

R Markdown - Palmer Penguins

Anurag Chauhan

2022-08-14

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.1

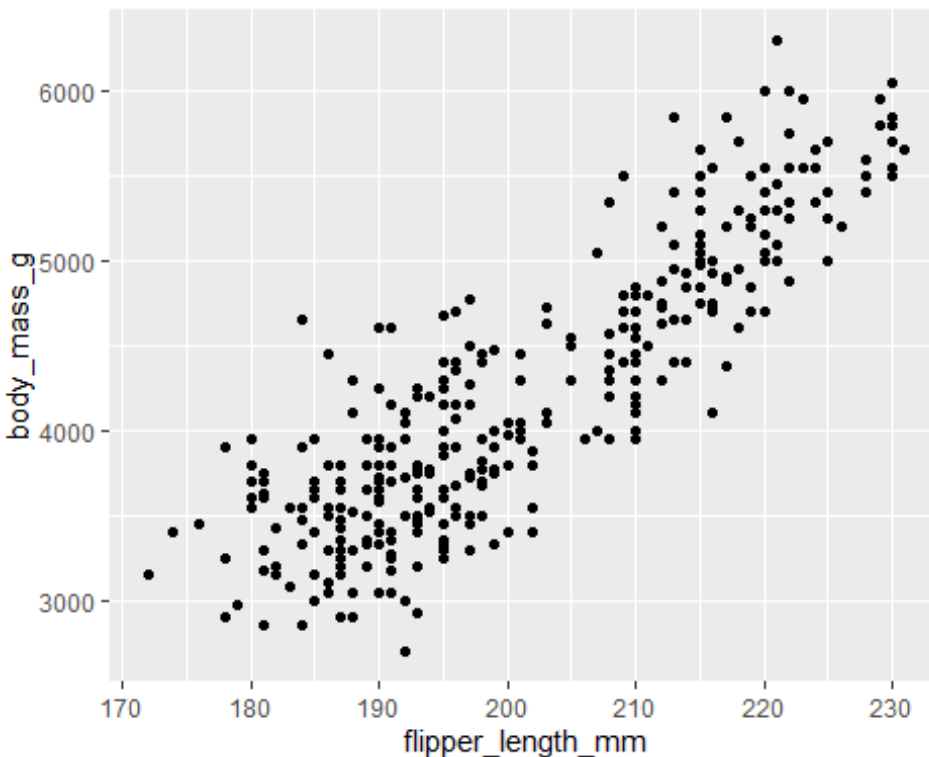
library(palmerpenguins)

## Warning: package 'palmerpenguins' was built under R version 4.2.1

data("penguins")
View(penguins)

ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g))

