

DSP 539: Assignment 3

Step-1: Data loading

```
surveys <- read_csv("data/portal_data_joined.csv")
```

Step-2: Data Cleaning

New Dataset is created by removing missing data (weight, hindfoot_length, sex)

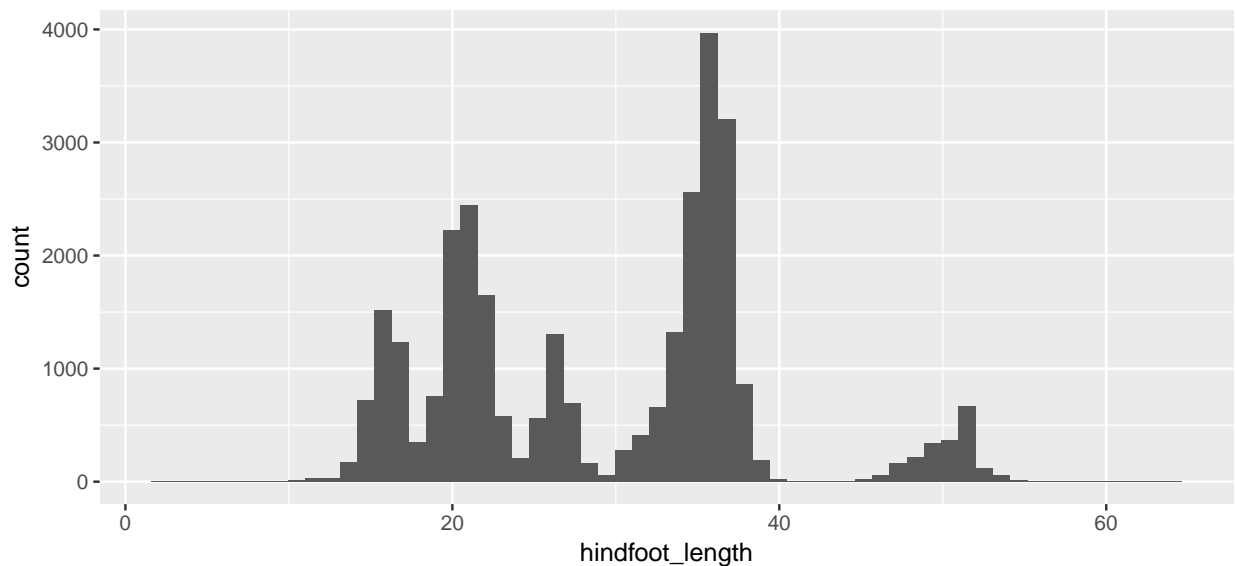
```
surveys_complete <- surveys %>%  
  filter(!is.na(weight),  
         !is.na(hindfoot_length),  
         !is.na(sex))
```

Step-3: Data Transformation

Rare species are omitted from final data set. Hence species frequency is counted first in descending order, rare species are then filtered out (less than 200) and finally most observed species are extracted.

