

Palmer Penguins

H Crutchfield

2024-02-26

Penguin Plots: Practice Analysis

Setting up my environment

Notes: setting up my R environment by loading the ‘tidyverse’ and ‘palmerpenguins’ packages:

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(palmerpenguins)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring:

- HTML
- PDF
- MS Word documents.

For more details on using R Markdown click here [Link] (<http://rmarkdown.rstudio.com>).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Visualizations

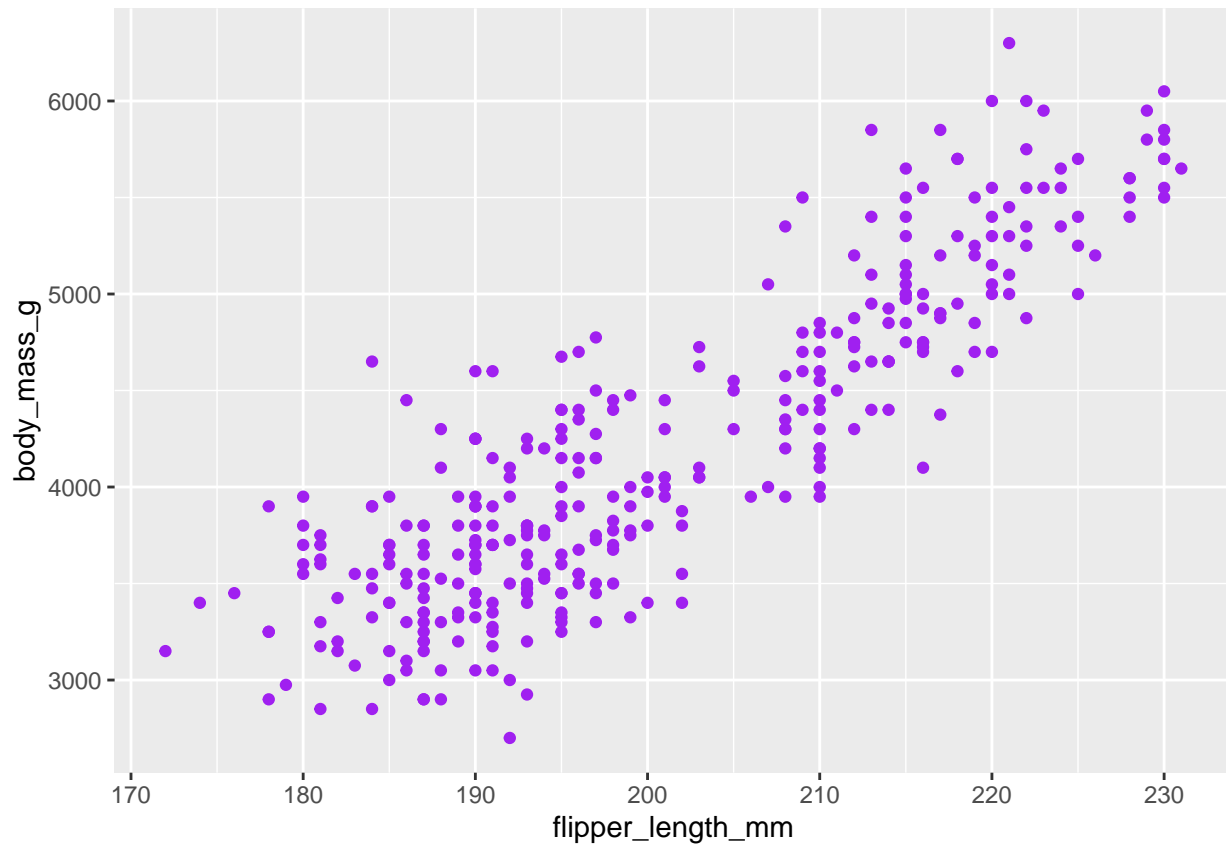
A series of visualizations

Including Plots

Plotted flipper length against body mass

```
ggplot(data=penguins,aes(x=flipper_length_mm,y=body_mass_g))+
  geom_point(color="purple")
```