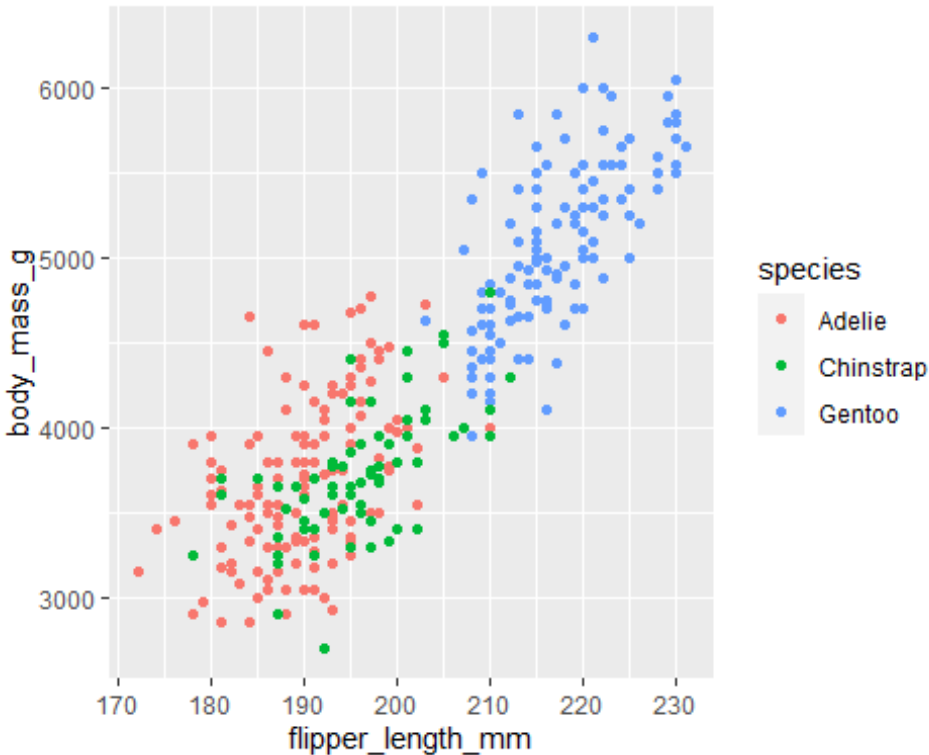
## Warning: Removed 2 rows containing missing values (geom_point).
```



- The plot shows positive relationship between the two variables i.e flipper_length_mm and body_mass_g.

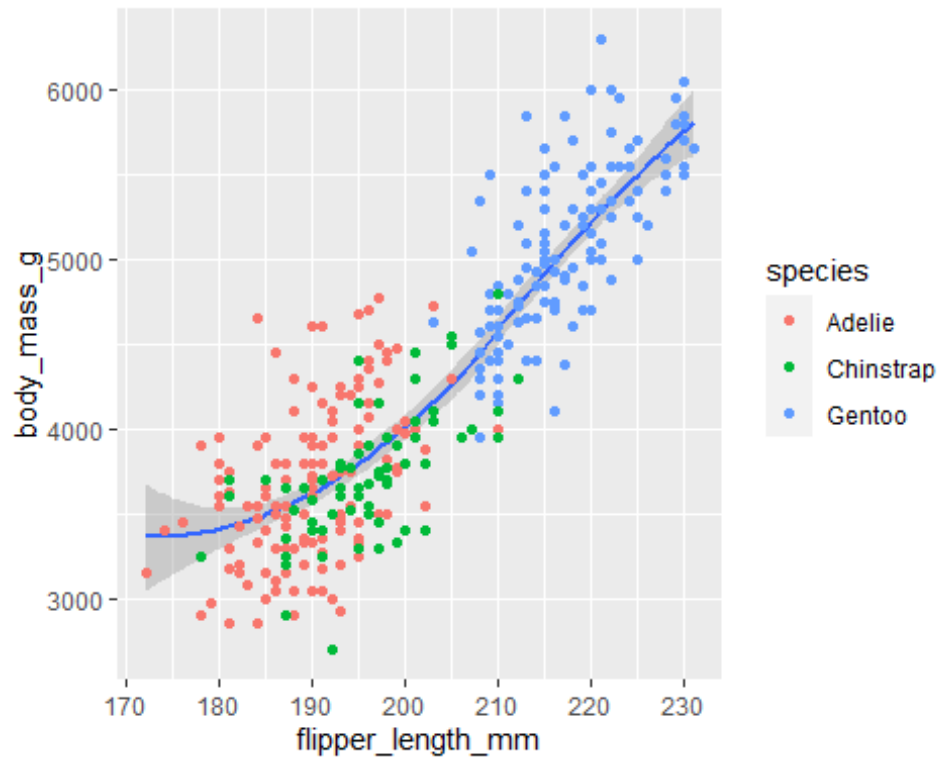
- But the plot has certain limitations. We can not tell which data points refer to which of the three penguin species.

```
ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))
## Warning: Removed 2 rows containing missing values (geom_point).
```



- Here the plot is more clear. One can easily distinguish between the penguin species.
- One can infer from the plot that the Gentoo species have the longest flipper length and highest body mass.

