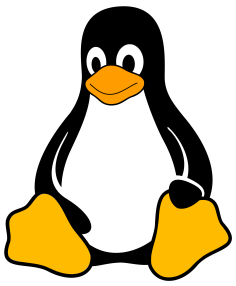

Penguin Prediction

March 2, 2022

The Dataset

Palmer Archipelago (Antarctica) penguin data. Data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network.

Thank you to Dr. Gorman, Palmer Station LTER and the LTER Network! Special thanks to Marty Downs (Director, LTER Network Office) for help regarding the data license & use.

The variables in the dataset are shown below:

- species: penguin species (Chinstrap, Adélie, or Gentoo)
- culmen_length_mm: culmen length (mm)
- culmen_depth_mm: culmen depth (mm)
- flipper_length_mm: flipper length (mm)
- body_mass_g: body mass (g)
- island: island name (Dream, Torgersen, or Biscoe) in the Palmer Archipelago (Antarctica)
- sex: penguin sex

What are culmen length & depth? The culmen is "the upper ridge of a bird's beak" (definition from Oxford Languages). For this penguin data, the culmen length and culmen depth are measured as shown below (thanks Kristen Gorman for clarifying!):

1. Visualize and describe in detail the distribution for female adilie penguins. What would you describe as a 'typical' weight?
2. Create a pairwise boxplot for the body mass of penguins across the different species. Colour fill each category by penguin gender. Describe in detail what insights may be drawn from the plot.
3. Create a pairwise boxplot of the culmen length for penguins across the different islands. Colour fill each category by species. Describe in detail what insights may be drawn from the plot.
4. Create a scatter plot for the body mass versus flipper length for adelie penguins in Torgersen. Size the points based on culmen length.
   (a) If you were told that a adelie penguin in Torgersen had a flipper length of 180 mm use your plot to predict the body mass of the penguin.
   (b) Color the plot you made in the first part of this problem by Gender. How would knowing the gender of the penguin to be 'Male' change your answer to part (a)?

5. Create a cool visualizing of your choosing using the dataset and describe in detail what insights you were able to draw from it.

Solution:

```
### Load Packages
library(ggplot2)
library(tidyverse)

### Load Dataset
pen <- read.csv('/data/datasets/penguin.csv')

### Question 1: Distribution for penguin body mass of female adelie penguins
pen %>% filter(Gender=='FEMALE' & Species == 'Adelie Penguin (Pygoscelis adeliae)') %>%
  ggplot(aes(x=Body_mass_g)) + geom_histogram()+
  theme_classic()+
  labs(x='body mass of penguins in grams', title='Distribution for penguin weights')

### Question 2: Pairwise boxplot of penguins by species, filled by color
pen %>% ggplot(aes(x=Species, y=Body_mass_g, fill=Gender)) +
  geom_boxplot() +
  theme_classic() +
  labs(x='Species', y='Body mass in grams', title='Body mass of species across species
    and gender')


### Question 3: Pariwise Boxplot of penguin Clumen Length across the different Island
    locations
#               filled by species
pen %>% ggplot(aes(x=Island, y=Culmen_Length_mm, fill=Species)) +
  geom_boxplot() +
  theme_classic() +
  labs(x='Species', y='Body mass in grams', title='Body mass of species across species
    and gender')


### Question 4: Predicting Body mass with flipper length for specific type and island
pen %>% filter(Island == 'Torgersen' & Species == "Adelie Penguin (Pygoscelis adeliae)")
    %>%
  ggplot(aes(x=Flipper_Length_mm, y=Body_mass_g,size = Culmen_Length_mm)) +
  geom_point(alpha=0.7) +
  theme_classic() +
  geom_smooth(method='lm')+
  labs(x='Flipper length in mm', y='Body mass in g', title="Body mass vs flipper length
    for Adelie Penguin's on the island of Torgersen")


### Question 5: Predicting Body mass with flipper length for specific type and island
    color by gender
#               does gender account for any variation?
pen %>% filter(Island == 'Torgersen' & Species == "Adelie Penguin (Pygoscelis adeliae)")
    %>%
  ggplot(aes(x=Flipper_Length_mm, y=Body_mass_g,size = Culmen_Length_mm, col=Gender)) +
  geom_point(alpha=0.7) +
  theme_classic() +
  labs(x='Flipper length in mm', y='Body mass in g', title="Body mass vs flipper length
    for Adelie Penguin's on the island of Torgersen")

### Question 6: Make a cool plot of your own that conveys some sort of message
```