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



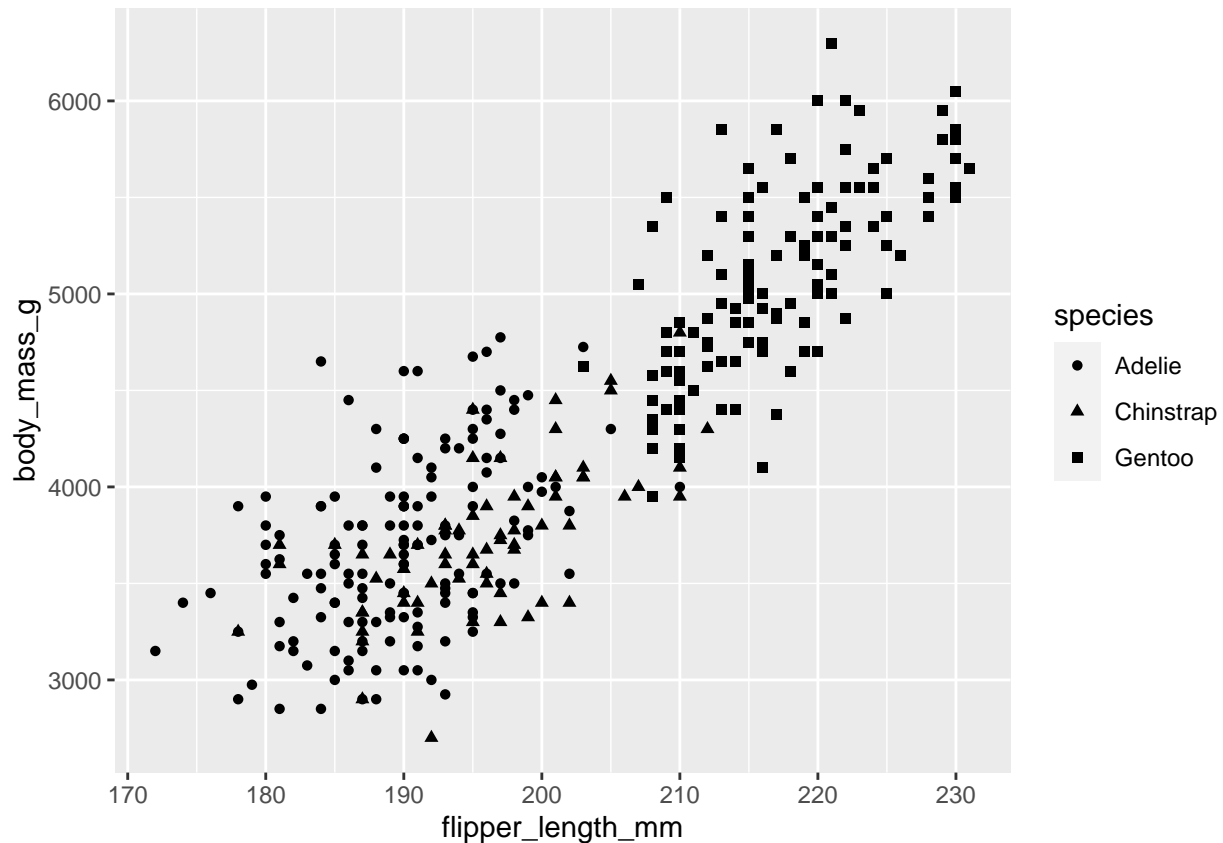
```
## Warning: Removed 2 rows containing missing values (geom_point).
```

Flipper and body mass by species

Plotted flipper length against body mass, looked at the breakdown by species.

