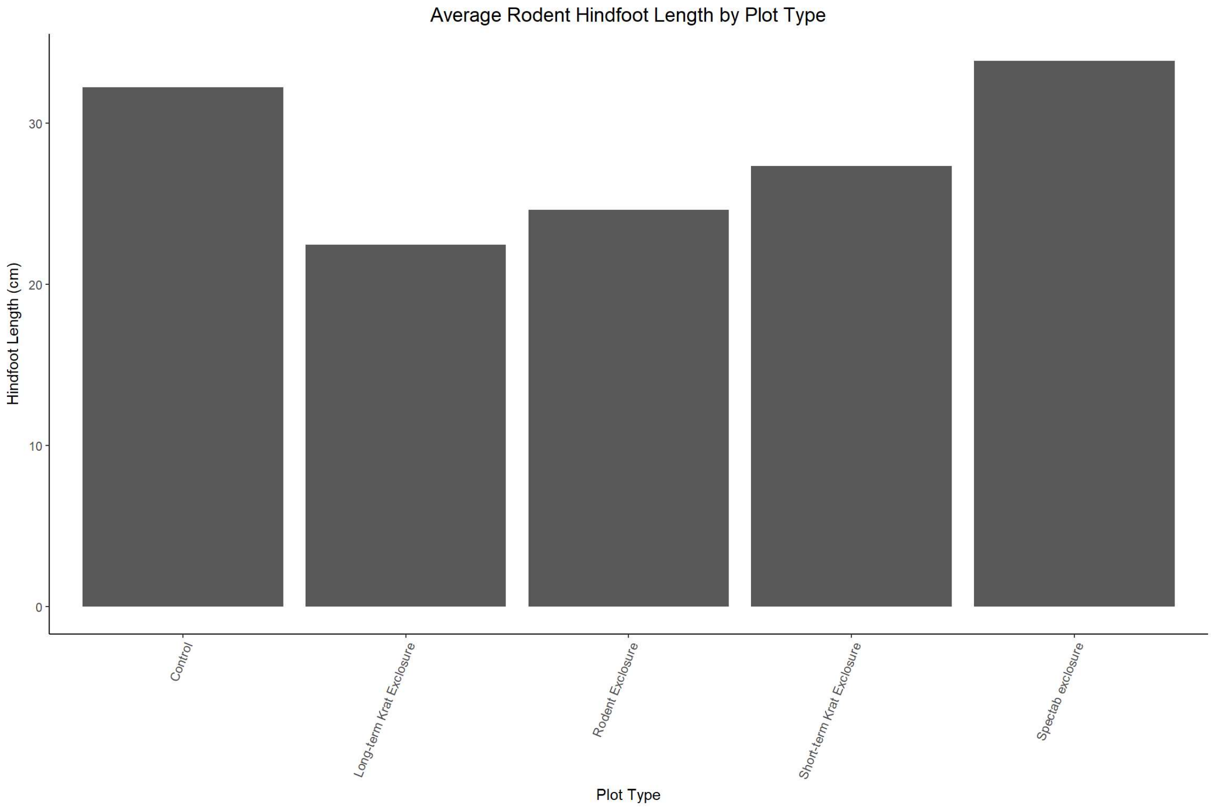


Figure 2.4 | The bar plot showed average hindfoot length by plot type. Control and Spectab exclosure contain similar average rodent hindfoot length information.

