Final Exam IF-H

2024-11-25

Final Exam

Instruction

1. Read the instruction carefully and answer accordingly
2. Do not cooperate with your friend
3. The working time is 2 hours (01.00 - 03.00 PM) with additional 10 minutes to submit your work (03.10 PM)
4. Submit the knitted pdf with name **Student ID_Name_Final Exam.pdf**

Import Library (5 points)

Import all the required library.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidymodels)
```

```
## -- Attaching packages --------------------------------------- tidymodels 1.2.0 --
## v broom        1.0.7     v rsample      1.2.1
## v dials        1.3.0     v tune         1.2.1
## v infer        1.0.7     v workflows    1.1.4
## v modeldata    1.4.0     v workflowsets 1.1.0
## v parsnip      1.2.1     v yardstick    1.3.1
## v recipes      1.1.0
## -- Conflicts ------------------------------------------ tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
```

```
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
## * Dig deeper into tidy modeling with R at https://www.tmwr.org
```

```
library(dplyr)
```

Import Dataset (5 points)

Import zoo1.csv and zoo2.csv that are available on SPADA.

```
zoo1 <- read_csv("dataset/zoo1.csv")
```

```
## Rows: 96 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): animal_name
## dbl (9): hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, b...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
zoo2 <- read_csv("dataset/zoo2.csv")
```

```
## Rows: 96 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): animal_name, class_type
## dbl (7): breathes, venomous, fins, legs, tail, domestic, catsize
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Preprocessing Data

Join Table (10 points)

Join the datasets based on the same column.

```
zoo <- inner_join(zoo1, zoo2, by = "animal_name")
```

```
## Warning in inner_join(zoo1, zoo2, by = "animal_name"): Detected an unexpected many-to-many relationsh
## i Row 26 of `x` matches multiple rows in `y`.
## i Row 26 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

Data Imputation (20 points)

The dataset contains N/A value in fins, legs, and tail columns. To fill the empty value, first, create a new dataset that is grouped by the class_type. Don't forget to create new columns to store the rounded mean of the fins, legs, and tail for each group.

```
zoo_grouped_means <- zoo %>%
  group_by(class_type) %>%
  summarise(
    mean_fins = round(mean(fins, na.rm = TRUE)),
    mean_legs = round(mean(legs, na.rm = TRUE)),
    mean_tail = round(mean(tail, na.rm = TRUE))
  )
head(zoo_grouped_means, 5)
```

```
## # A tibble: 5 x 4
##   class_type   mean_fins mean_legs mean_tail
##   <chr>            <dbl>     <dbl>     <dbl>
## 1 Amphibian            0         4         0
## 2 Bird                 0         2         1
## 3 Bug                  0         6         0
## 4 Fish                 1         0         1
## 5 Invertebrate         0         4         0
```

Then, left join the zoo dataset with the grouped dataset. After that, if the value of the column is N/A, the value will be filled with the mean value of the respective class_type, otherwise, the value remains the same.

```
zoo <- zoo %>%
  left_join(zoo_grouped_means, by = "class_type") %>%
  mutate(
    fins = ifelse(is.na(fins), mean_fins, fins),
    legs = ifelse(is.na(legs), mean_legs, legs),
    tail = ifelse(is.na(tail), mean_tail, tail)
  )
head(zoo, 5)
```

```
## # A tibble: 5 x 21
##   animal_name  hair feathers  eggs  milk airborne aquatic predator toothed
##   <chr>       <dbl>    <dbl> <dbl> <dbl>    <dbl>   <dbl>    <dbl>   <dbl>
## 1 aardvark        1        0     0     1        0       0        1       1
## 2 antelope        1        0     0     1        0       0        0       1
## 3 bass            0        0     1     0        0       1        1       1
## 4 bear            1        0     0     1        0       0        1       1
## 5 boar            1        0     0     1        0       0        1       1
## # i 12 more variables: backbone <dbl>, breathes <dbl>, venomous <dbl>,
## #   fins <dbl>, legs <dbl>, tail <dbl>, domestic <dbl>, catsize <dbl>,
## #   class_type <chr>, mean_fins <dbl>, mean_legs <dbl>, mean_tail <dbl>
```

Choosing Columns (10 points)

Discard the mean columns.