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

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



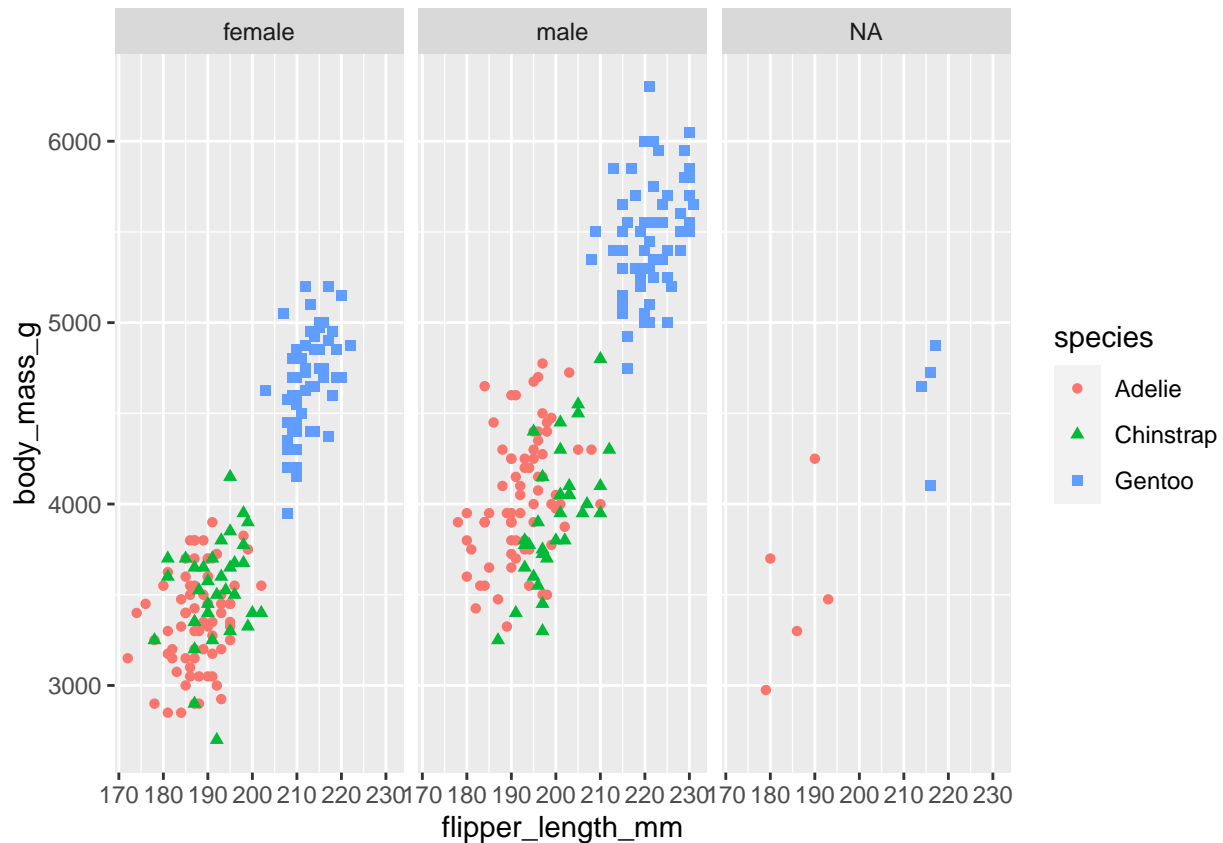
#3

Flipper and body mass by species and sex

Plotted flipper length against body mass, looked at the breakdown by species and sex