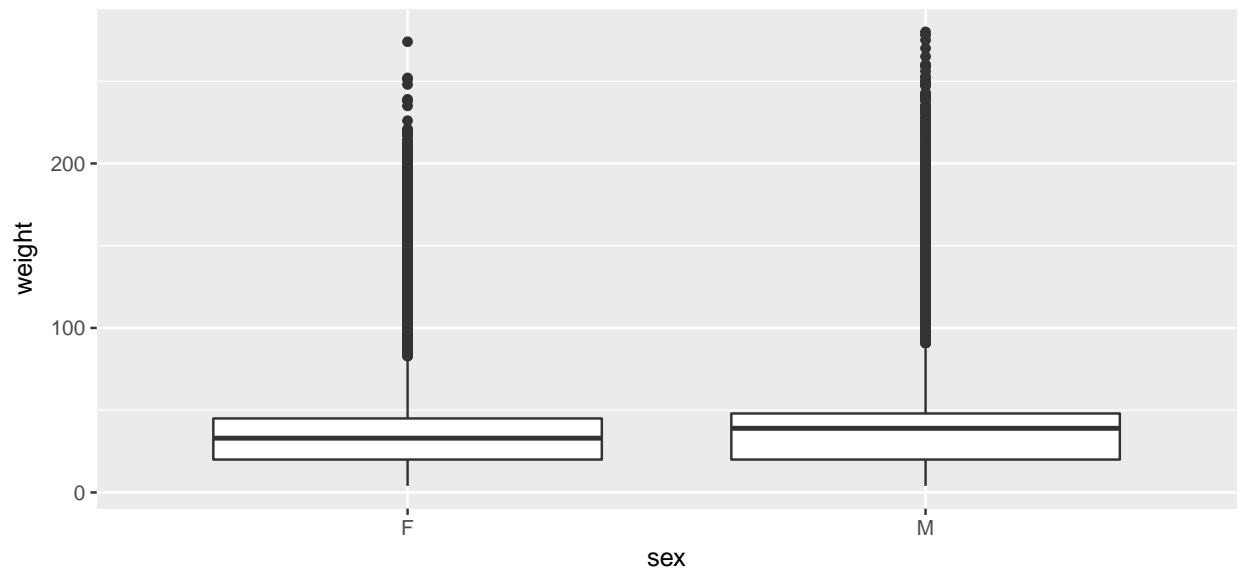
Step-4: Plots and reports

a. Histogram of hindfoot lengths



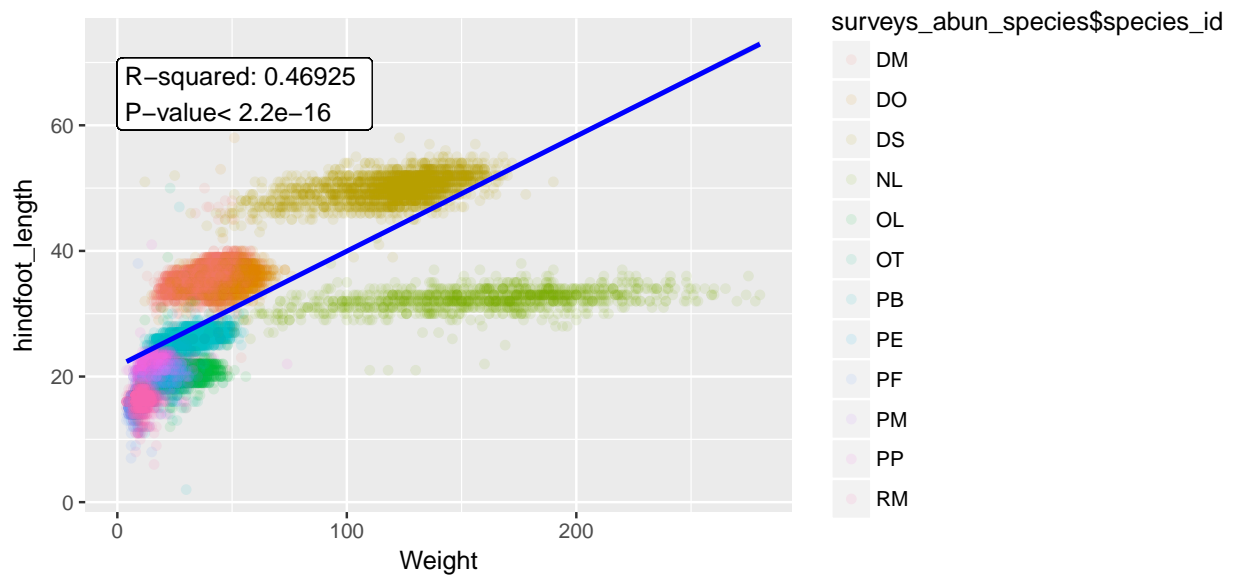
Highest number of species has length 36. Overall data is sparsely distributed.

