# R Markdown on Palmer Penguins

Ayushi

2023-02-06

## Introduction to the data Palmer Penguin


Palmer Penguins is a built in dataset provided by R Studio to make data exploration and visualization easier. The data is about Antarctica's three penguin species namely :

- Adelie
- Gentoo
- Chinstrap


**NOTE : Firstly we need to install and load the library for 'palmer penguins' . Various packages such as ggplot2 , tidyverse and dyplyr are required to start with the analysis and data visualization.**

### Data representation , first few rows

```
library(knitr)
library(palmerpenguins)
knitr::kable(head(penguins))
```

| species | island | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g | sex | year |
|---|---|---|---|---|---|---|---|
| Adelie | Torgersen | 39.1 | 18.7 | 181 | 3750 | male | 2007 |
| Adelie | Torgersen | 39.5 | 17.4 | 186 | 3800 | female | 2007 |
| Adelie | Torgersen | 40.3 | 18.0 | 195 | 3250 | female | 2007 |
| Adelie | Torgersen | NA | NA | NA | NA | NA | 2007 |

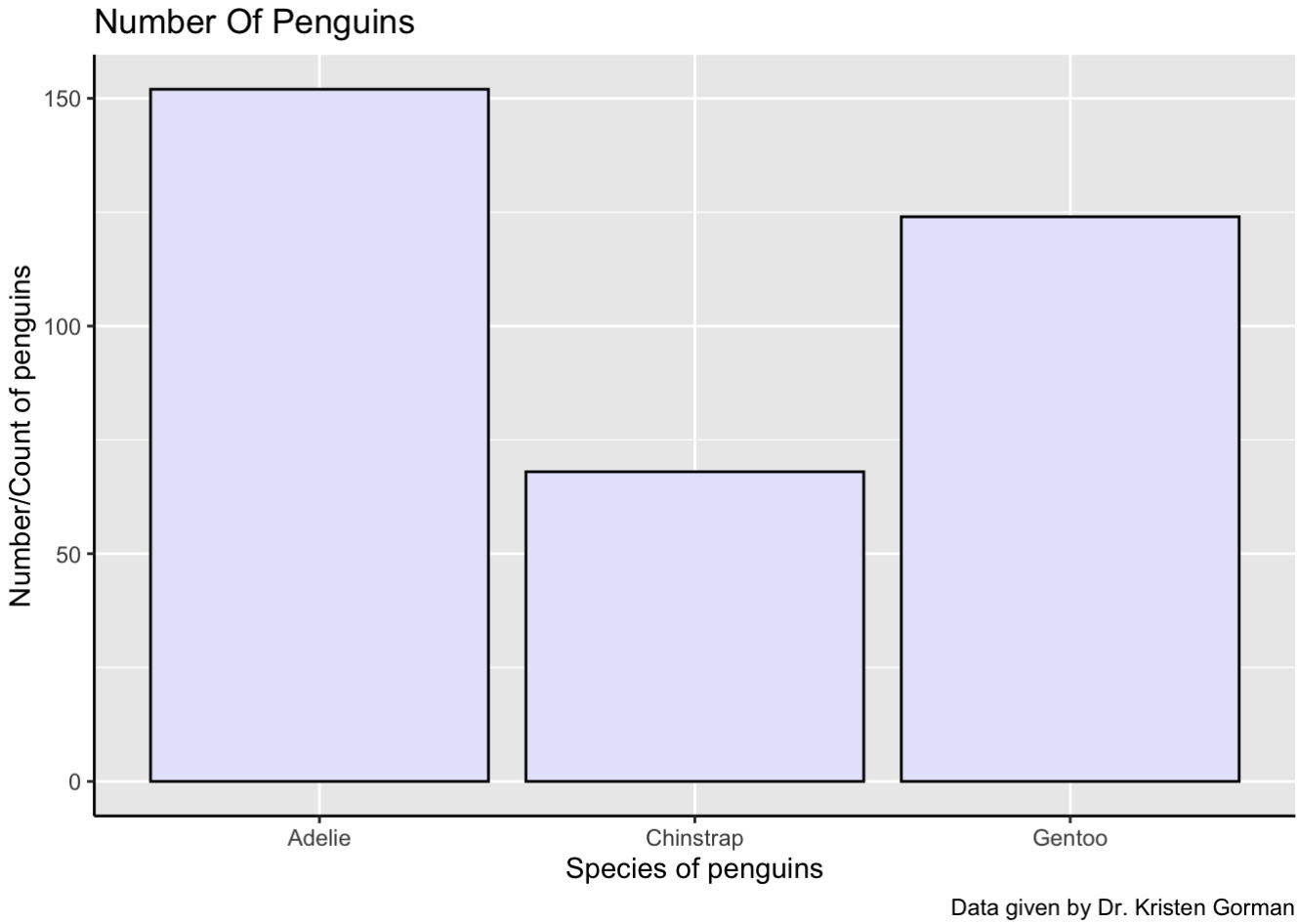
| species | island | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g | sex | year |
|---|---|---|---|---|---|---|---|
| Adelie | Torgersen | 36.7 | 19.3 | 193 | 3450 | female | 2007 |
| Adelie | Torgersen | 39.3 | 20.6 | 190 | 3650 | male | 2007 |