```r
zoo <- zoo %>%
  select(-mean_fins, -mean_legs, -mean_tail)

head(zoo, 5)
```

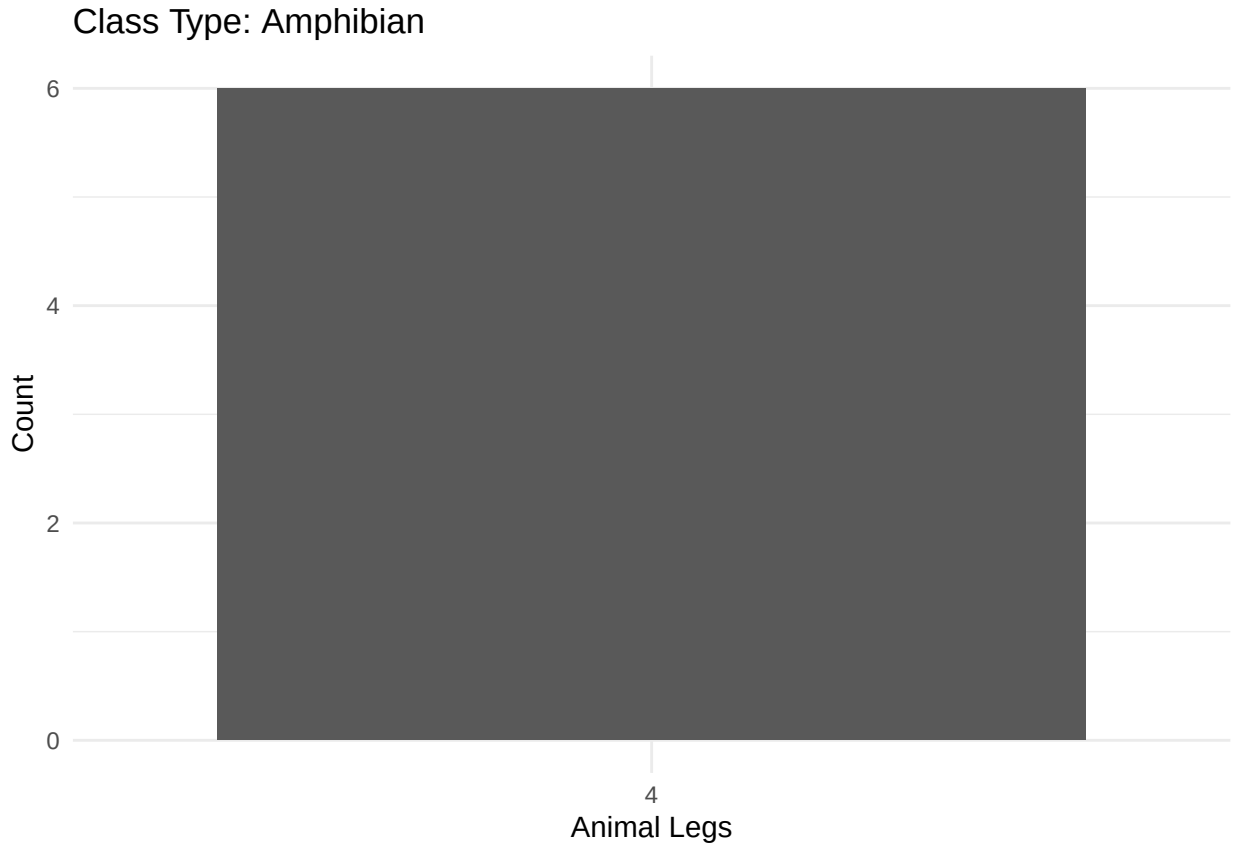
```
## # A tibble: 5 x 18
##   animal_name  hair feathers  eggs  milk airborne aquatic predator toothed
##   <chr>       <dbl>    <dbl> <dbl> <dbl>    <dbl>   <dbl>    <dbl>   <dbl>
## 1 aardvark        1        0     0     1        0       0        1       1
## 2 antelope        1        0     0     1        0       0        0       1
## 3 bass            0        0     1     0        0       1        1       1
## 4 bear            1        