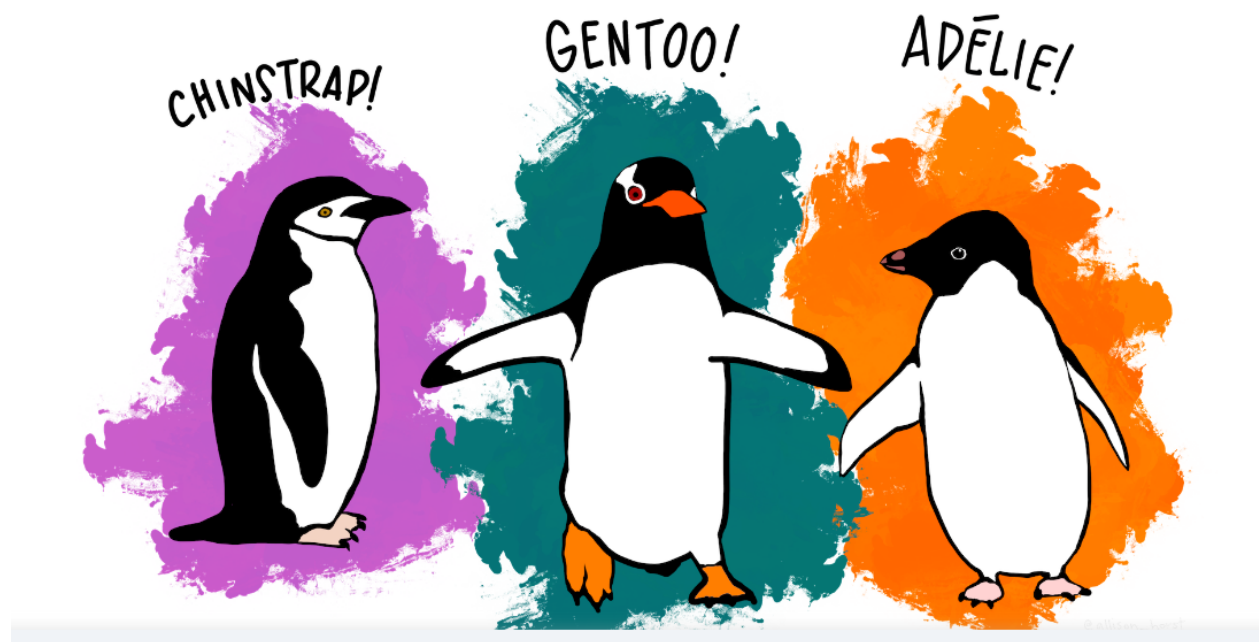


Analysis Of Palmer Penguins Dataset Using R

About the palmerpenguins dataset

It is a dataset comprising various measurements of three different penguin species, namely Adelie, Gentoo, and Chinstrap. The rigorous study was conducted in the islands of the Palmer Archipelago, Antarctica. These data were collected from 2007 to 2009 by Dr. Kristen Gorman with the Palmer Station Long Term Ecological Research Program, part of the US Long Term Ecological Research Network.



Setting up my environment for Data Exploration by loading the necessary R packages

- 'tidyverse' package for data exploration
- 'ggplot2' package for data visualization