```
ggplot(data=penguins)+
  geom_smooth(mapping=aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
## Warning: Removed 2 rows containing missing values (geom_point).
```

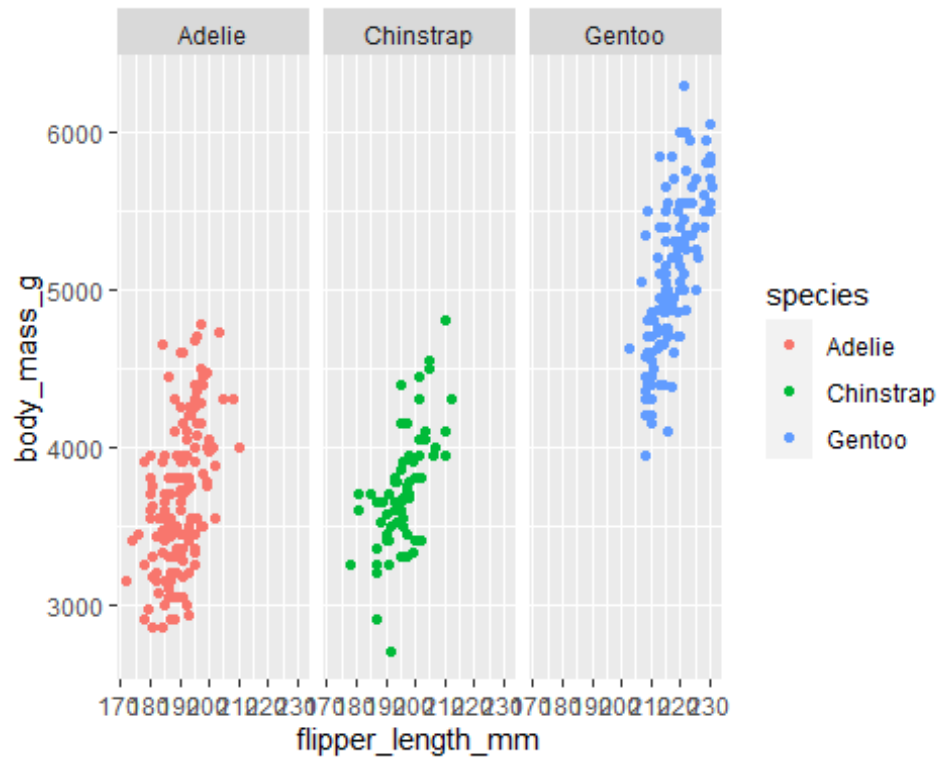


- `geom_smooth()` function is used to plot smooth curve through the data points.

facet_function

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