b. Boxplots: weight by sex



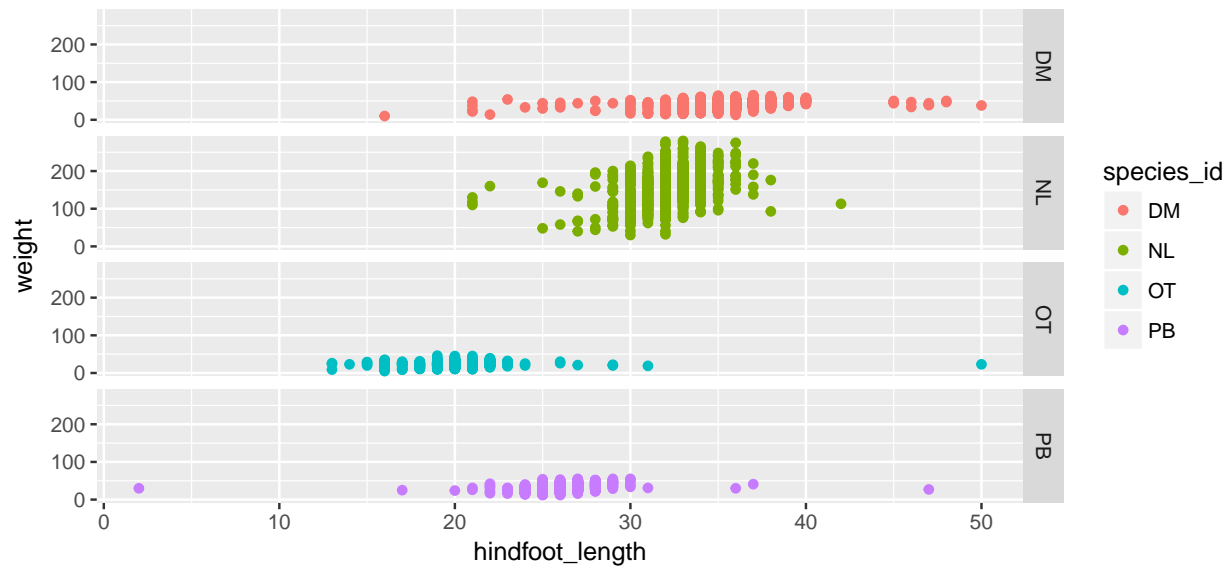
50 % species of both male and female are having weight within 50-25

c. Scatterplots: hindfoot length vs weight



It reflects that data are not that much correlated; low p value rejects the null hypothesis

d. Scatterplots: hindfoot length vs weight for 4 species

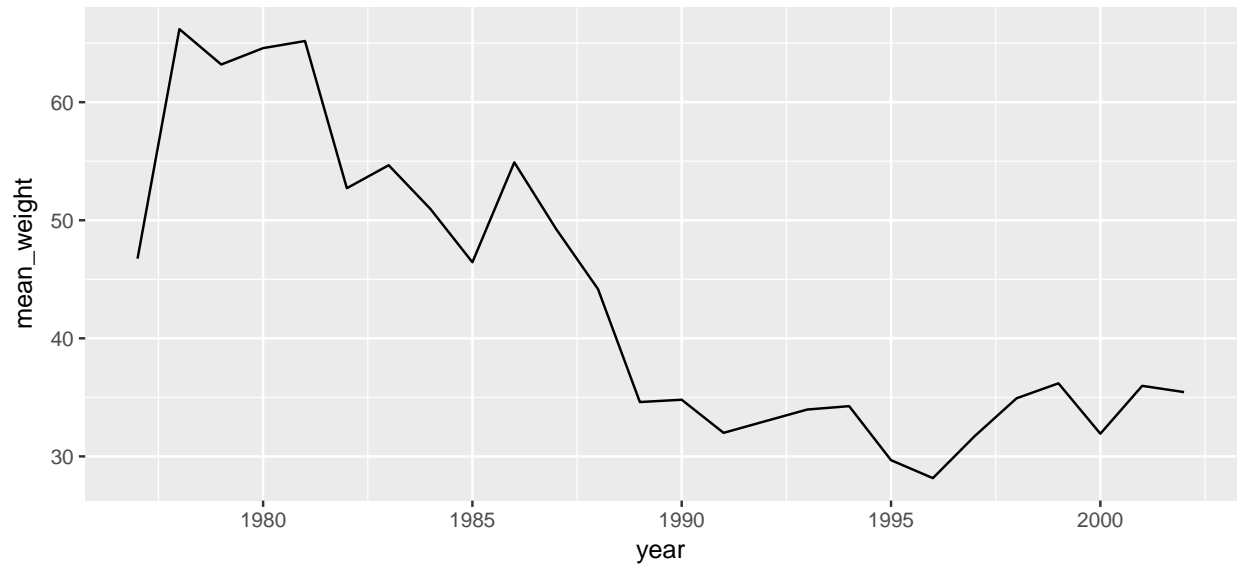


e. Plot: Yearly counts of 'Rodent Exclosure'