### Objectives of this Markdown includes :

- To find which species has the highest number of penguins and which species has the lowest.
- To know how many males and females are there in each species
- To know which species are the largest by comparing *flipper_length_mm* and *body_mass_g*. (NOTE that here mm and g represents milimetre and gram respectively)
- To find out which species lives in which island.
- To find out which penguin specie has the :
    - longest bill length

# Objective 1 - To find which species has the highest number of penguins and which species has the lowest

```
library(ggplot2)
ggplot(penguins)+
  geom_bar(mapping = aes(x = species),color = "black",fill = "lavender")+
  theme_gray()+
  labs(x= "Species of penguins", y ="Number/Count of penguins", title = "Number Of Pe
nguins",caption = "Data given by Dr. Kristen Gorman")+
  theme(axis.line = element_line("black"))
```



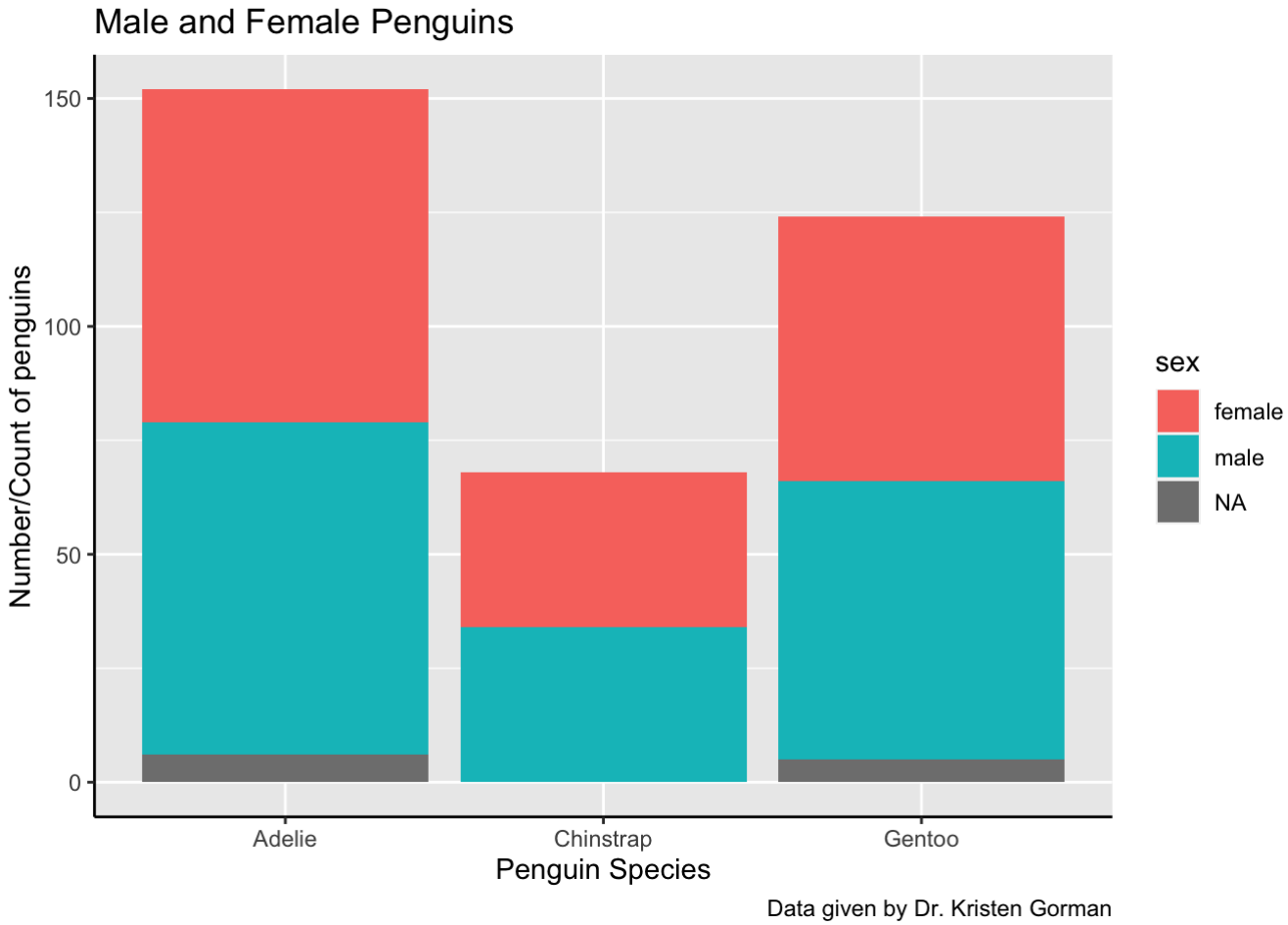

### Description of the Objective 1

Through the above **Bar Graph** we can conclude :

- **Adelie** species has the most number of penguins.
- **Chinstrap** species has the least number of penguins.

# Objective 2 - To know how many males and females are there in each specie

```
ggplot(penguins)+
  geom_bar(mapping = aes(x = species, fill = sex))+
  labs(x = "Penguin Species", y = "Number/Count of penguins", title = "Male and Femal
e Penguins", caption = "Data given by Dr. Kristen Gorman")+
  theme(axis.line = element_line("black"))
```




## Description of the Objective 2

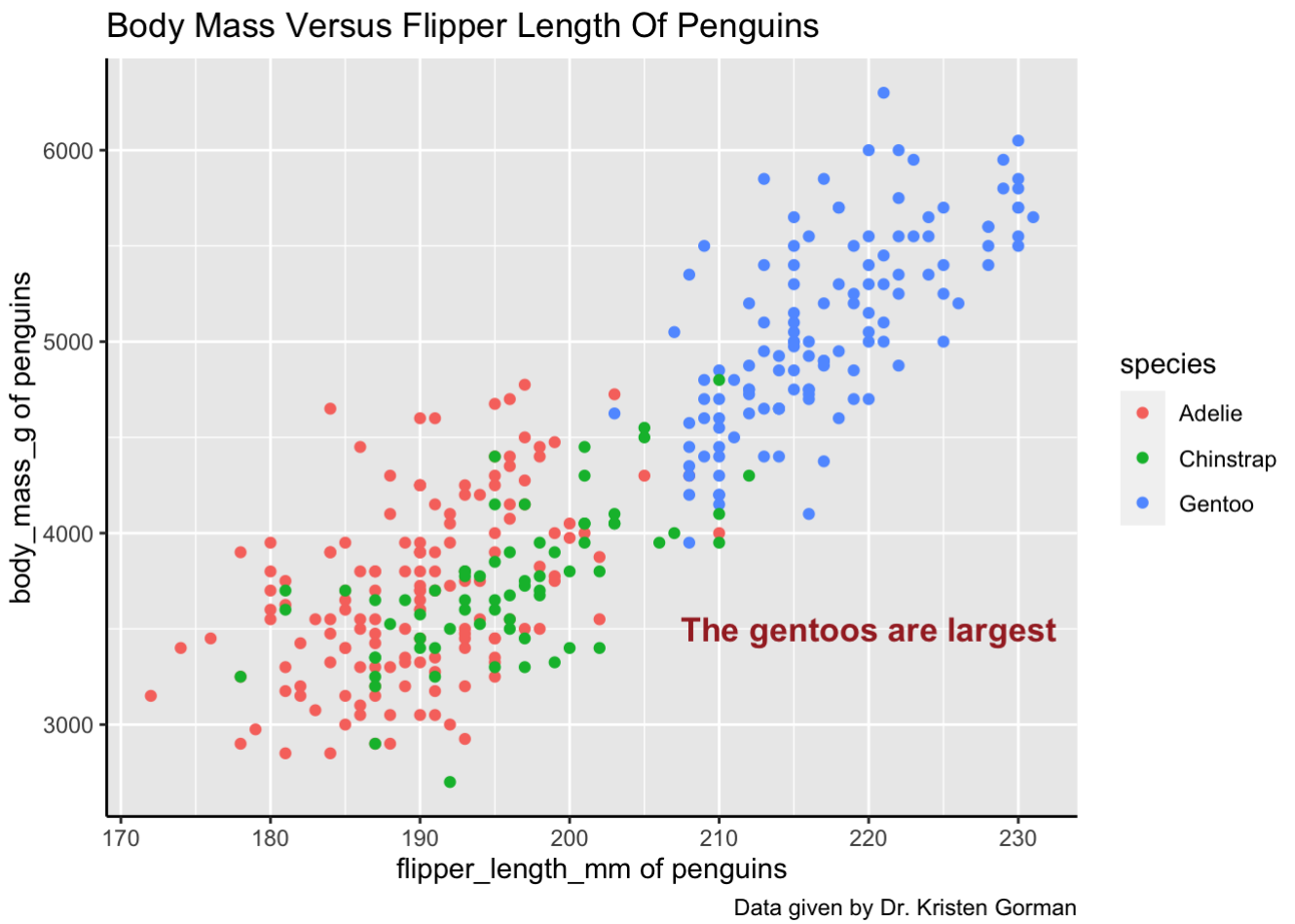
Through the above **Stacked Bar Graph** we can conclude :