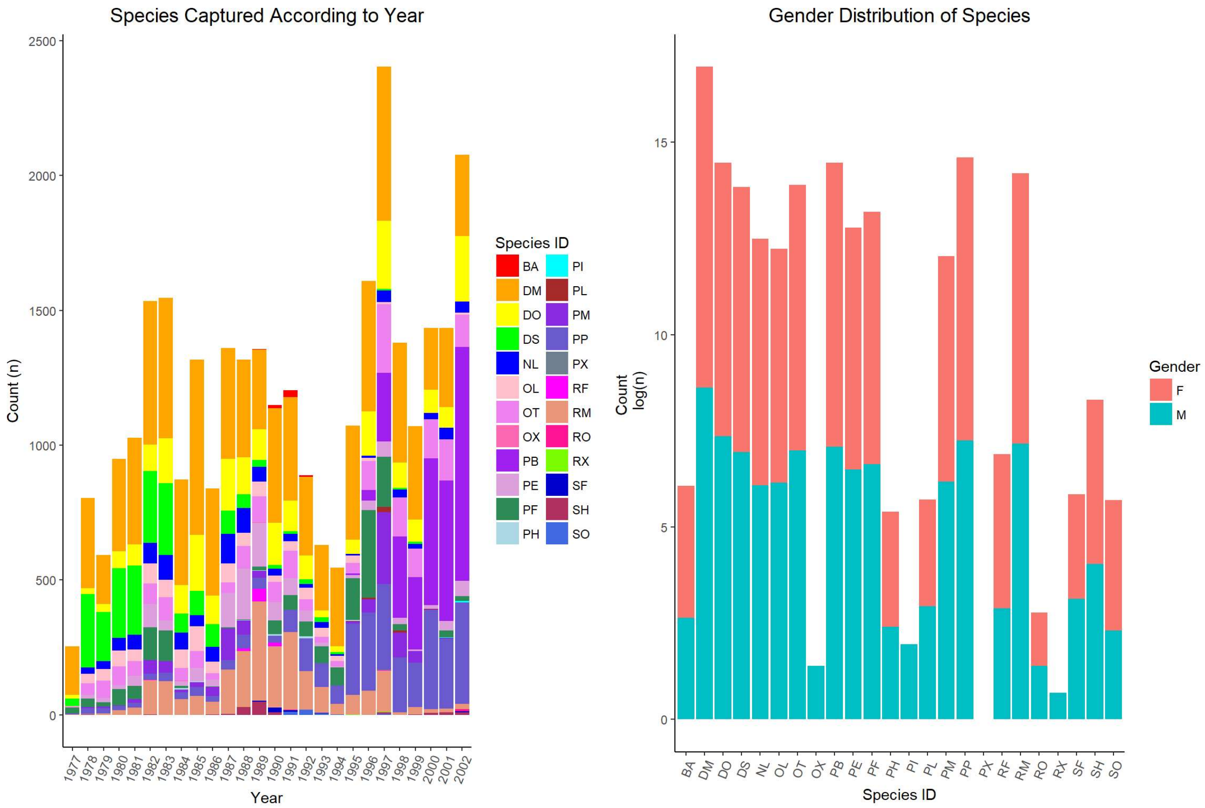


Figure 2.5 | The stacked bar plot on the left panel showed the total of rodent species captured per year. DM has the highest count from the year 1977 to 1999. The amount of PM overtook starting the year 2000 to 2002. The right panel showed gender distribution in terms of species IDs. Overall, most species have a even female and male distribution except OX, PI and RX only have male. PX has one female and male, however, the number is insignificant to be shown on the plot.