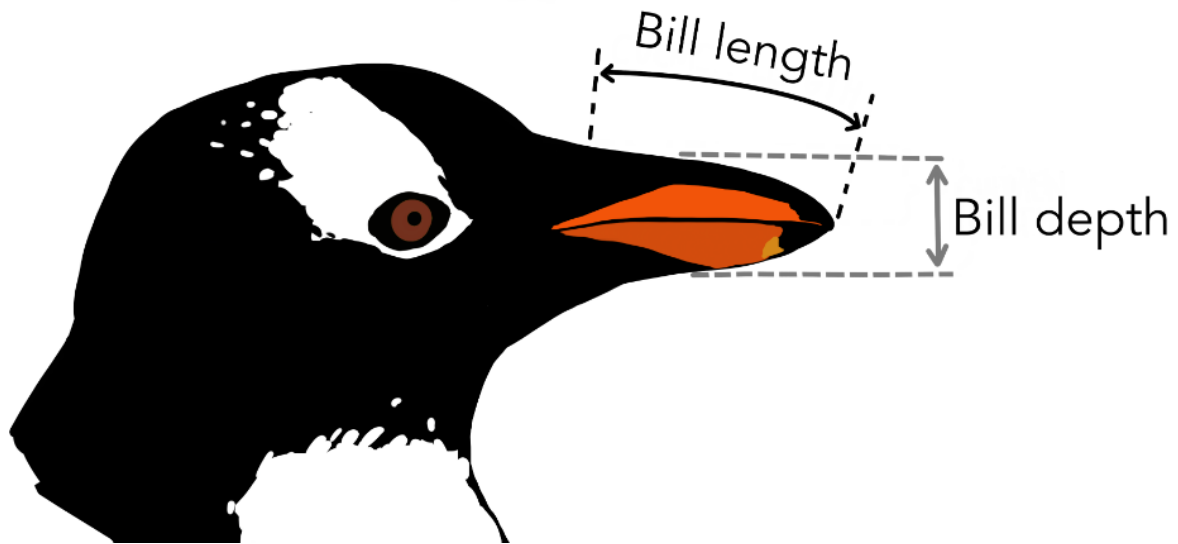
```
library(tidyverse)
library(ggplot2)
library(palmerpenguins)
```

In the next step, I have read the dataset using view() function.

```
view(penguins)
```

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	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
7	Adelie	Torgersen	38.9	17.8	181	3625	female	2007
8	Adelie	Torgersen	39.2	19.6	195	4675	male	2007
9	Adelie	Torgersen	34.1	18.1	193	3475	NA	2007
10	Adelie	Torgersen	42.0	20.2	190	4250	NA	2007

The below given image explains what is a bill_length & bill_depth in a penguins.



Exploratory Data Analysis:

I have used Exploratory Data Analysis (EDA) to analyze the data and discover trends, patterns, or check assumptions in data.

Maximum, Minimum & Mean bill length by species:

```
penguins%>%
group_by(species)%>%
drop_na()%>%
summarise(max_bill_length_mm=max(bill_length_mm),
           min_bill_length_mm=min(bill_length_mm),
           mean_bill_length_mm=mean(bill_length_mm))
```

```
## # A tibble: 3 x 4
##   species    max_bill_length_mm min_bill_length_mm mean_bill_length_mm
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 Adelie         46             32.1           38.8
## 2 Chinstrap      58             40.9           48.8
## 3 Gentoo        59.6           40.9           47.6
```

Maximum, Minimum & Mean bill depth by species:

```
penguins%>%
group_by(species)%>%
drop_na()%>%
summarise(max_bill_depth_mm=max(bill_depth_mm),
           min_bill_depth_mm=min(bill_depth_mm),
           mean_bill_depth_mm=mean(bill_depth_mm))
```

```
## # A tibble: 3 x 4
##   species    max_bill_depth_mm min_bill_depth_mm mean_bill_depth_mm
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 Adelie        21.5           15.5           18.3
## 2 Chinstrap     20.8           16.4           18.4
## 3 Gentoo        17.3           13.1           15.0
```

Maximum, Minimum & Mean flipper_length by species:

```
penguins%>%
group_by(species)%>%
drop_na()%>%
summarise(max_flipper_length_mm=max(flipper_length_mm),
           min_flipper_length_mm=min(flipper_length_mm),
           mean_flipper_length_mm=mean(flipper_length_mm))
```

```
## # A tibble: 3 x 4
##   species    max_flipper_length_mm min_flipper_length_mm mean_flipper_length_mm
##   <fct>          <int>          <int>          <dbl>
## 1 Adelie         210            172            190.
## 2 Chinstrap      212            178            196.
## 3 Gentoo        231            203            217.
```

