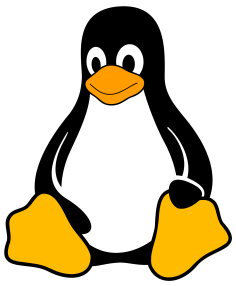

# Penguin Prediction

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 pairwise boxplots for flipper length vs. Gender. Would this plot be inappropriate for any reason?
3. Create pairwise boxplots for flipper length vs. gender for the adelie species of penguin. Describe what insights can be drawn from the plot.
4. Create a pairwise boxplots for the body mass of penguins across the various species.
   (a) Which species appears to be the heaviest?
   (b) If I were to tell you that I selected a random 'gentoo' penguin, what would you predict it's body mass to be.
   (c) Color each species in your plot by gender. If I were to tell you that the penguin described in part (b) was female, would this change your prediction?
5. Create a custom plot of your own using the dataset and describe in details what insights you can draw from it.

**Solution:**

```r
### 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 for flipper length across gender. Inappropriate?
pen %>% ggplot(aes(y=Flipper_Length_mm, fill=Gender))+geom_boxplot()

### Question 3 Pairwise boxplot for flipper length across gender for adelie penguins
pen %>% filter(Species=='Adelie Penguin (Pygoscelis adeliae)') %>%
  ggplot(aes(y=Flipper_Length_mm, fill=Gender))+geom_boxplot()+
  theme_classic()

### Question 4: Pairwise boxplot for bodymass across species, which is heaviest?
pen %>% ggplot(aes(y=Body_mass_g, fill=Species))+geom_boxplot()

### Question 5: Pairwise boxplot for bodymass across species and gender
pen %>% ggplot(aes(y=Body_mass_g,x=Species, fill=Gender))+geom_boxplot()
```