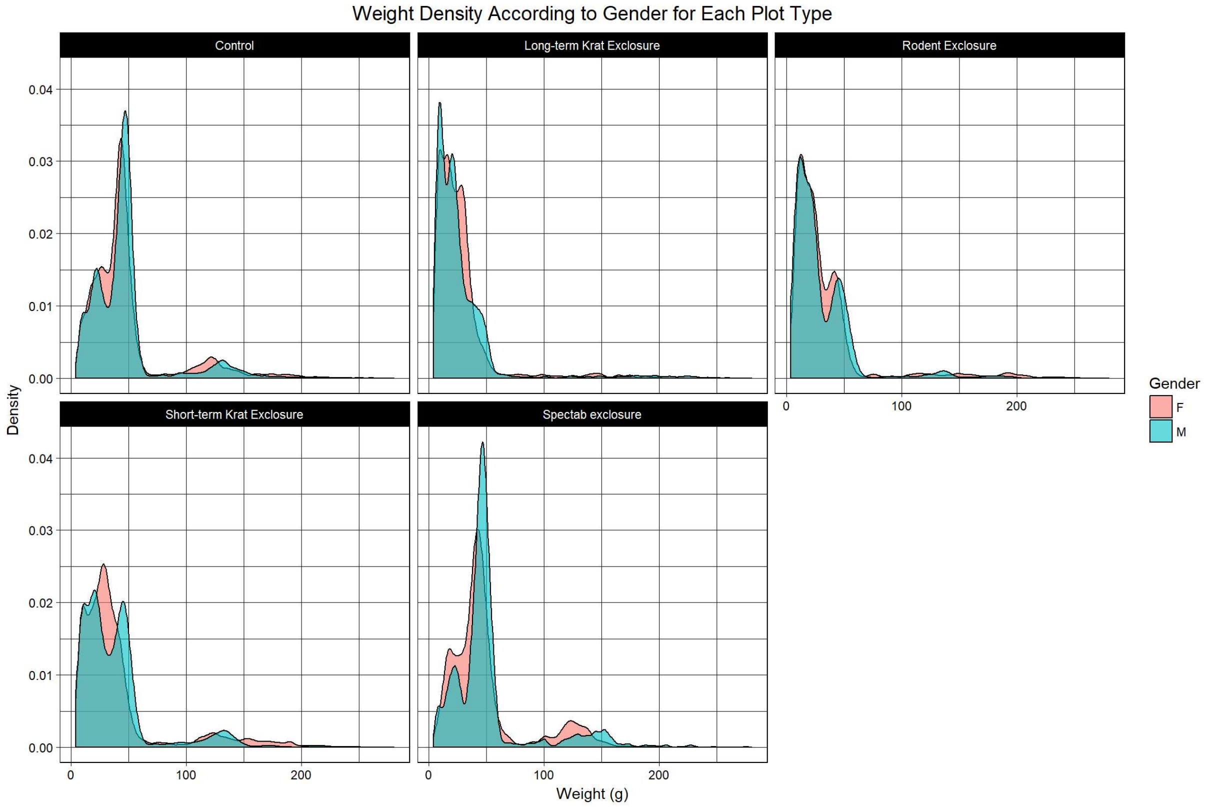


Figure 2.6 | Weight density distribution by gender for five different plot types. Most plot type have the density skewed to left, showing the weight distribution of both gender is between 0-100g. Spectab exclosure plot type, however, has male outweighed female. From Control, Rodent Exclosure and Short-term Krat Exclosure plot types, although female has lower density but they are more heavier than the male.