Maximum, Minimum & Mean body__mass by species:

```
penguins%>%
  group_by(species)%>%
  drop_na()%>%
  summarise(max_body_mass_g=max(body_mass_g),
            min_body_mass_g=min(body_mass_g),
            mean_body_mass_g=mean(body_mass_g))
```

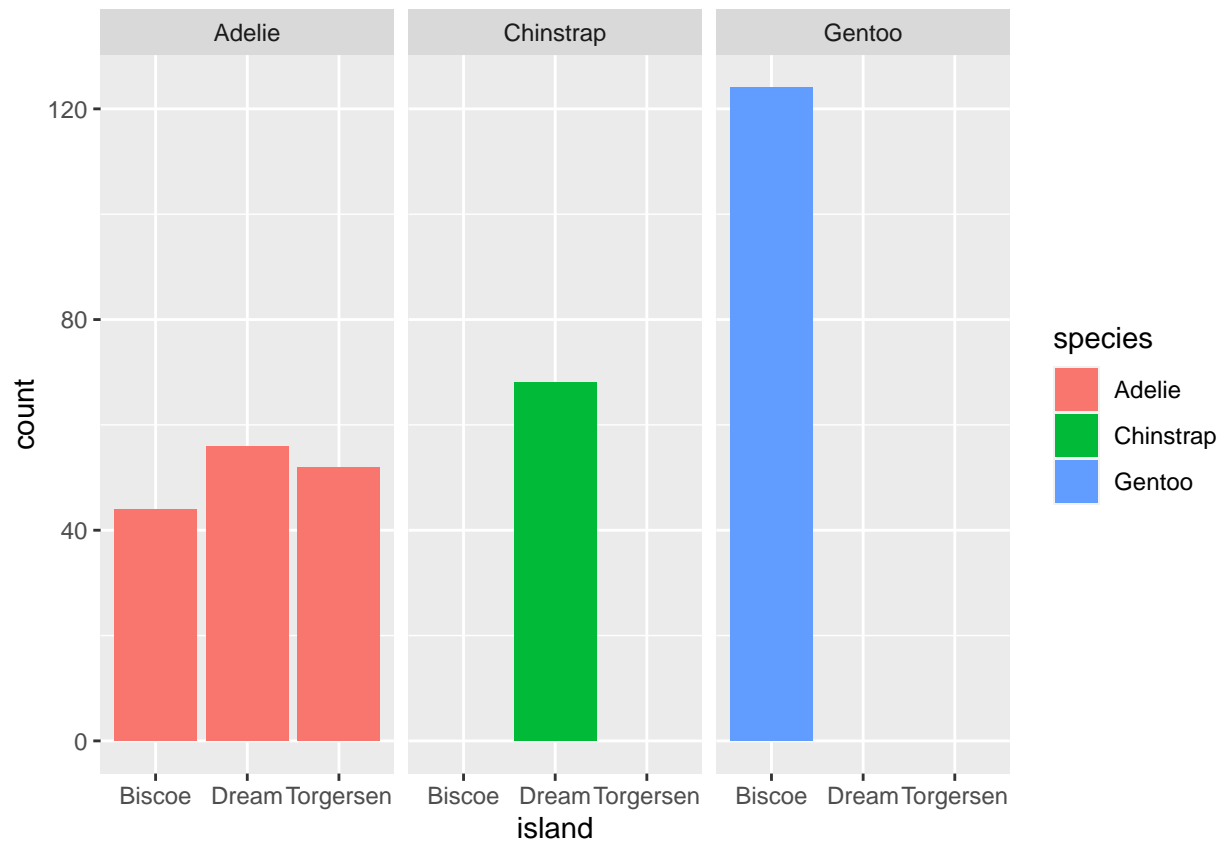
```
## # A tibble: 3 x 4
##   species    max_body_mass_g min_body_mass_g mean_body_mass_g
##   <fct>          <int>          <int>          <dbl>
## 1 Adelie         4775            2850            3706.
## 2 Chinstrap      4800            2700            3733.
## 3 Gentoo        6300            3950            5092.
```

Data Visualization

I am using 'ggplot2' package for plotting relationship between variables.