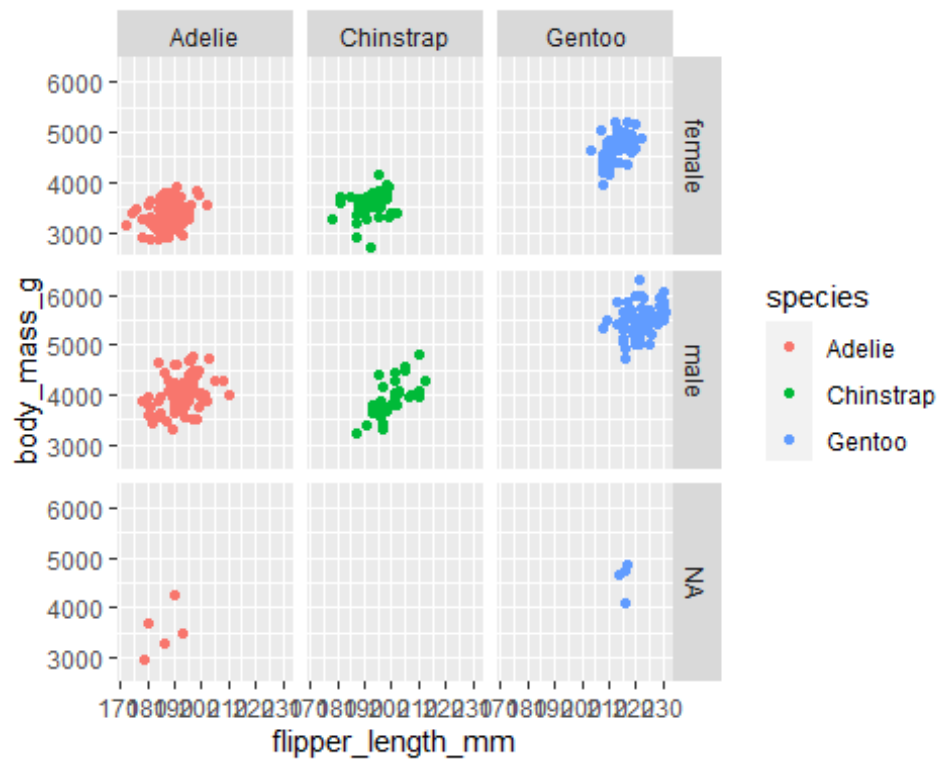
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



- facet function is used to create smaller groups or subsets of your data which makes analysis easier.
- facet_wrap() is used when you want to facet your plot by a single variable.
- facet_grid() is used if there are more than 1 variable. But facet_grid() can also be used with single variable.