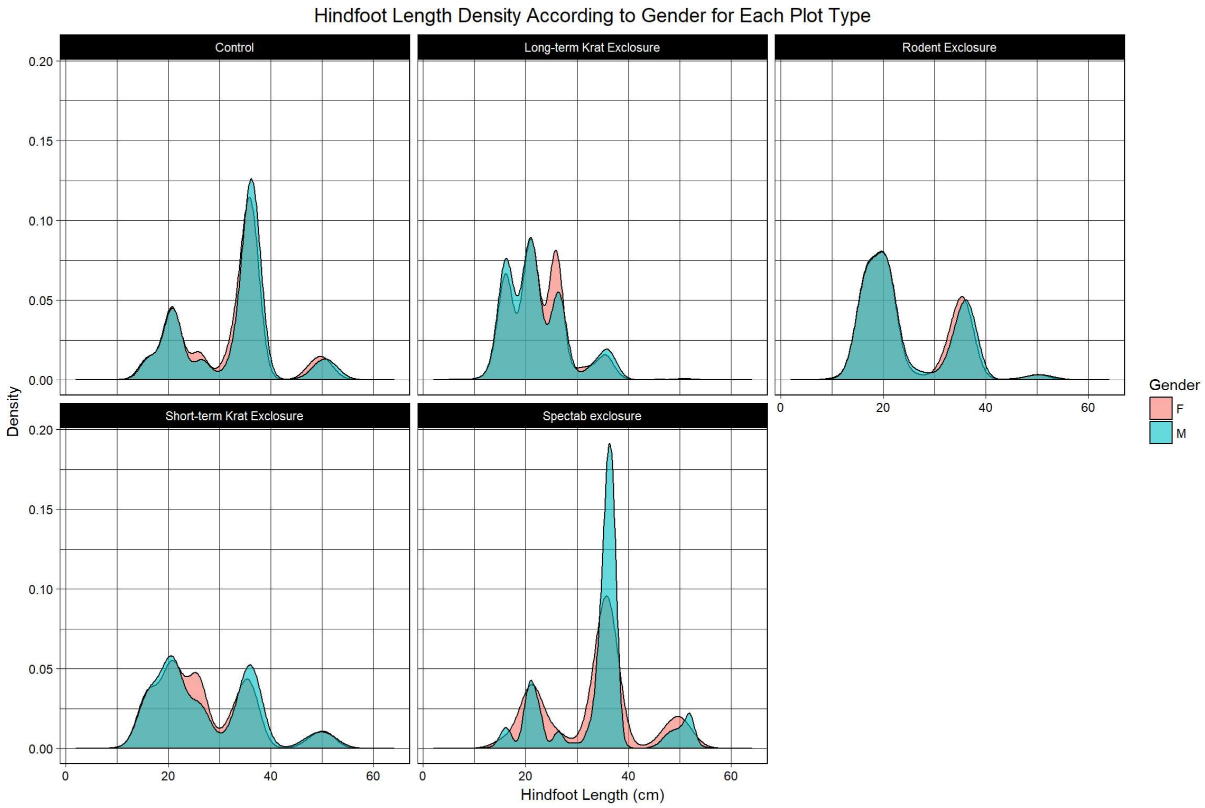


Figure 2.7 | Hindfoot length density distribution by gender for five different plot types. Similar pattern of density distribution for Control and Rodent Exclosure was observed. Female and male are overlapped in both plots. Spectab exclosure has the male more densely distributed for hindfoot length within the range of 30 - 40 cm.