- Chinstrap specie has roughly equal number of male and female penguins.
- In penguin specie Adelie and Gentoo we cannot say about number of male and female penguins as we have some blank or 'NA' values.

# Objective 3 - To know which species are the largest by comparing *flipper_length_mm* and

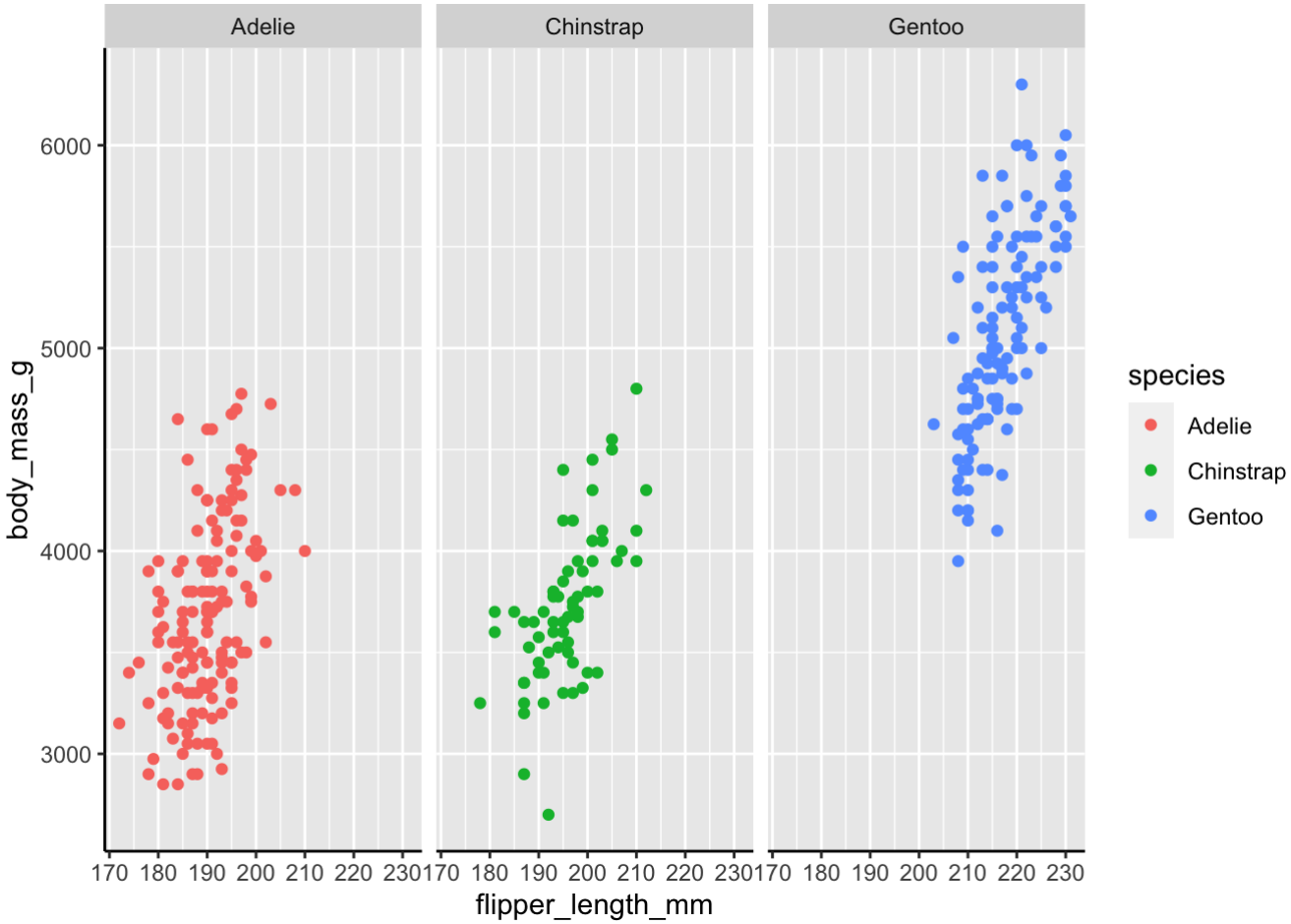
# *body_mass_g.*

```
ggplot(data = penguins)+
 geom_point(mapping = aes(x = flipper_length_mm, y= body_mass_g,color = species))+
 theme_gray()+
 labs(x = "flipper_length_mm of penguins", y = "body_mass_g of penguins", title = "Bo
dy Mass Versus Flipper Length Of Penguins",caption = "Data given by Dr. Kristen Gorma
n")+
  annotate("text",x=220,y = 3500, label = "The gentoos are largest",color = "brown",
fontface = "bold",size = 4.5)+
 theme(axis.line = element_line("black"))
```




## We can view this individually also

```
ggplot(data = penguins)+
  geom_point(mapping = aes(x = flipper_length_mm, y = body_mass_g, color = species))+
  facet_wrap(~species)+
  theme_gray()+
  theme(axis.line = element_line("black"))
```