0     0     1        0       0        1       1
## 5 boar            1        0     0     1        0       0        1       1
## # i 9 more variables: backbone <dbl>, breathes <dbl>, venomous <dbl>,
## #   fins <dbl>, legs <dbl>, tail <dbl>, domestic <dbl>, catsize <dbl>,
## #   class_type <chr>
```

Data Visualization (12 points)

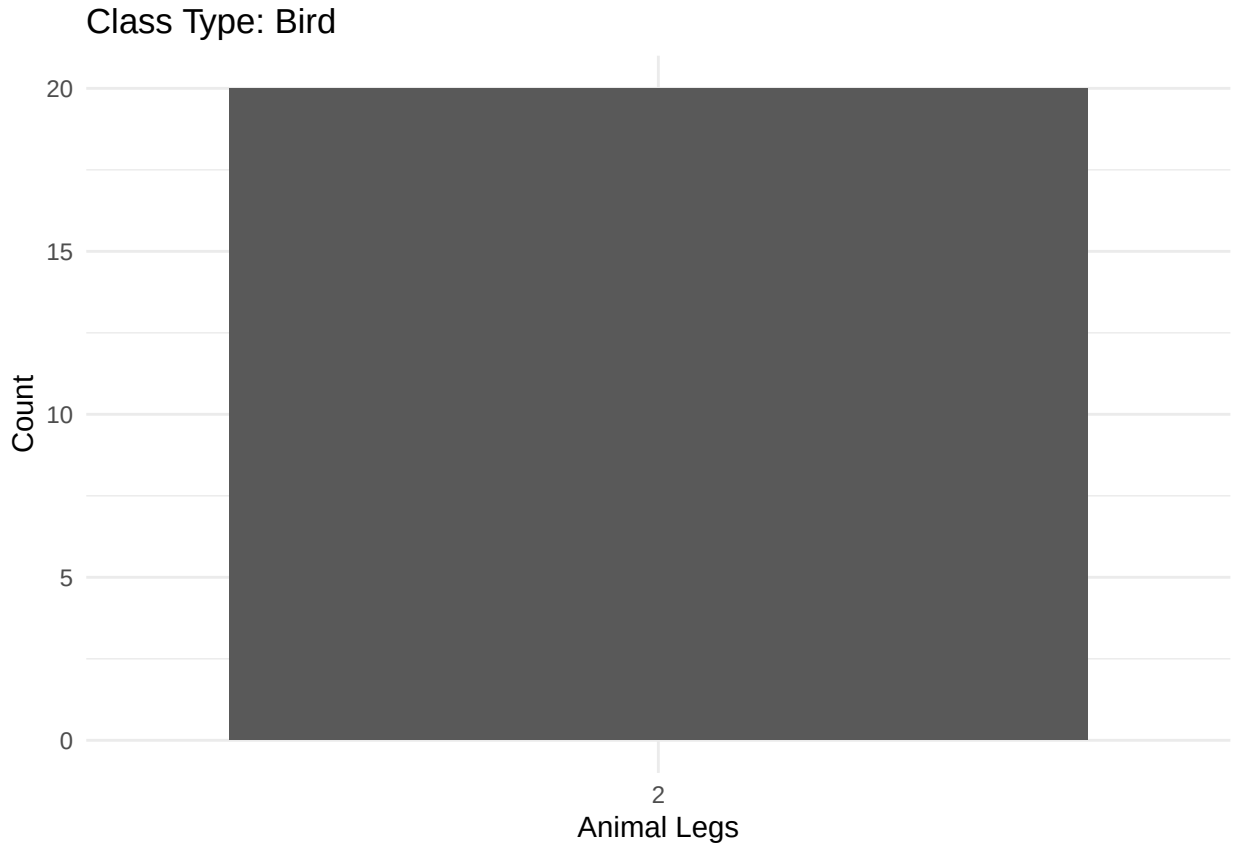
Create bar charts of legs for every class_type Because we have 16 class, there should be 6 charts. Use the function that we learned the other day, not manually. Feel free to use the help menu.