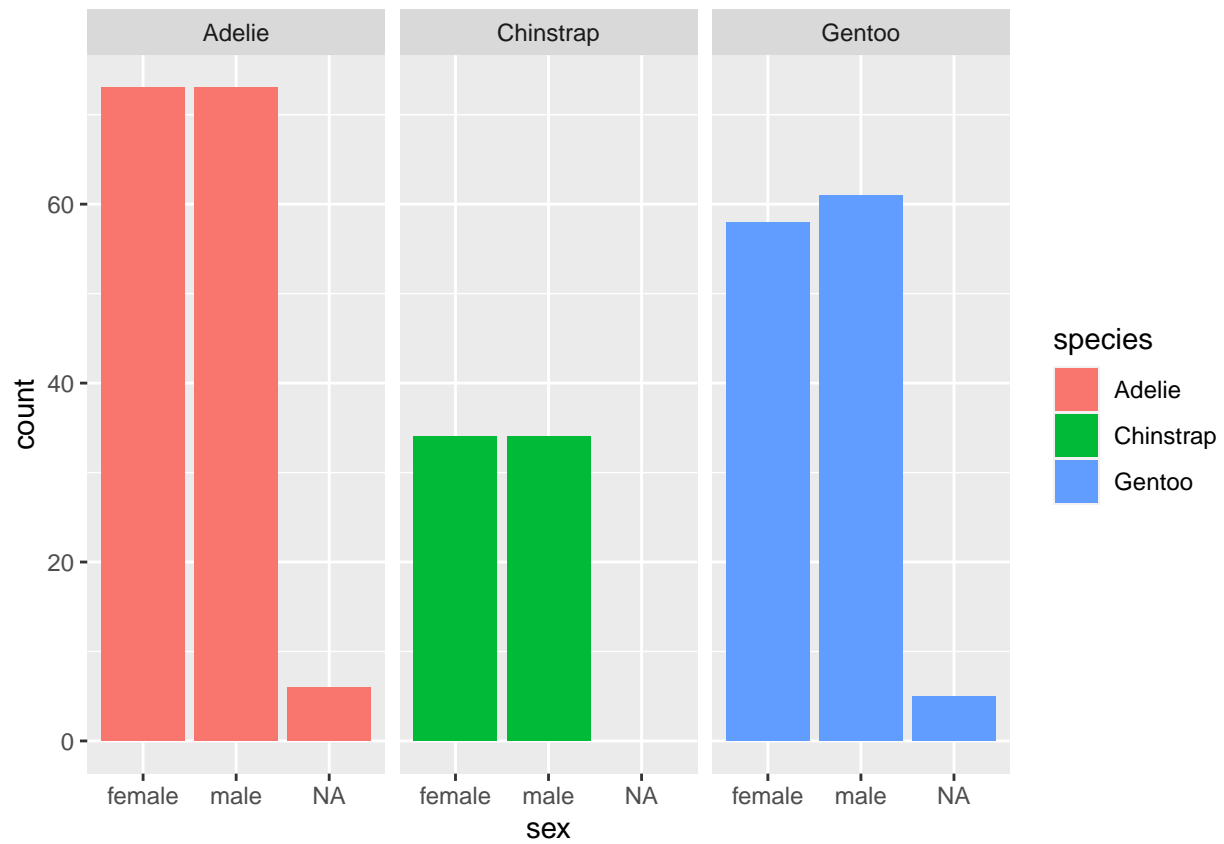
Species Based Island Count Plot:

```
ggplot(data = penguins)+  
  geom_bar(mapping = aes(x=island,fill=species))+  
  facet_wrap(~species)
```