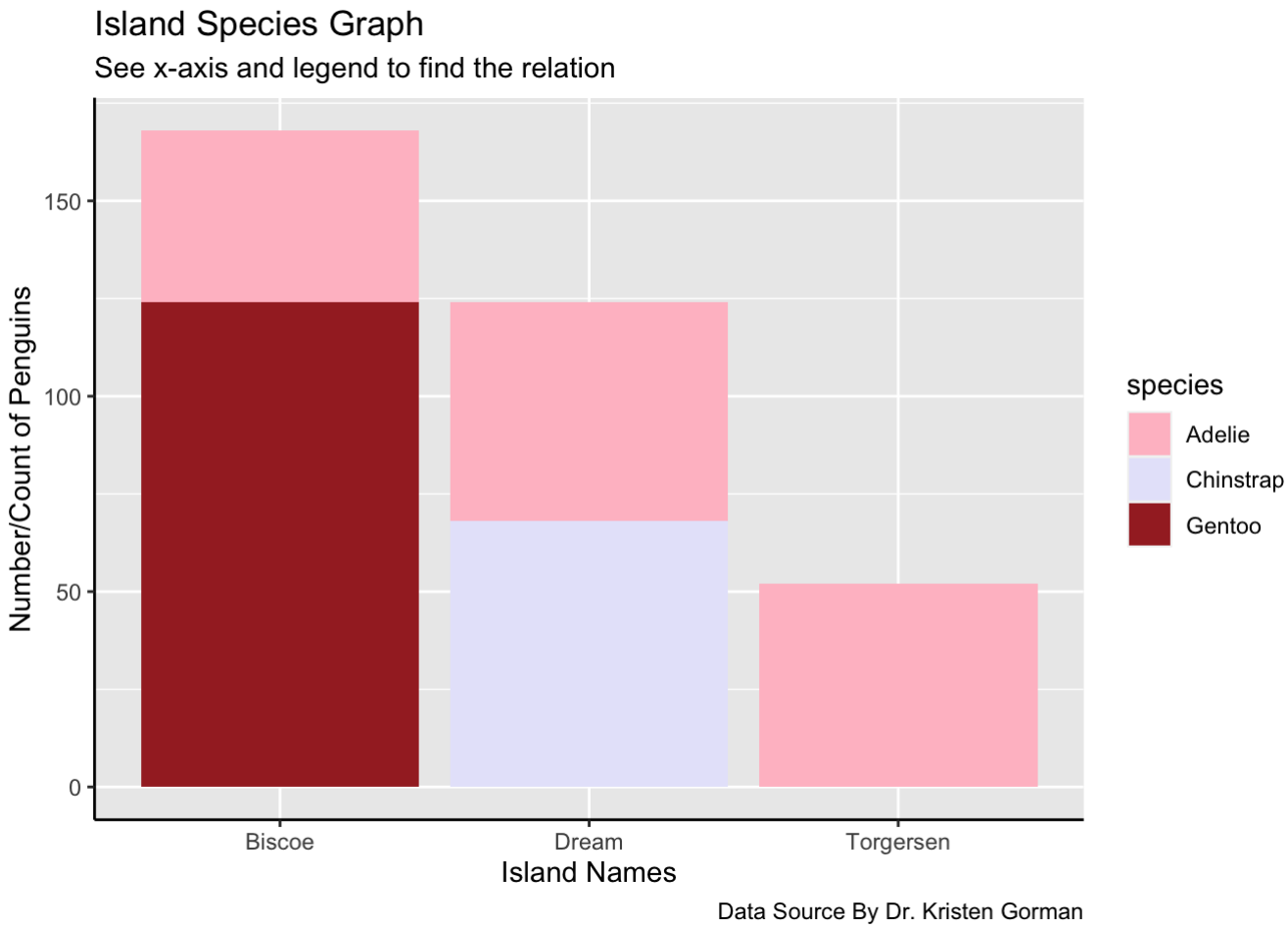



### Description of the Objective 3

- From the above **Scattered Plot**, it can be concluded that Gentoos specie is the largest.
- The **longest flipper length** can be concluded as **231 mm** .
- The **heaviest body mass** can be concluded somewhere above **6000 g**.

# Objective 4 - To find out which species lives in which island.

```
ggplot(data = penguins)+
  geom_bar(mapping = aes(x = island , fill = species))+
   scale_fill_manual(values = c("pink","lavender","brown"))+
  labs(x = "Island Names", y = "Number/Count of Penguins", title = "Island Species Gr
aph", subtitle = "See x-axis and legend to find the relation",caption = "Data Source
By Dr. Kristen Gorman")+
  theme(axis.line = element_line("black"))
```