```r
zoo_type <- split(zoo, f = zoo$class_type)

plot_bar_chart <- function(data, class_label) {
  ggplot(data, aes(x = factor(legs))) +
    geom_bar() +
    labs(
      title = paste("Class Type:", class_label),
      x = "Animal Legs",
      y = "Count"
    ) +
    theme_minimal()
}

bar_charts <- lapply(names(zoo_type), function(class_label) {
  plot_bar_chart(zoo_type[[class_label]], class_label)
})

bar_charts[1:6]
```

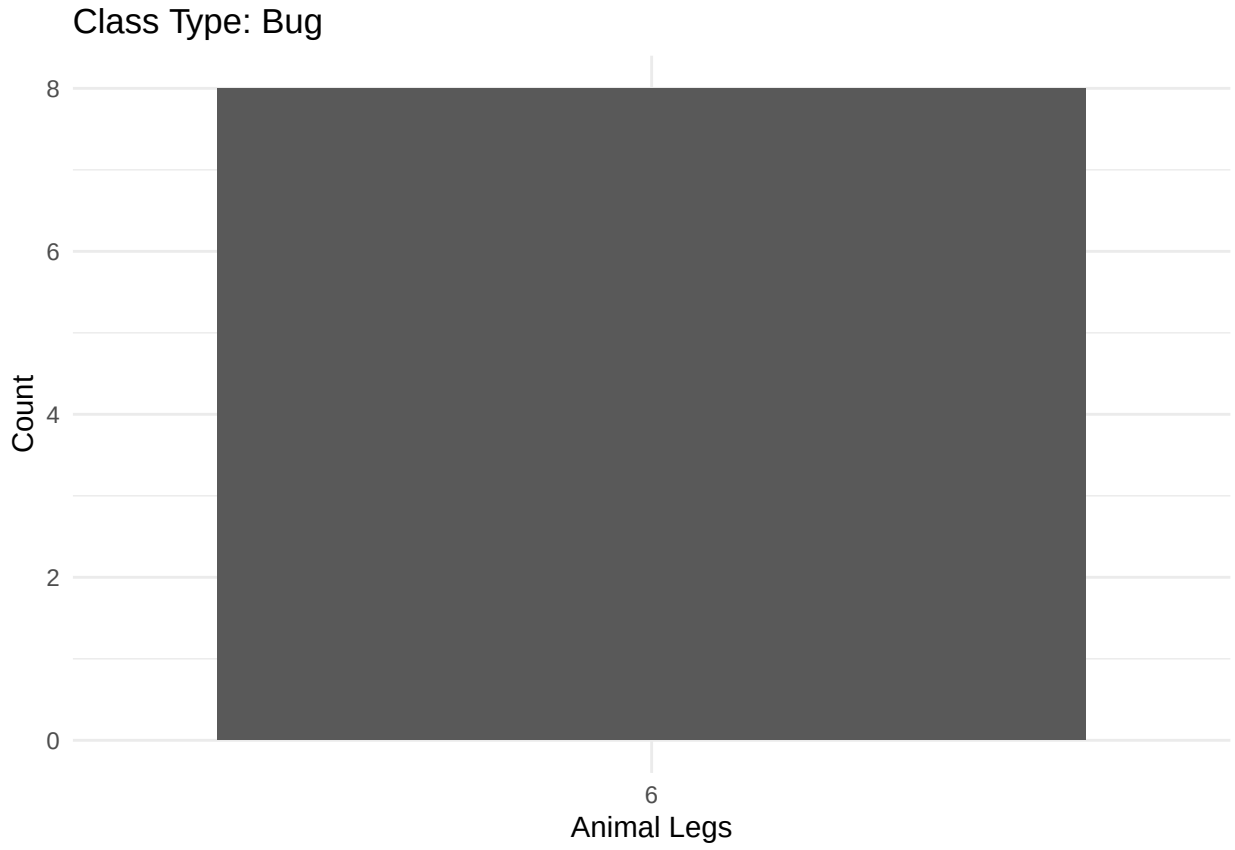
```
## [[1]]
```



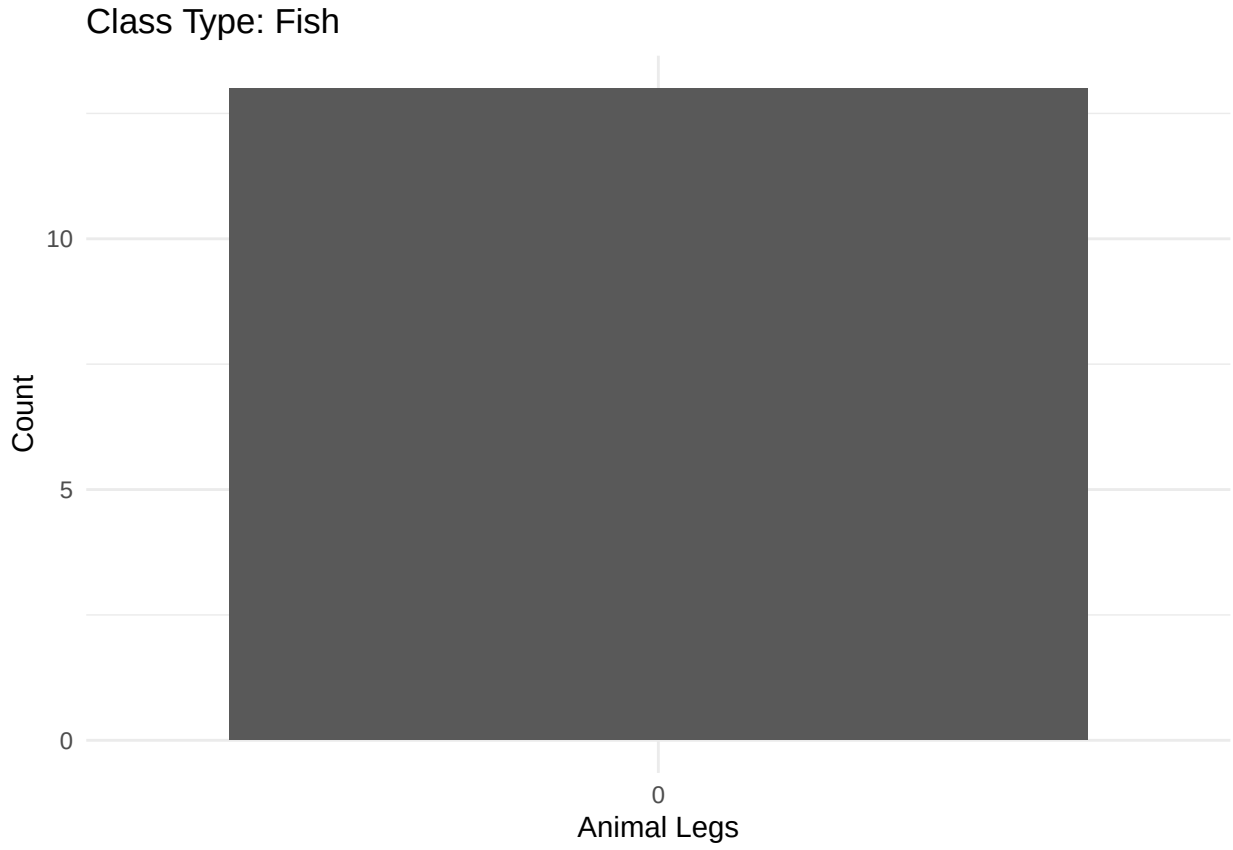