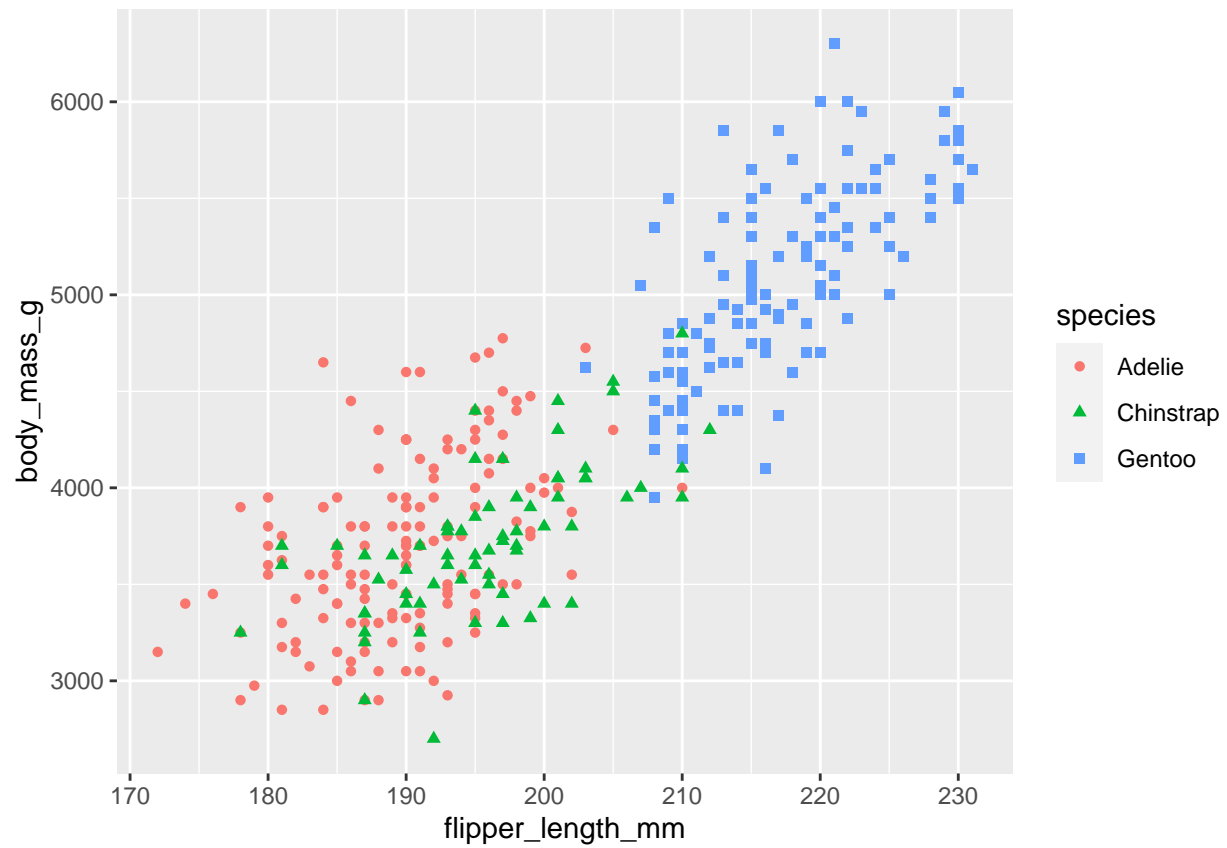
Species Based Gender Count Plot:

```
ggplot(data = penguins)+  
  geom_bar(mapping = aes(x=sex,fill=species))+  
  facet_wrap(~species)
```