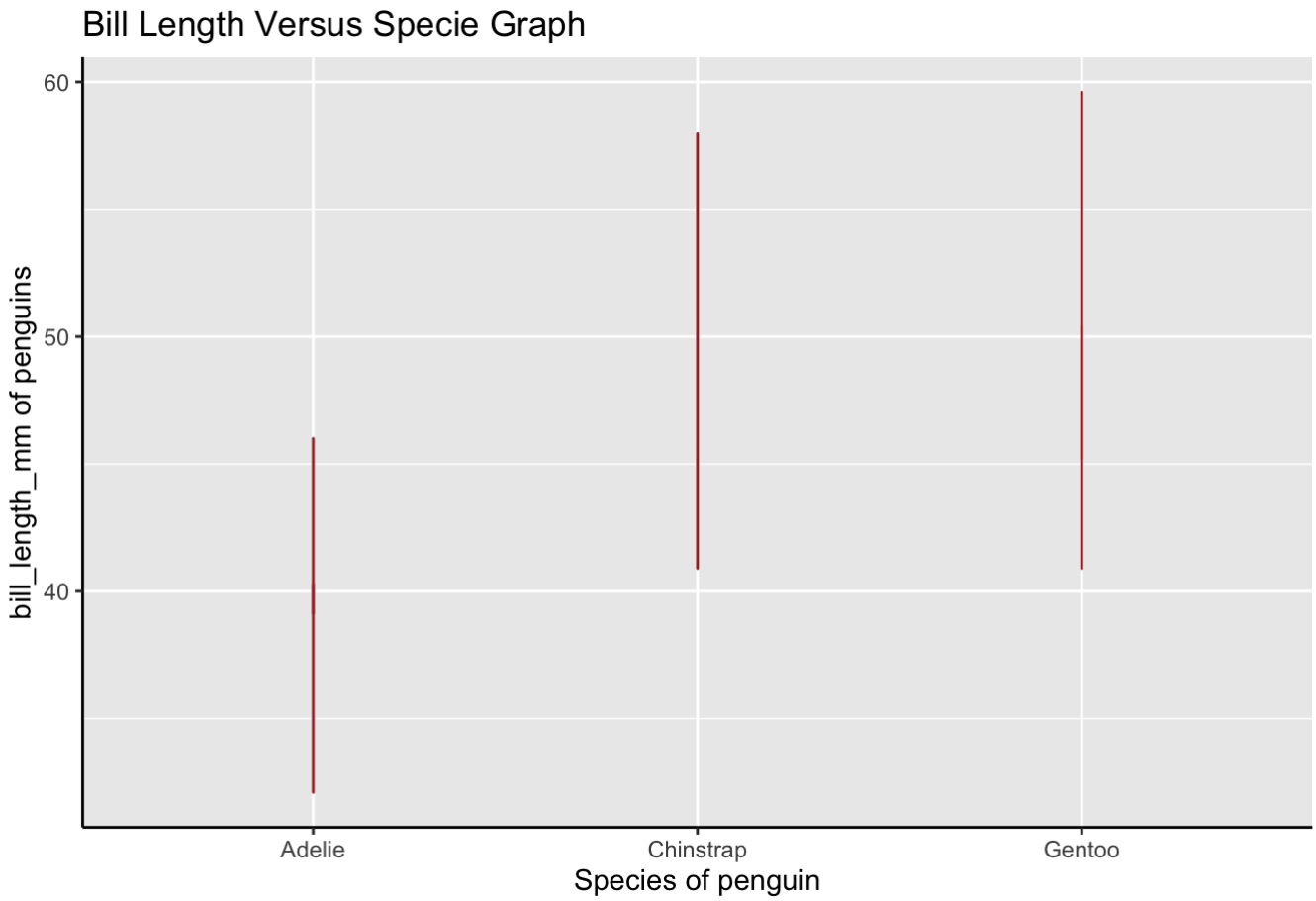

### Description of Objective 4

Through the above **Stacked Bar Graph** we can conclude :

- **Adelie Specie** - Found on all three islands, i.e.. on Biscoe , Dream and Torgersen
- **Chinstrap Specie** - Found only on Dream island.
- **Gentoo Specie** - Found on Biscoe island
- ***Gentoo specie penguins and Adelie specie penguins have good bondings as they co-exist on same island***
- ***Chinstrap specie penguins and Adelie specie penguins may also have good understanding as they co-exist on Dream island***
- It can also be said that Adelie specie are more adaptive as compared to others because they can survive on every island.

# Objective 5 - To find out which penguin specie has the longest bill length.

```
ggplot(penguins)+
  geom_line(aes(x = species,y = bill_length_mm),color = "brown")+
  theme_gray()+
  labs(x = "Species of penguin" , y = "bill_length_mm of penguins", title = "Bill Len
gth Versus Specie Graph", caption = "Data Source by Dr. Kristen Gorman")+
  theme(axis.line = element_line("black"))
```




Data Source by Dr. Kristen Gorman

### Description of objective 5

From the above **Line Graph** we get an idea for the range of bill length and it can be concluded that :

- **Gentoo Specie** of penguins has the highest bill length , going *near 60 mm* .

# Here marks the end of R Markdown

- data source - R Studio