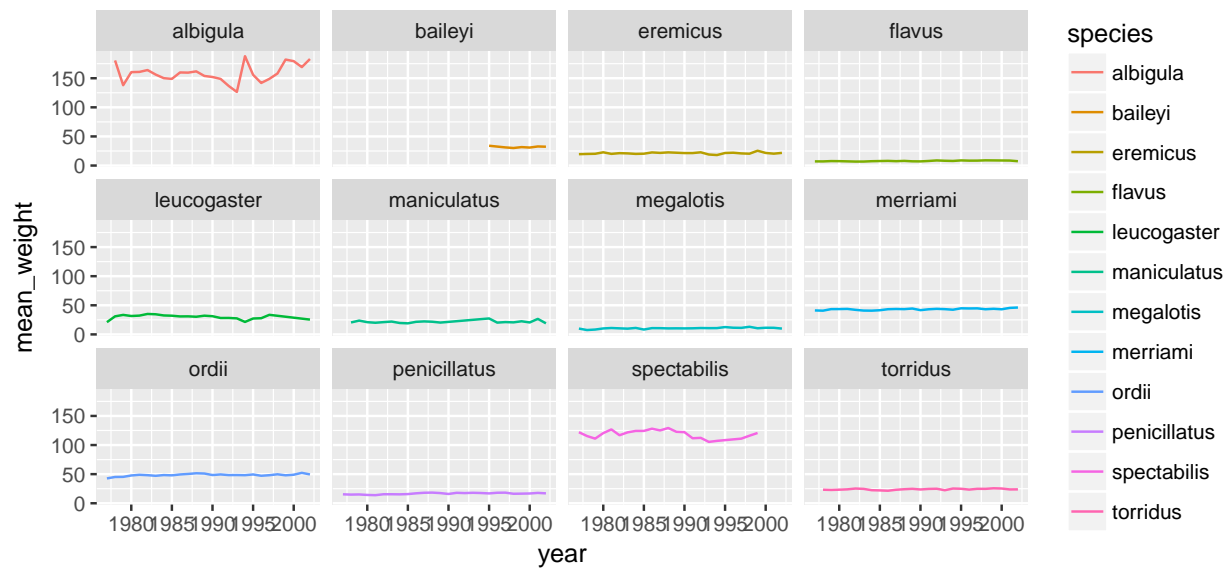
###f. Plot: Average weight (all animals)