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

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



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

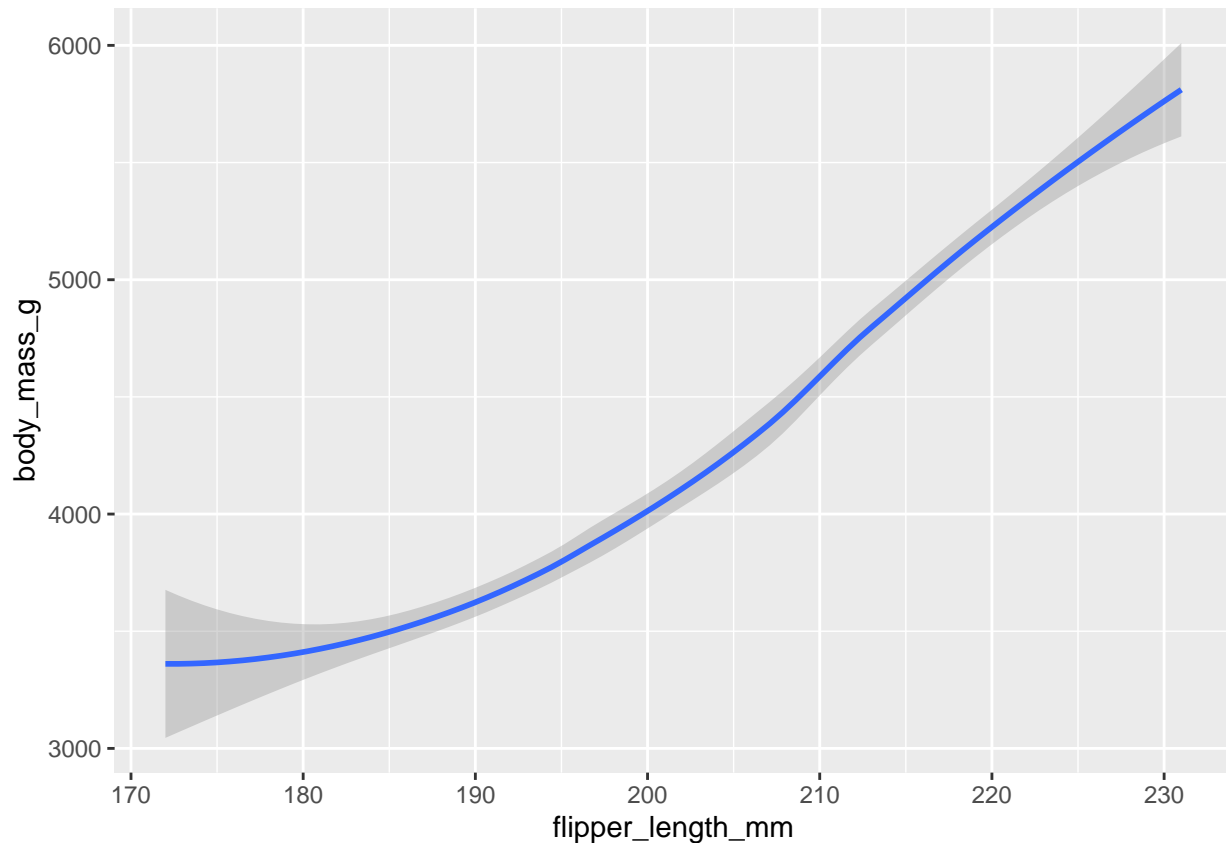
Creating a Line Chart

Another way to visualize data on Flipper Length vs Body Mass, we create a line chart that creates a trend line given the two variables. This is useful to see general trends between two variables and look for both positive, and negative correlations between them. This was done using the `geom_smooth()` function.

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

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```



This line graph shows a clear positive correlation between flipper length and body mass for the penguins included in this dataset.

All Together

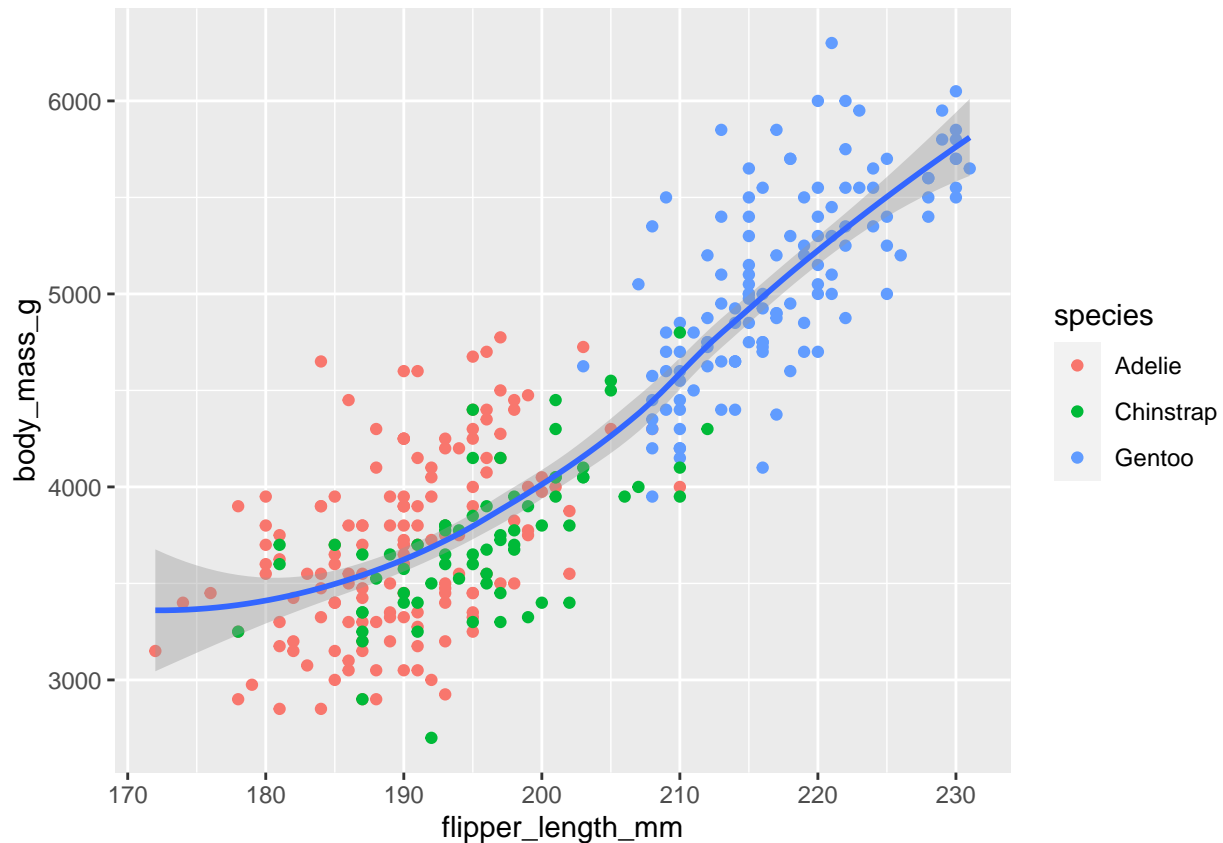
We now have both a scatter plot and a line graph, showing correlation between flipper length and body mass. The next chart combines the trend line with the scatter plot, showing positive correlation between flipper length and body mass amongst all species of penguins within the given dataset.

