Homework 1

Harsh Mehta (Student Number: 24208383)

Penguins Dataset

The **penguins dataset** contains data on **344 adult penguins** across **three species** (Adelie, Chinstrap, and Gentoo) found on **three islands** (Biscoe, Dream, and Torgersen) in the Palmer Archipelago, Antarctica. The data includes information on their **physical characteristics** (bill length, bill depth, flipper length, and body mass), **sex** (female or male), and **year** (2007, 2008, or 2009).

This dataset is a cleaned subset of the more detailed penguins_raw dataset, which also includes nesting behavior and isotope data.

```
  species    island bill_len bill_dep flipper_len body_mass    sex year
1  Adelie Torgersen     39.1     18.7         181      3750   male 2007
2  Adelie Torgersen     39.5     17.4         186      3800 female 2007
3  Adelie Torgersen     40.3     18.0         195      3250 female 2007
4  Adelie Torgersen       NA       NA          NA        NA   <NA> 2007
5  Adelie Torgersen     36.7     19.3         193      3450 female 2007
6  Adelie Torgersen     39.3     20.6         190      3650   male 2007
```
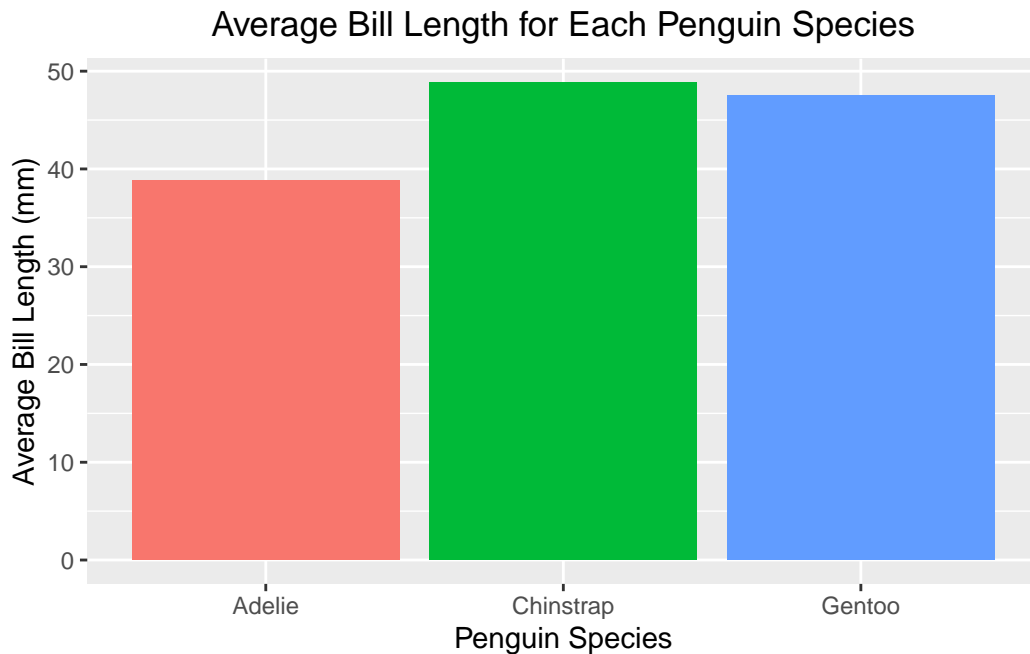
The dataset includes **8 variables**:

- ***species (factor)***: Penguin species (Adelie, Chinstrap, and Gentoo)
- ***island (factor)***: Island where the penguin was observed (Biscoe, Dream, and Torgersen)
- ***bill_len (numeric)***: Bill length in millimeters
- ***bill_depth (numeric)***: Bill depth in millimeters
- ***flipper_len (integer)***: Flipper length in millimeters
- ***body_mass (integer)***: Body mass in grams
- ***sex (facotr)***: Sex of the penguin (female or male)

- *__year (integer)__*: Year the observation was recorded (2007, 2008, or 2009)

Here, we loaded the **palmerpenguins library** since it has the dataset of penguin species from the Palmer Archipelago. Then we loaded the **dplyr library** for data manipulation and transformation. After that we loaded the **ggplot2 library** for data visualization.

Now, we will create a barplot to visualize the average bill length for each penguin species. For that, we will **first clean the dataset** by **removing entries with missing values** for **bill length and species**. Then, we will **calculate the average bill length** for each penguin species, and then we will visualize these averages using a bar plot to clearly compare bill lengths across different species.

Average Bill Length for Each Penguin Species

__The following can be interpreted from the above plot:__

- **Chinstrap penguins** have the **longest average bill length**, measuring just under 50 mm.

- **Gentoo penguins** follow closely, with an average slightly lower than Chinstrap.

- **Adelie penguins** have the **shortest average bill length**, at around 39 mm.

This suggests that **bill length varies significantly between species**, potentially reflecting differences in **feeding habits**, **ecological adaptation**, or **evolutionary traits**.