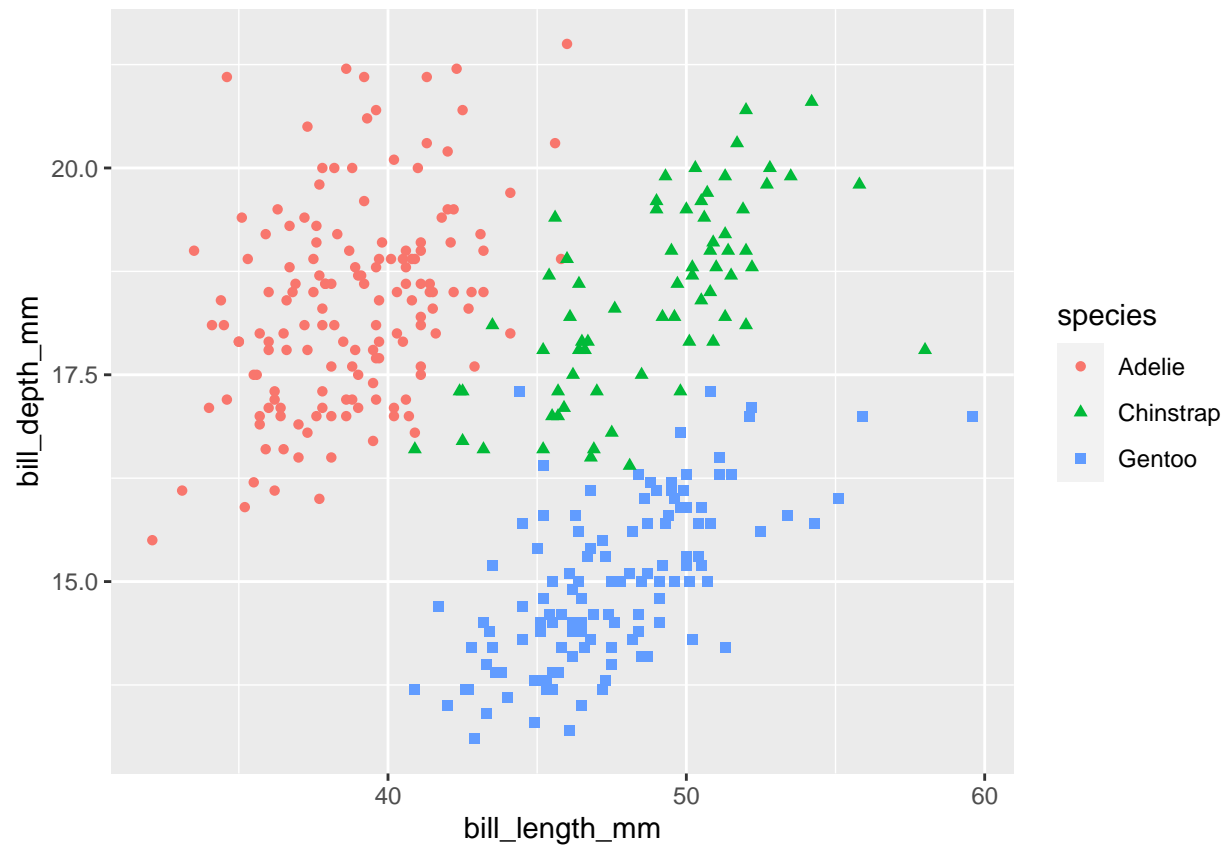
flipper length and body mass scatter Plot:

```
ggplot(data = penguins)+  
  geom_point(mapping = aes(x=flipper_length_mm,y=body_mass_g,color=species,shape=species))
```