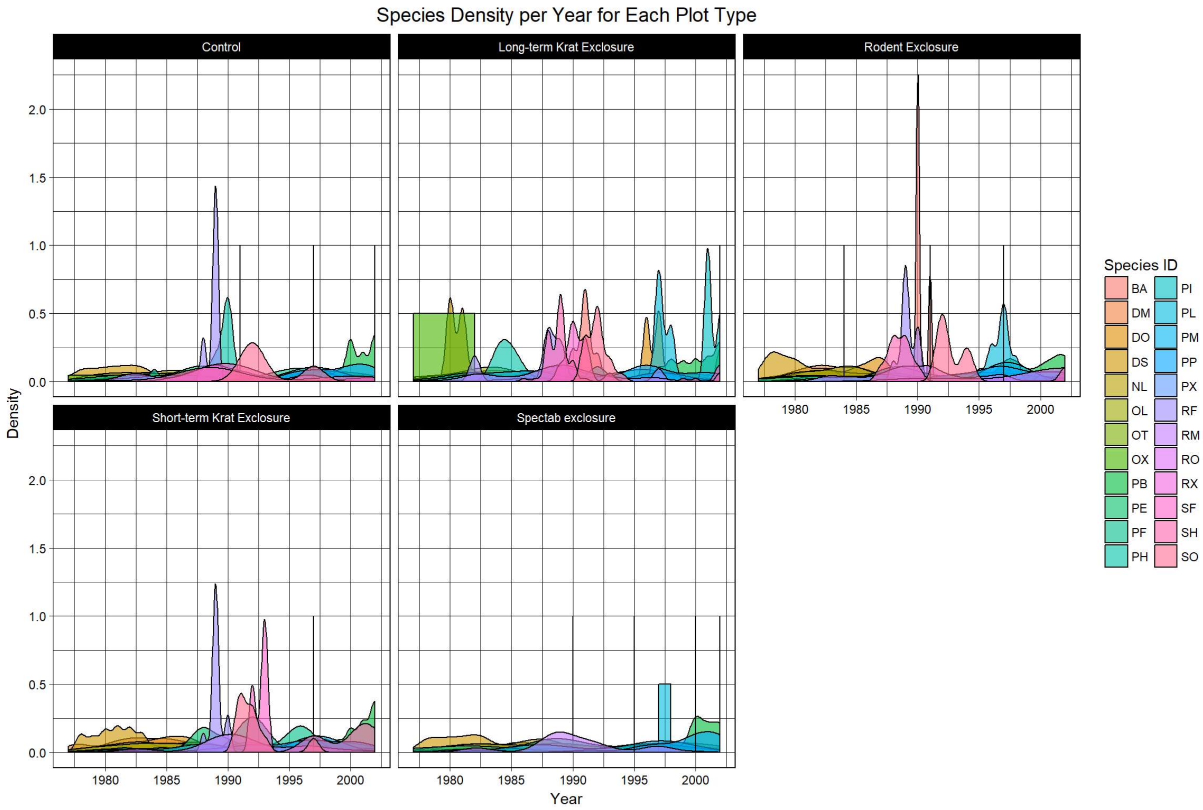


Figure 2.8 | Species density distribution by year for five different plot types. Most plot types have species concentrated within the range of year 1985 and 1995. The density of species distribution of Spectab exclosure tends to skewed towards right.

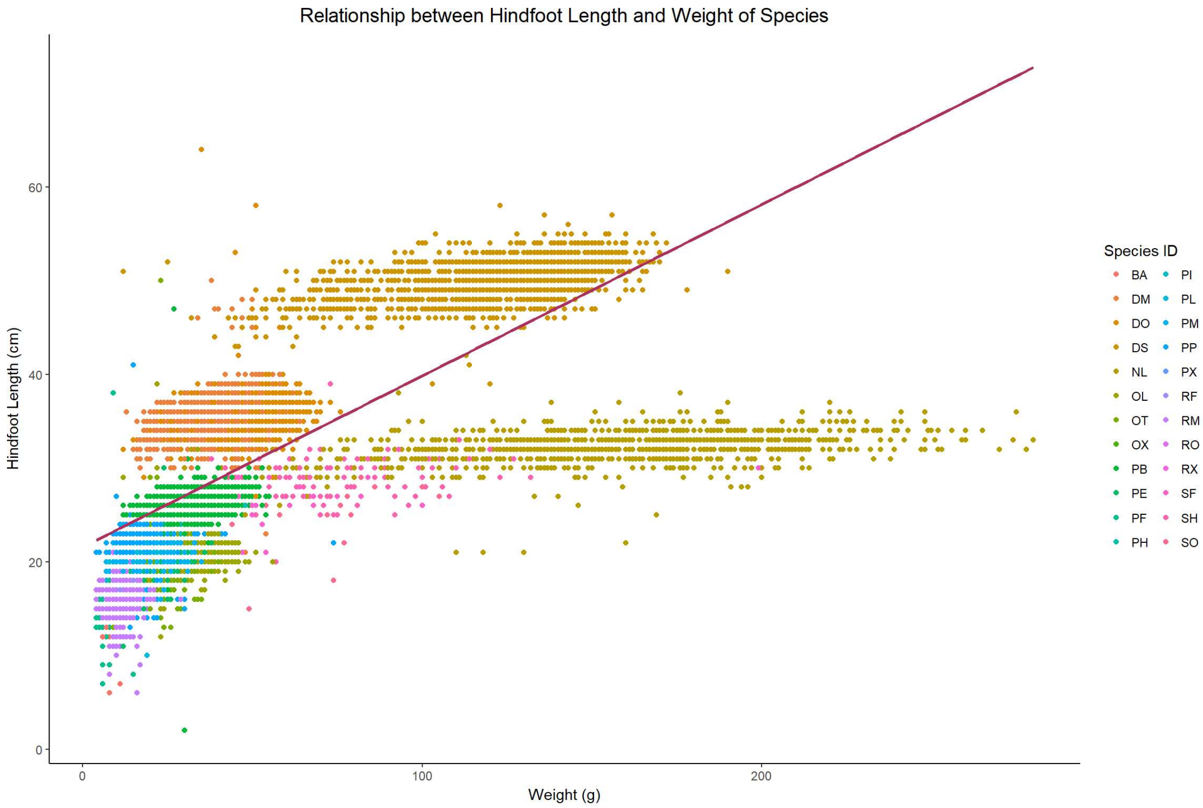


Figure 2.9 | Correlation of hindfoot length and weight by species IDs.