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

```
## `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()`).
```



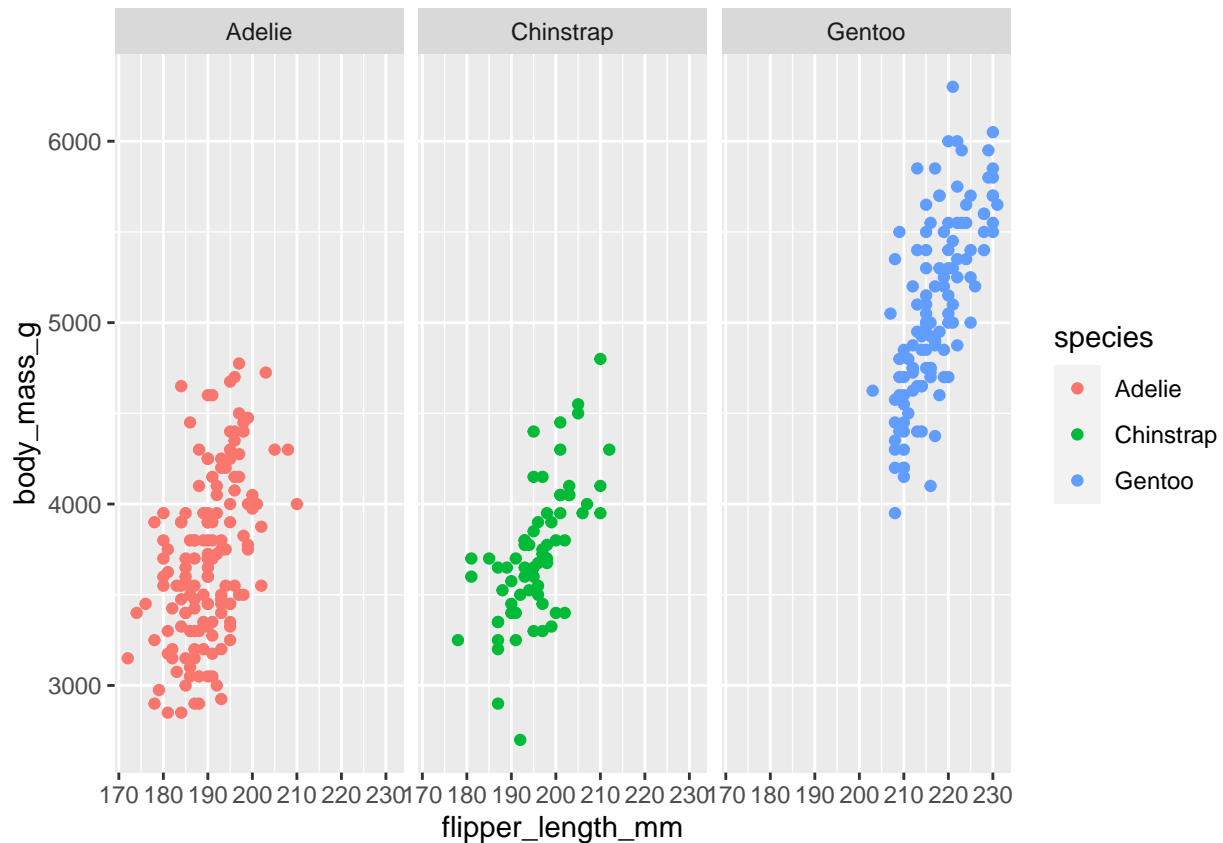
**Warning: Removed 2 rows containing non-finite values (geom_smooth),
Removed 2 rows containing missing values (geom_point).**