```
##
## [[2]]
```



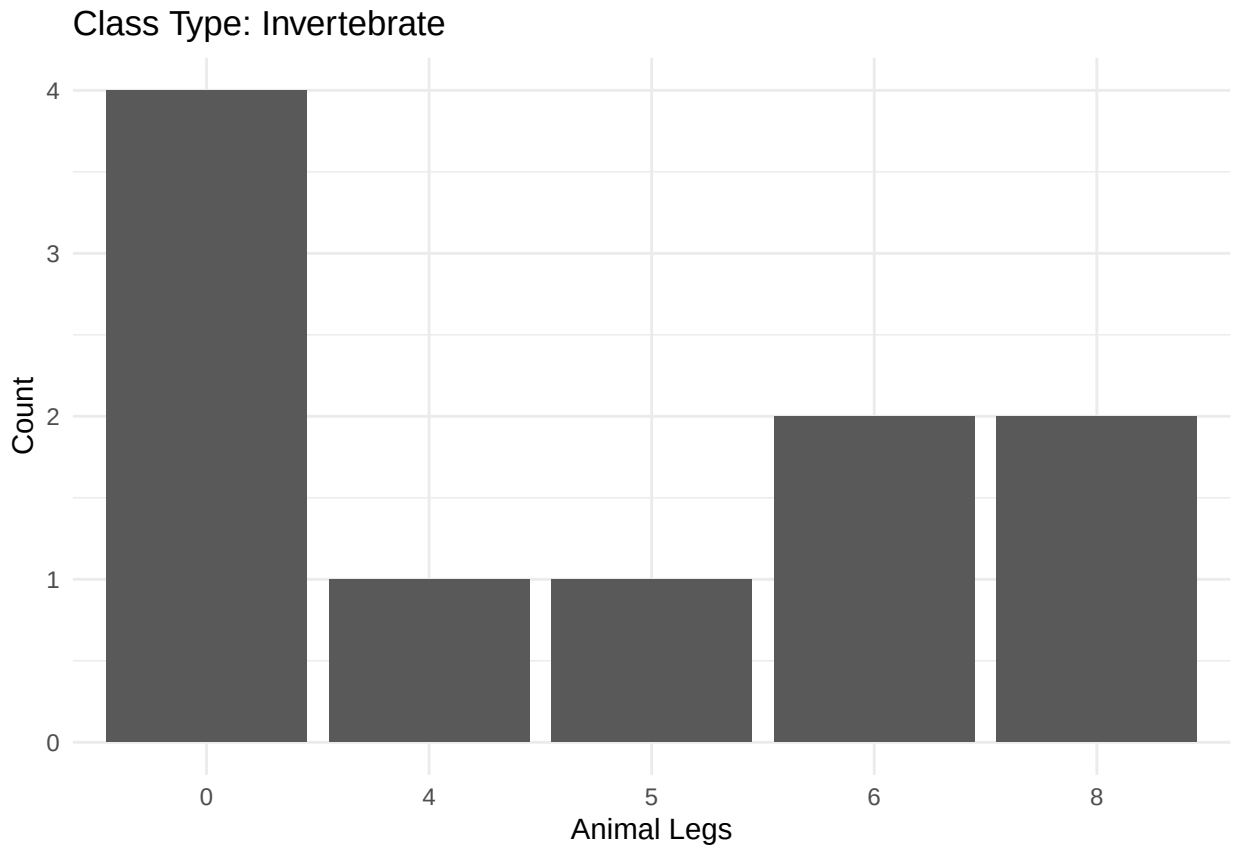

```
##
## [[3]]
```



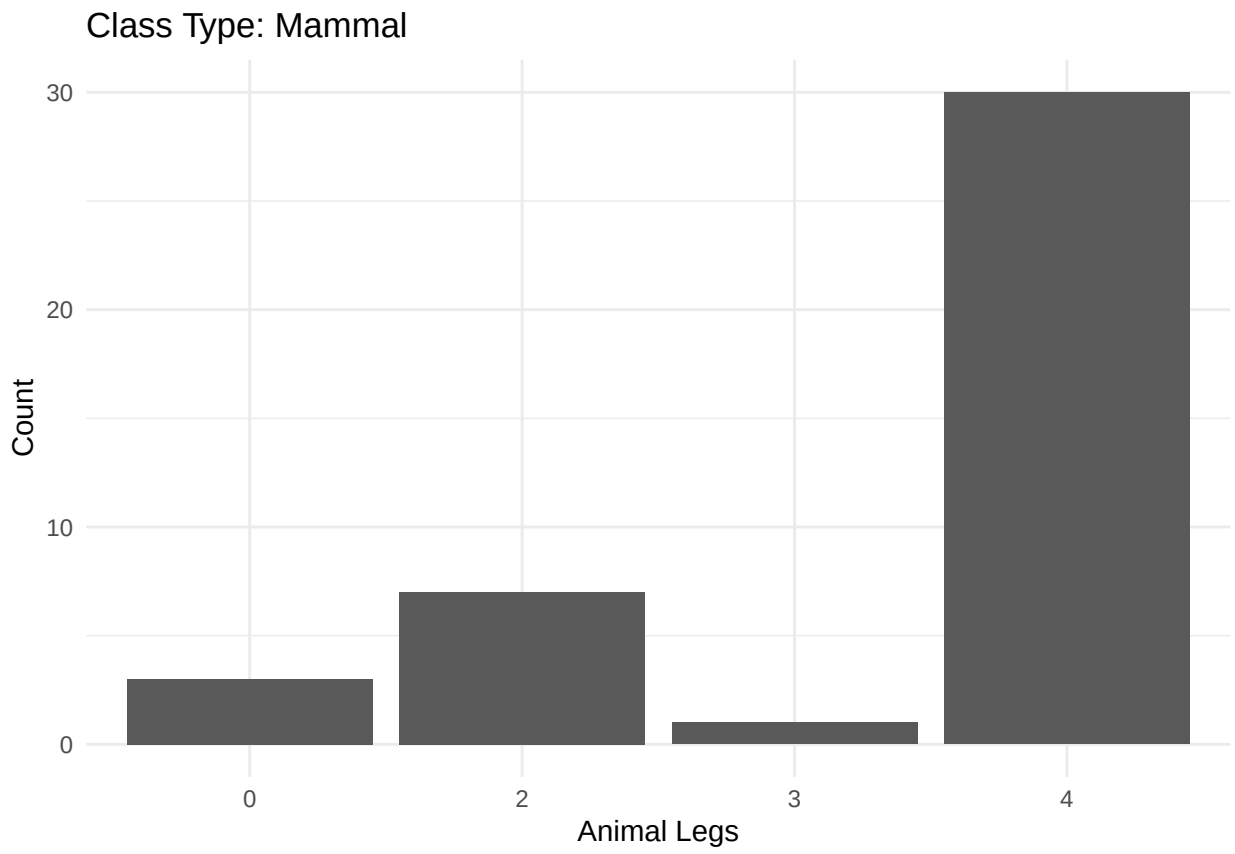

```
##
## [[4]]
```




```
##
## [[5]]
```




```
##
## [[6]]
```



Data Modelling

Data Splitting (5 points)

Split the zoo dataset for training and testing. Set the training ratio to 80% and split evenly based on the class_type. Show the summary of the training and testing data.

```
set.seed(125)
split <- initial_split(zoo, prop = 0.8, strata = class_type)
training <- training(split)
testing <- testing(split)

cat("\nSummary Training \n")
```

```
##
## Summary Training
```

```
summary(training)
```

```
##  animal_name            hair           feathers           eggs
##  Length:77          Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.4675   Mean   :0.1948   Mean   :0.5844
```