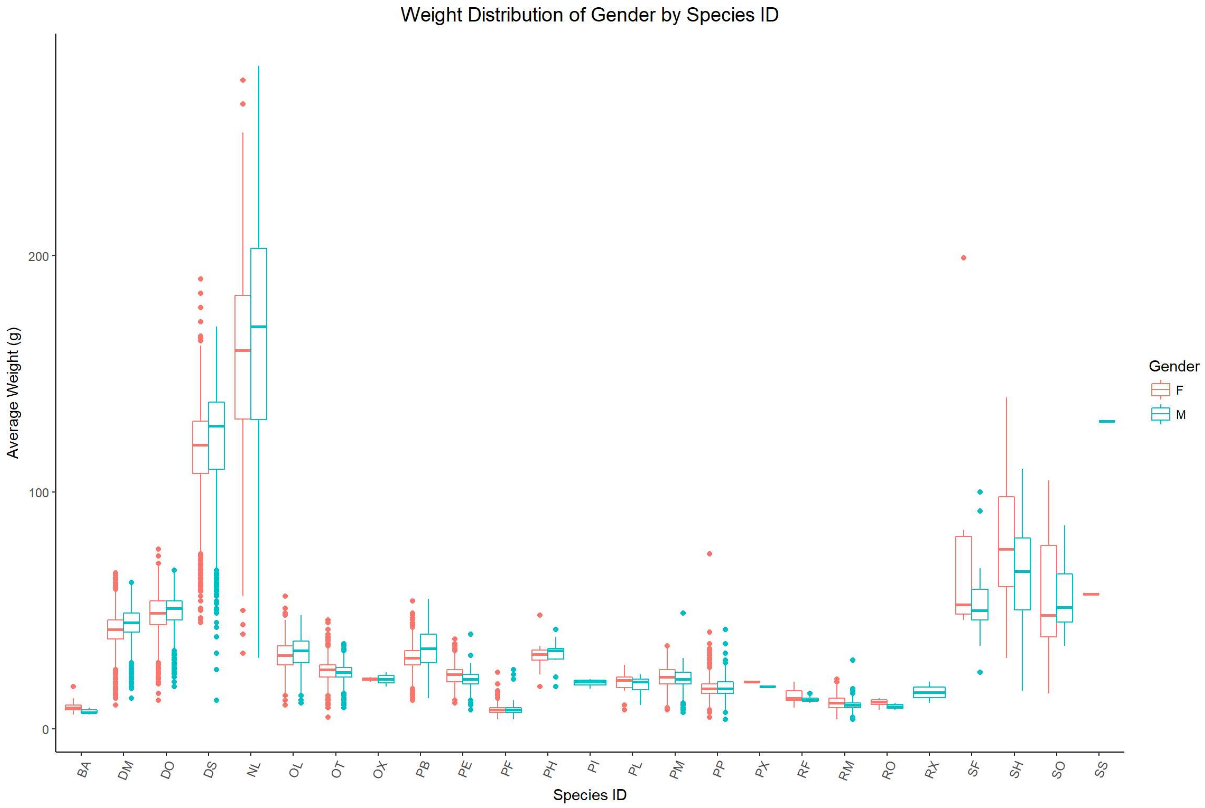


Figure 3.0 | Rodent weight distribution for female and male according to species IDs.

Chapter 3. Statistical Analysis

The above figure of "Weight for Male and Female Rodents by Species ID" suggests there is a relationship between sex and weight (g) in rodents. We conducted a Student's t-test to compare the weights of males to that of females.

```
##
##  Welch Two Sample t-test
##
## data:  Rodent_Female$weight and Rodent_Male$weight
## t = -2.0226, df = 31751, p-value = 0.04312
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.62412287 -0.02552529
## sample estimates:
## mean of x mean of y
##  42.17055  42.99538
```

Because the p-value is less than the alpha of 0.05, we reject the null hypothesis that the mean weights of the two sexes are the same. In the output above, "mean of x" is the mean of females and "mean of y" is the mean of males.