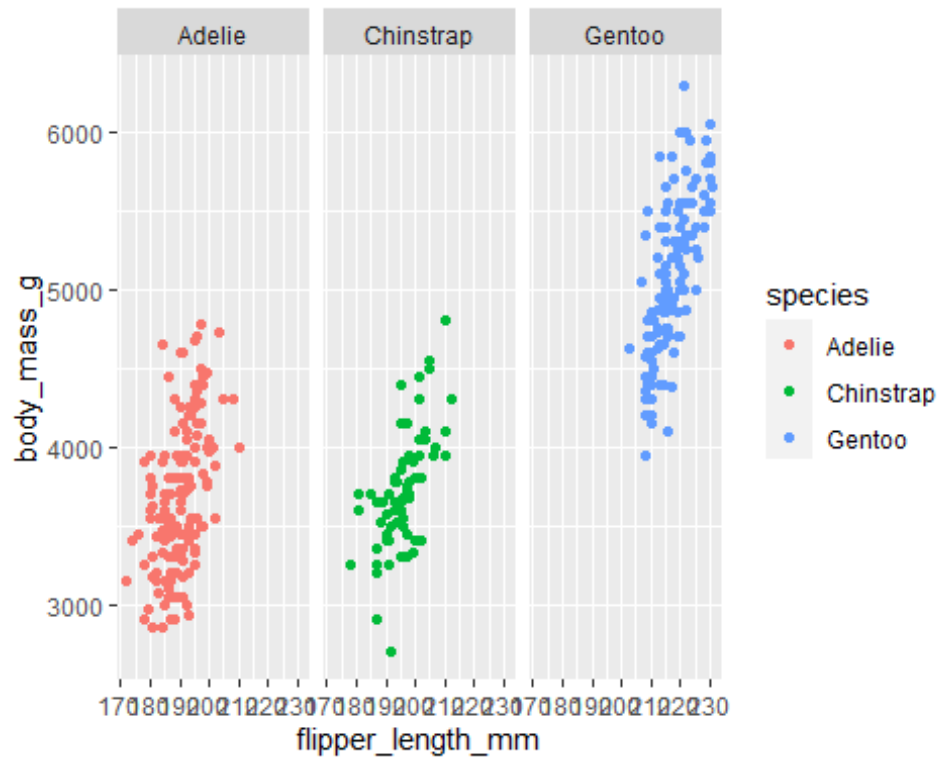
```
ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+
  facet_grid(sex~species)
```

Warning: Removed 2 rows containing missing values (geom_point).



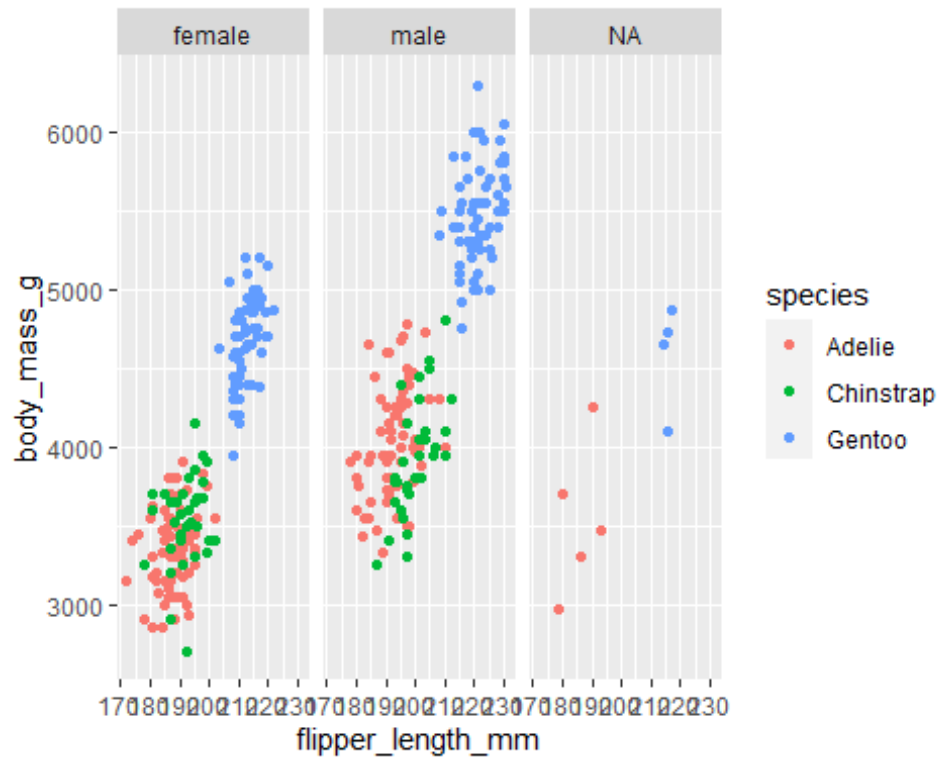
```
ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+
  facet_grid(~species)