It shows decreasing trend

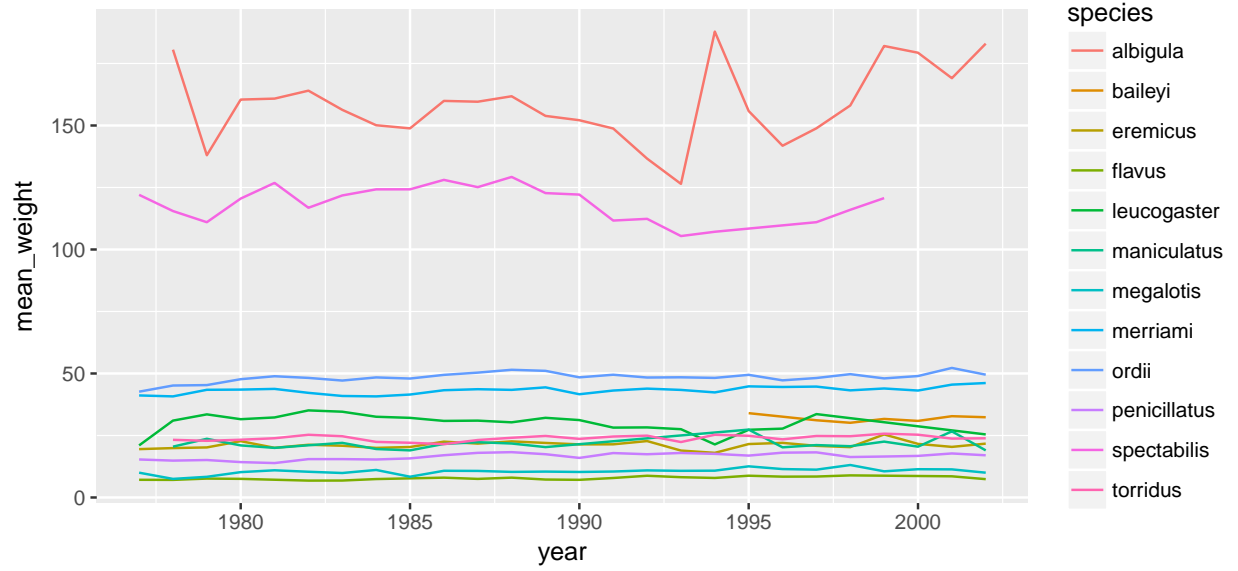


g. Plot: Yearly average weight (each species)