The above figure of "Relationship between Hindfoot Length and Weight of Species" suggests there is a relationship between hindfoot length (cm) and weight (g) in rodents. We created a linear model to determine if there is indeed a relationship between these variables.

```
##
## Call:
## lm(formula = hindfoot_length ~ weight, data = rodent_complete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.400  -5.584  -0.509   6.125  36.028
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.572622   0.061245   352.2   <2e-16 ***
## weight       0.182831   0.001115   164.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.964 on 30674 degrees of freedom
## Multiple R-squared:  0.4673, Adjusted R-squared:  0.4673
## F-statistic: 2.69e+04 on 1 and 30674 DF,  p-value: < 2.2e-16
```

Because the p-value is less than the alpha of 0.05, we reject the null hypothesis that the slope of the linear regression model does not differ significantly from zero. In addition, the multiple R-squared value is used to describe how well a given model explains variation in the data. In this case, this model explains 46.73% of the variation in the data.

The above figure of "Average Rodent Hindfoot Length by Species ID" suggests hindfoot length (cm) is different according to species ID. We conducted several tests to determine if this was the case. We began by conducting a Bartlett Test of Homogeneity of Variances to evaluate the homogeneity of variance assumption of ANOVA.

```
##
##  Bartlett test of homogeneity of variances
##
## data:  hindfoot_length by species_id
## Bartlett's K-squared = 2269.8, df = 24, p-value < 2.2e-16
```

Because the p-value is less than the alpha of 0.05, we reject the null hypothesis that variances of the levels of species ID are equal. Because of this, we know our data does not meet the assumptions of the ANOVA and a non-parametric alternative must be used. In this case, we used a Kruskal-Wallis rank sum test.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  hindfoot_length by as.factor(species_id)
## Kruskal-Wallis chi-squared = 28746, df = 24, p-value < 2.2e-16
```

Because the p-value is less than the alpha of 0.05, we reject the null hypothesis that all species have the same average hindfoot lengths. However, this test did not inform us as to which species are different from which others. In order to determine this, we conducted a non-parametric Pairwise Wilcoxon rank sum test. In the output below, pairs with values <0.05 are significantly different from each other and we reject the null that the group means are the same.