## Warning: Removed 2 rows containing missing values (geom_point).
```



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

## Warning: Removed 2 rows containing missing values (geom_point).
```



Annotation

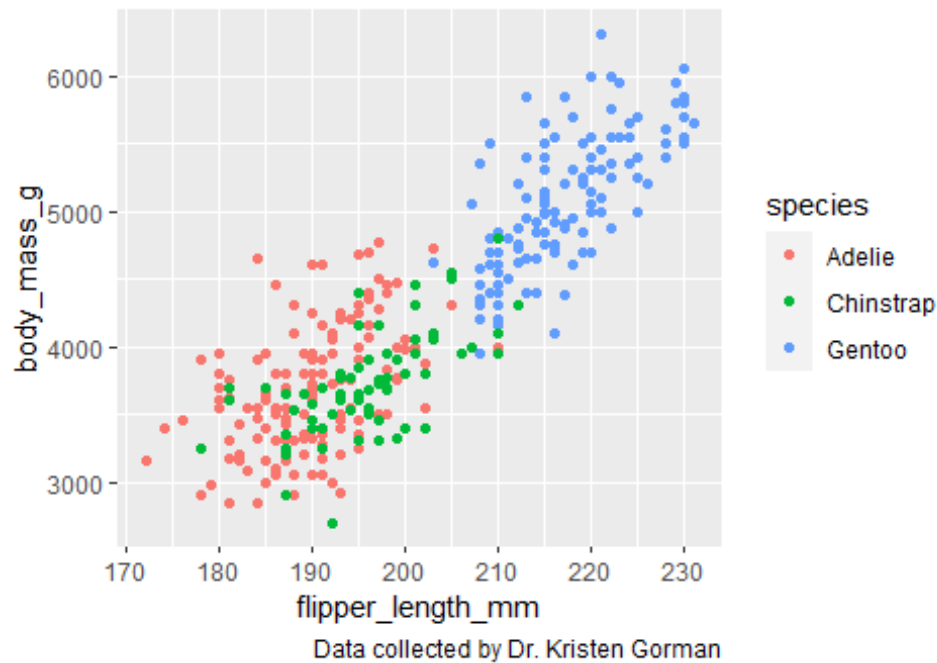
- Annotation refers to add notes to a document or diagram to explain or comment upon it.
- Annotation helps to quickly understand the plot.

```
ggplot(data=penguins)+
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle=
"Sample of Three Penguin Species",caption="Data collected by Dr. Kristen
Gorman")
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

Palmer Penguins: Body Mass vs. Flipper Length

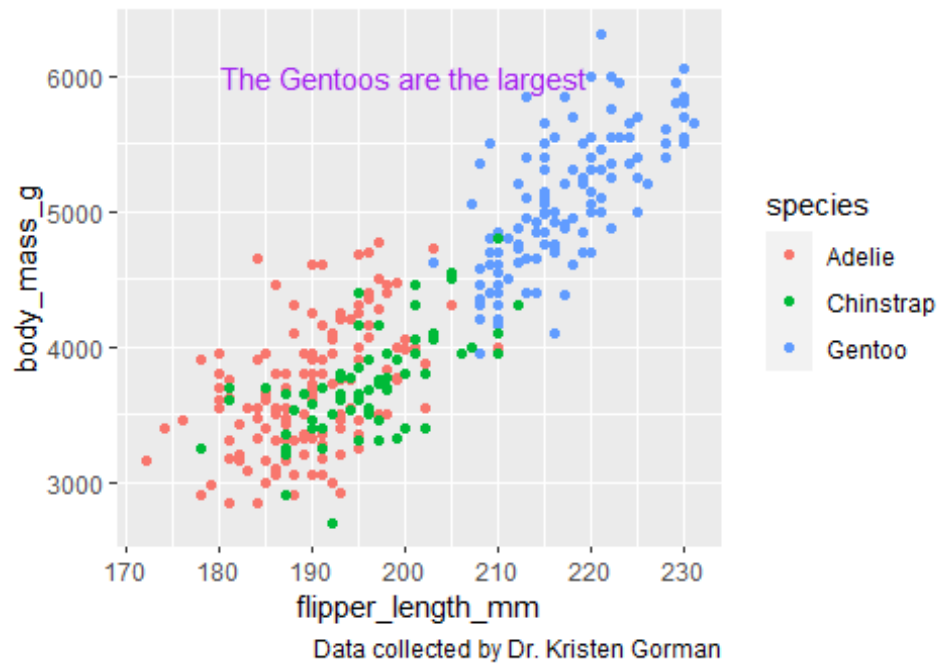
Sample of Three Penguin Species



```
ggplot(data=penguins)+  
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle=  
"Sample of Three Penguin Species",caption="Data collected by Dr. Kristen  
Gorman")+  
  annotate("text",x=200,y=6000,label="The Gentoos are the  
largest",color="Purple")  
  