```
##                    3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                    Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       milk           airborne         aquatic          predator
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :1.0000
##  Mean   :0.4286   Mean   :0.2727   Mean   :0.3636   Mean   :0.5325
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##     toothed          backbone         breathes         venomous
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:1.0000   1st Qu.:0.00000
##  Median :1.0000   Median :1.0000   Median :1.0000   Median :0.00000
##  Mean   :0.5974   Mean   :0.8052   Mean   :0.8182   Mean   :0.06494
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##       fins             legs             tail           domestic
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :4.000   Median :1.0000   Median :0.0000
##  Mean   :0.1558   Mean   :2.974   Mean   :0.6883   Mean   :0.1429
##  3rd Qu.:0.0000   3rd Qu.:4.000   3rd Qu.:1.0000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :8.000   Max.   :1.0000   Max.   :1.0000
##     catsize        class_type
##  Min.   :0.0000   Length:77
##  1st Qu.:0.0000   Class :character
##  Median :0.0000   Mode  :character
##  Mean   :0.4545
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

```
cat("\nSummary Testing \n")
```

```
##
## Summary Testing
```

```
summary(testing)
```

```
##  animal_name            hair           feathers           eggs
##  Length:21          Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :1.0000
##                     Mean   :0.3333   Mean   :0.2381   Mean   :0.5714
##                     3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
##                     Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       milk           airborne         aquatic          predator
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
##  Median :0.000   Median :0.0000   Median :0.0000   Median :1.000
##  Mean   :0.381   Mean   :0.1429   Mean   :0.4286   Mean   :0.619
##  3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##     toothed          backbone         breathes         venomous
```

```
##  Min.   :0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.000   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :1.000   Median :1.0000   Median :1.0000   Median :0.00000
##  Mean   :0.619   Mean   :0.8571   Mean   :0.7143   Mean   :0.09524
##  3rd Qu.:1.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##       fins            legs            tail          domestic
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:2.000   1st Qu.:1.0000   1st Qu.:0.00000
##  Median :0.0000   Median :2.000   Median :1.0000   Median :0.00000
##  Mean   :0.2381   Mean   :2.714   Mean   :0.8095   Mean   :0.09524
##  3rd Qu.:0.0000   3rd Qu.:4.000   3rd Qu.:1.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :8.000   Max.   :1.0000   Max.   :1.00000
##     catsize       class_type
##  Min.   :0.000   Length:21
##  1st Qu.:0.000   Class :character
##  Median :0.000   Mode  :character
##  Mean   :0.381
##  3rd Qu.:1.000
##  Max.   :1.000
```

Creating Model (13 points)

Create a multinomial classification model.

```
zoo_recipe <- recipe(class_type ~ ., data = training)

multinom_spec <- multinom_reg() %>%
  set_engine("nnet") %>%
  set_mode("classification")

zoo_workflow <- workflow() %>%
  add_recipe(zoo_recipe) %>%
  add_model(multinom_spec)

zoo_model <- zoo_workflow %>%
  fit(data = training)
```

Evaluasi Model

Testing Model (15 points)

Test the model using the testing data.

```
zoo_predictions <- predict(zoo_model, testing, type = "class")

zoo_results <- testing %>%
  bind_cols(predictions = zoo_predictions)

head(zoo_results, 10)
```

```
## # A tibble: 10 x 19
```

```
##    animal_name  hair feathers  eggs  milk airborne aquatic predator toothed
##    <chr>       <dbl>    <dbl> <dbl> <dbl>    <dbl>   <dbl>    <dbl>   <dbl>
##  1 aardvark        1        0     0     1        0       0        1       1
##  2 boar            1        0     0     1        0       0        1       1
##  3 carp            0        0     1     0        0       1        0       1
##  4 catfish         0        0     1     0        0       1        1       1
##  5 chicken         0        1     1     0        1       0        0       0
##  6 crab            0        0     1     0        0       1        1       0
##  7 dolphin         0        0     0     1        0       1        1       1
##  8 frog            0        0     1     0        0       1        1       1
##  9 haddock         0        0     1     0        0       1        0       1
## 10 hare            1        0     0     1        0       0        0       1
## # i 10 more variables: backbone <dbl>, breathes <dbl>, venomous <dbl>,
## #   fins <dbl>, legs <dbl>, tail <dbl>, domestic <dbl>, catsize <dbl>,
## #   class_type <chr>, .pred_class <fct>
```

Confusion Matrix (5 points)

Create a confusion matrix for the test result.

```
# zoo_conf_mat <- conf_mat(zoo_results)
# zoo_conf_mat <- conf_mat(zoo_results, truth = class_type, estimate = predictions$.pred_class)
# tidak bisa
```