```
##
##  Pairwise comparisons using Wilcoxon rank sum test
##
## data:  surveys_hfoot_id$hindfoot_length and as.factor(surveys_hfoot_id$species_id)
##
##    AH      BA      DM      DO      DS      NL      OL      OT      OX
## BA 0.90881 -       -       -       -       -       -       -       -
## DM 1.00000 < 2e-16 -       -       -       -       -       -       -
## DO 1.00000 < 2e-16 < 2e-16 -       -       -       -       -       -
## DS 0.90881 < 2e-16 < 2e-16 < 2e-16 -       -       -       -       -
## NL 1.00000 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -       -       -       -
## OL 0.83935 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 -       -       -
## OT 0.79119 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 7.9e-09 -       -
## OX 1.00000 0.00878 5.7e-05 6.6e-05 9.5e-05 8.1e-05 1.00000 1.00000 -
## PB 0.79119 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 2.8e-05
## PE 0.80992 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 2.5e-08 1.00000 1.00000
## PF 0.79119 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.01116
## PH 1.00000 1.3e-11 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00148
## PI 1.00000 0.00016 1.2e-05 1.4e-05 2.2e-05 2.2e-05 0.00546 0.00074 0.15616
## PL 0.90881 2.2e-11 < 2e-16 < 2e-16 < 2e-16 < 2e-16 1.00000 1.00000 1.00000
## PM 0.80819 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.77879 0.03630 1.00000
## PP 0.80992 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00546
## PX 1.00000 0.90881 0.86583 0.87455 0.90881 0.88924 1.00000 1.00000 1.00000
## RF 0.80819 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.05796
## RM 0.79119 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00995
## RO 1.00000 0.00878 5.7e-05 6.6e-05 9.5e-05 7.9e-05 6.0e-05 3.5e-05 0.88924
## RX 1.00000 0.90881 0.86583 0.87455 0.90881 0.88924 1.00000 1.00000 1.00000
## SF 1.00000 1.4e-13 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00135
## SH 1.00000 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00027
## SO 1.00000 1.6e-13 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 < 2e-16 0.00678
##    PB      PE      PF      PH      PI      PL      PM      PP      PX
## BA -       -       -       -       -       -       -       -       -
## DM -       -       -       -       -       -       -       -       -
## DO -       -       -       -       -       -       -       -       -
## DS -       -       -       -       -       -       -       -       -
## NL -       -       -       -       -       -       -       -       -
## OL -       -       -       -       -       -       -       -       -
## OT -       -       -       -       -       -       -       -       -
## OX -       -       -       -       -       -       -       -       -
## PB -       -       -       -       -       -       -       -       -
## PE < 2e-16 -       -       -       -       -       -       -       -
## PF < 2e-16 < 2e-16 -       -       -       -       -       -       -
## PH 1.00000 < 2e-16 < 2e-16 -       -       -       -       -       -
## PI 6.1e-06 0.00083 5.2e-06 0.00237 -       -       -       -       -
## PL < 2e-16 1.00000 < 2e-16 2.1e-10 0.00711 -       -       -       -
## PM < 2e-16 0.01626 < 2e-16 < 2e-16 0.00171 1.00000 -       -       -
## PP < 2e-16 < 2e-16 < 2e-16 < 2e-16 1.00000 4.9e-12 < 2e-16 -       -
## PX 0.79119 1.00000 0.81019 1.00000 1.00000 1.00000 1.00000 1.00000 -
## RF < 2e-16 < 2e-16 < 2e-16 2.1e-14 3.5e-05 4.4e-13 < 2e-16 < 2e-16 0.99628
## RM < 2e-16 < 2e-16 < 2e-16 < 2e-16 4.7e-06 < 2e-16 < 2e-16 < 2e-16 0.80819
## RO 2.7e-05 5.3e-05 1.00000 0.00148 0.04809 0.00214 3.9e-05 4.7e-05 1.00000
## RX 0.79119 1.00000 0.93218 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
## SF 1.00000 < 2e-16 < 2e-16 1.00000 0.00719 2.0e-11 < 2e-16 < 2e-16 1.00000
## SH < 2e-16 < 2e-16 < 2e-16 4.6e-07 0.00021 < 2e-16 < 2e-16 < 2e-16 0.93218
## SO 1.00000 < 2e-16 < 2e-16 1.00000 0.04214 3.4e-09 < 2e-16 < 2e-16 1.00000
##    RF      RM      RO      RX      SF      SH
## BA -       -       -       -       -       -
## DM -       -       -       -       -       -
## DO -       -       -       -       -       -
## DS -       -       -       -       -       -
## NL -       -       -       -       -       -
## OL -       -       -       -       -       -
## OT -       -       -       -       -       -
## OX -       -       -       -       -       -
## PB -       -       -       -       -       -
## PE -       -       -       -       -       -
## PF -       -       -       -       -       -
## PH -       -       -       -       -       -
## PI -       -       -       -       -       -
## PL -       -       -       -       -       -
## PM -       -       -       -       -       -
## PP -       -       -       -       -       -
## PX -       -       -       -       -       -
## RF -       -       -       -       -       -
## RM < 2e-16 -       -       -       -       -
## RO 0.00379 0.73783 -       -       -       -
## RX 1.00000 1.00000 1.00000 -       -       -
## SF < 2e-16 < 2e-16 0.00094 1.00000 -       -
## SH < 2e-16 < 2e-16 0.00022 0.93218 0.00051 -
## SO 4.0e-15 < 2e-16 0.00164 1.00000 1.00000 2.5e-06
##
## P value adjustment method: holm
```