## Warning: Removed 2 rows containing missing values (geom_point).
```


Palmer Penguins: Body Mass vs. Flipper Length

Sample of Three Penguin Species

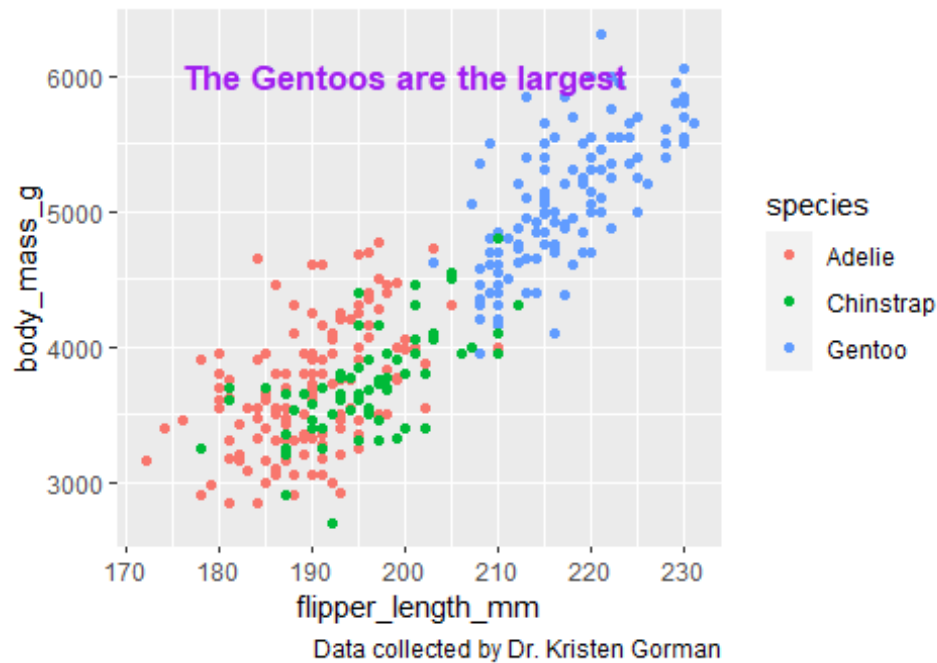


- `annotate()` function is used to add labels on the plot itself.

```
ggplot(data=penguins)+  
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle=  
"Sample of Three Penguin Species",caption="Data collected by Dr. Kristen  
Gorman")+  
  annotate("text",x=200,y=6000,label="The Gentoos are the  
largest",color="purple",fontface="bold",size=4.5)  
  
## Warning: Removed 2 rows containing missing values (geom_point).
```

Palmer Penguins: Body Mass vs. Flipper Length

Sample of Three Penguin Species

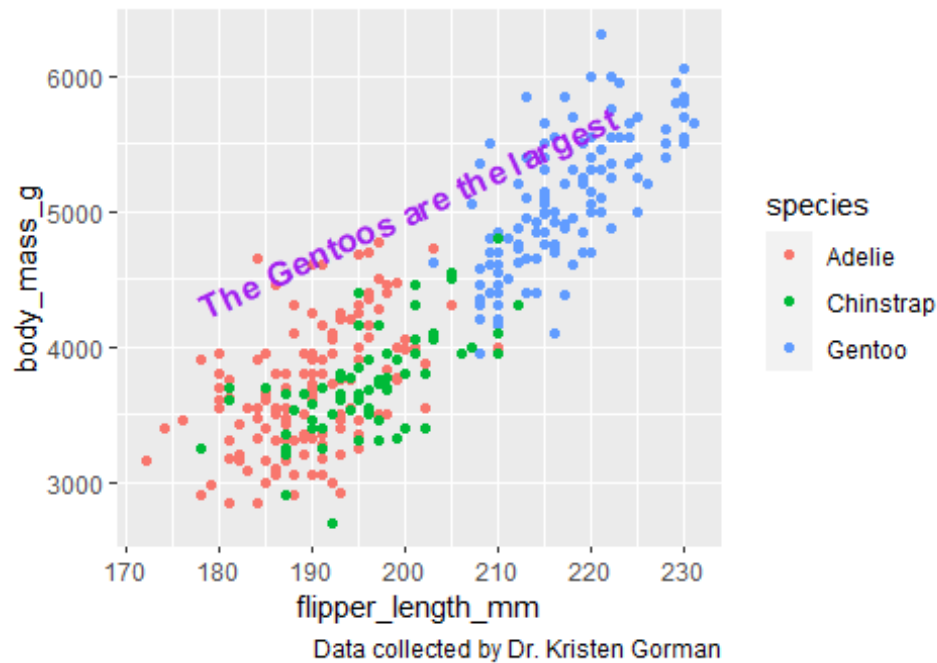


- We can also change the size of labels and make it bold with annotate function.

