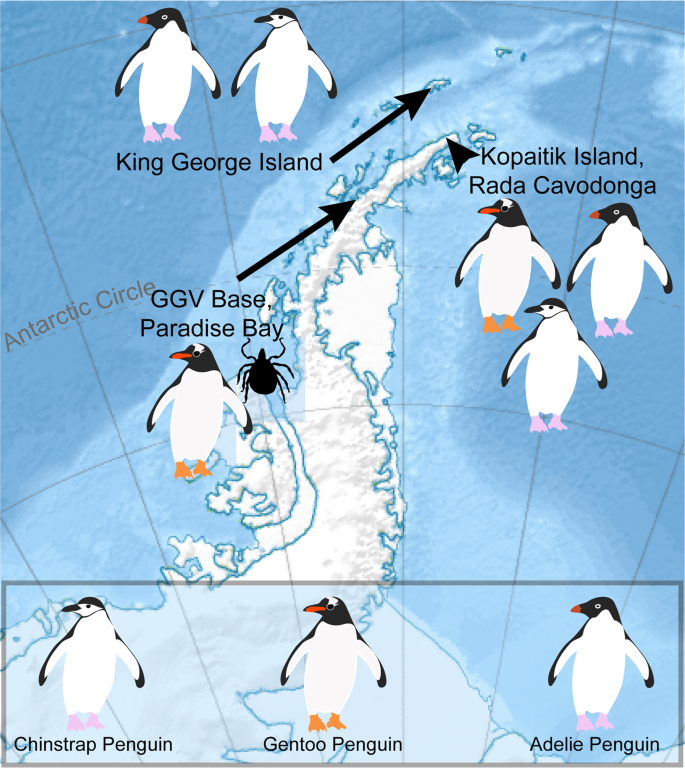
Your boss is a lover of Penguins and she sent you data that were collected by [Dr. Kristen Gorman](https://www.uaf.edu/cfos/people/faculty/detail/kristen-gorman.php) and the [Palmer Station, Antarctica LTER](https://pal.lternet.edu/), a member of the [Long Term Ecological Research Network](https://lternet.edu/) for analysis using R programming.

**Penguins Pictures**



**Source**: Wikipedia, developed by user Kikos

**Load and Display Data**

Read the file penguins.csv and store the result in a dataframe called penguin

**Data Cleaning**

Examine missing values (NA) in this data. How many are they? (Hint: You may use sum(is.na(dataframe\_name)) or inspectdf::inspect\_na(dataframe\_name)

**Derived Variable**

Assuming body\_mass\_g and flipper\_length\_mm are the weight and height of Penguins. Use mutate() function in the dplyr package to calculate the BMI for each Penguins. Please save your resulting dataframe as penguins\_bmi.

**Analyze Data**

* Create descriptive statistics for all variables inpenguins\_bmi dataframe
* Find and display the number of each Penguin specie in the penguins\_bmi dataframe (hint: you can use count() function in dplyr to do this)
* Create and display a dataframe that shows the number of penguins of each species on each island (hint: you can use count() function to do this or a combination of group\_by() and summarize(n = n()).

**Data Visualization**

You are going to use ggplot2 package to plot some data in this section. For example, let’s say I have a dataframe called mpg (the mpg dataframe is available when you load tidyverse or ggplot2 package). If I want to plot a scatter diagram where displ variable will be in the x-axis and hwy variable will be in the y-axis I will use:

**ggplot(mpg, aes(x = displ, y = hwy)) +**

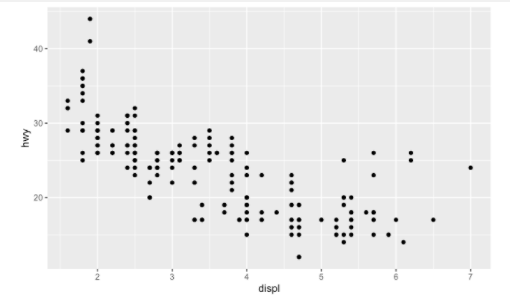
**geom\_point()**

or with pipe:

**mpg %>% ggplot(aes(x = displ, y = hwy)) +**

**geom\_point()**

The result of this is shown below:

****

You can read more on data visualization with ggplot2 via<https://r4ds.had.co.nz/data-visualisation.html> and <https://r4ds.had.co.nz/graphics-for-communication.html>

**Your Task:**

What is the relationship between body mass and flipper length? You can answer this question by using a scatter plot. Please use penguins\_bmi dataframe for this task.