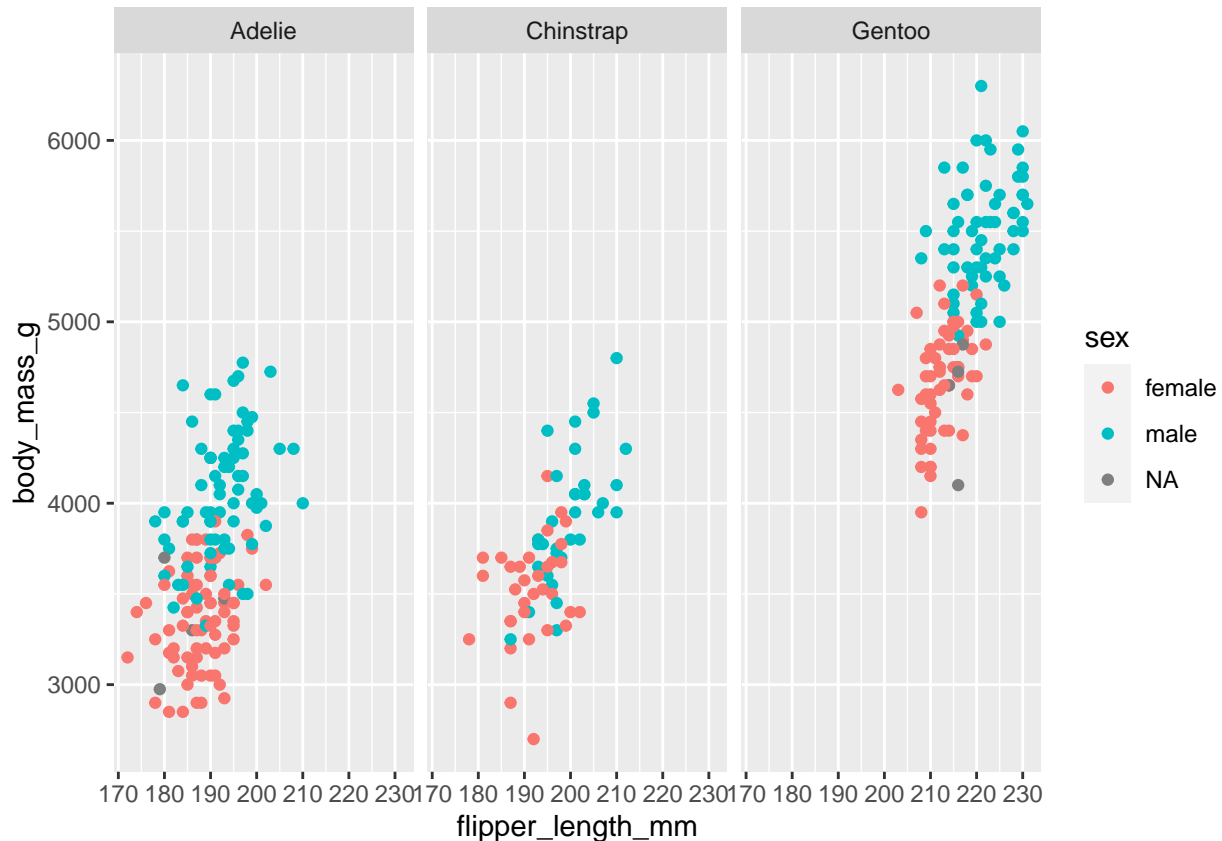
bill length and bill depth scatter Plot:

```
ggplot(data = penguins)+  
  geom_point(mapping = aes(x=bill_length_mm,y=bill_depth_mm,color=species,shape=species))
```



Species based gender scatter Plot for flipper_length Vs body_mass:

```
ggplot(data = penguins)+  
  geom_point(mapping = aes(x=flipper_length_mm,y=body_mass_g,color=sex))+  
  facet_wrap(~sex)+  
  facet_wrap(~species)
```



Conclusion

So, this was all about data exploration and visualization of Palmer Penguins data. Below are some key takeaways from the above analysis:

- Used functions from the 'tidyverse' package for data exploration, data cleaning & to do exploratory data analysis.
- Used functions from 'ggplot2' to plot different interactive plots describing relationships among variables.
- Used R Markdown to create this analysis report in html & pdf formats.