```
ggplot(data=penguins)+  
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle=  
"Sample of Three Penguin Species",caption="Data collected by Dr. Kristen  
Gorman")+  
  annotate("text",x=200,y=5000,label="The Gentoos are the  
largest",color="purple",fontface="bold",size=4.5,angle=25)  
  
## Warning: Removed 2 rows containing missing values (geom_point).
```

Palmer Penguins: Body Mass vs. Flipper Length

Sample of Three Penguin Species



#Storing a plot as variable

```
p<-ggplot(data=penguins)+  
  geom_point(mapping=aes(x=flipper_length_mm,y=body_mass_g,color=species))+  
  labs(title="Palmer Penguins: Body Mass vs. Flipper Length",subtitle=  
"Sample of Three Penguin Species",caption="Data collected by Dr. Kristen  
Gorman")
```

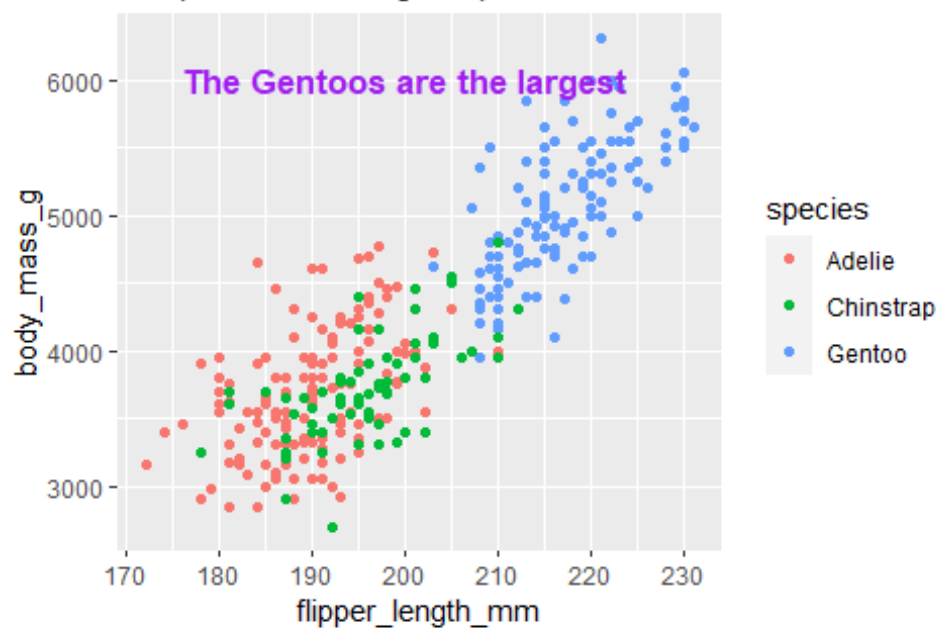
#adding annotation to previous saved plot as a variable

```
p+annotate("text",x=200,y=6000,label="The Gentoos are the  
largest",color="purple",fontface="bold",size=4.5)
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

Palmer Penguins: Body Mass vs. Flipper Length

Sample of Three Penguin Species



Data collected by Dr. Kristen Gorman