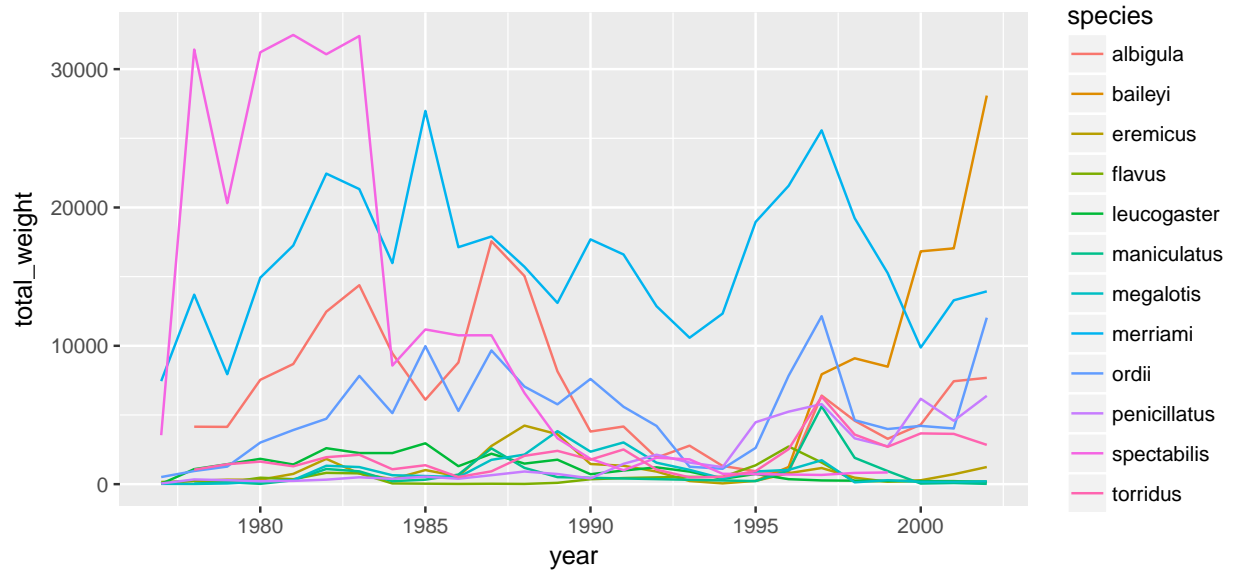
It remains constant



h. Plot: Yearly average weight



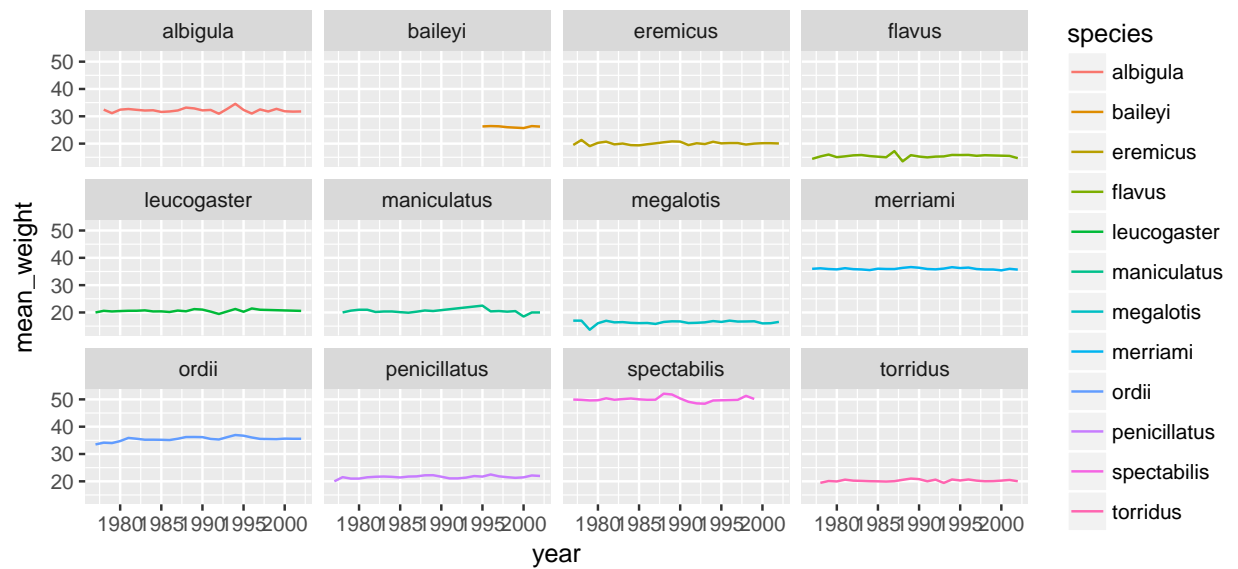
i. Yealy total weight by species



j. hindfoot length(All species) trend



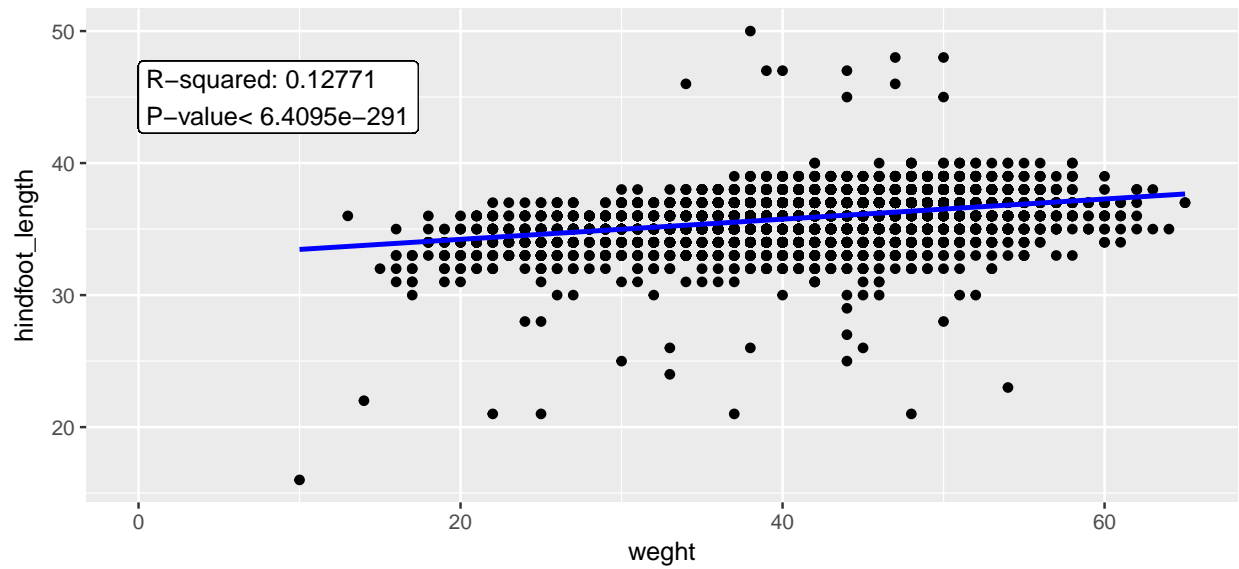
k. hindfoot length(Per species) trend



Plot f-k (Interpretation):

Due to smaller rodents captured, each species has relatively constant mean of hindfoot length and weight

1. hindfoot length vs weight (species id=DM)



Here species *Dipodomys merriami* has less variation and low p value also reject the null hypothesis