Facets

Differentiated between penguin species using the `facet_wrap()` 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()``).



Labels, Annotations, and Themes

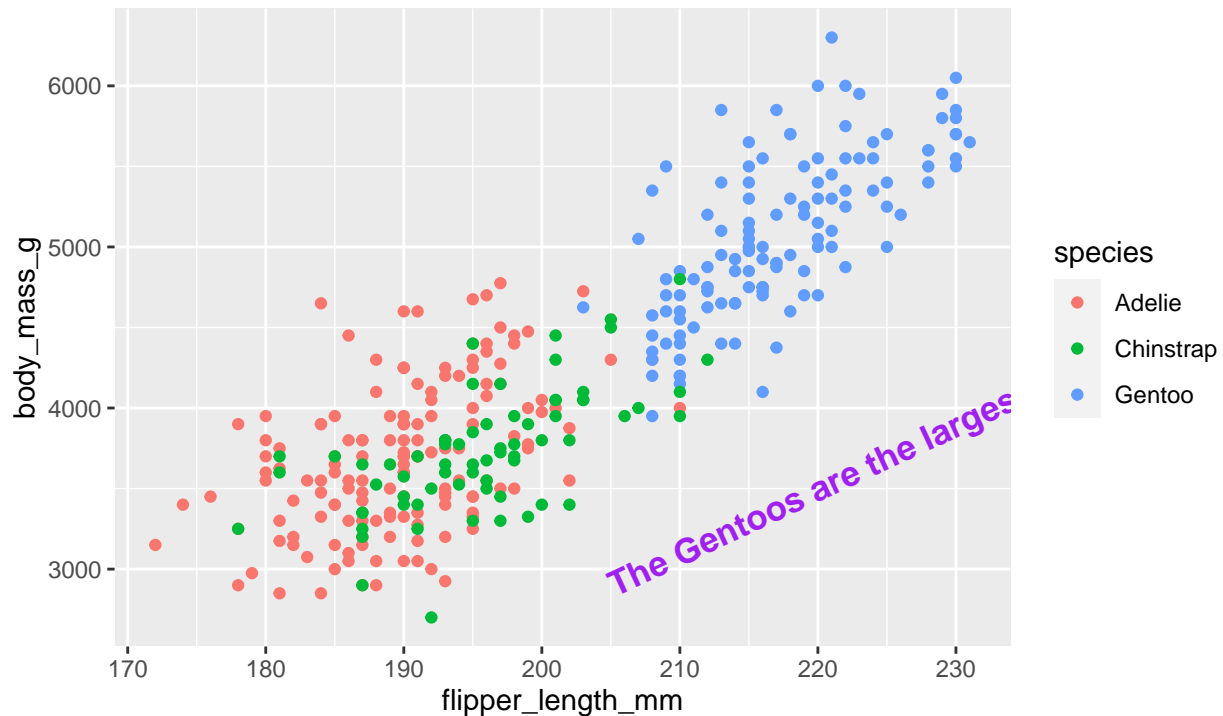
To make the visualization more comprehensive, we added titles, subtitles, captions, and annotations, emphasizing important information in our analysis.

```
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=220,y=3500,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



Data collected by Dr. Kristen Gorman

#

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

Conclusion

Through analysis, we have concluded there is a strong, positive correlation between flipper length and body mass of the penguins in the dataset. We were able to confirm that the correlation is found across all three species within the dataset. We created multiple visualizations to support these findings and provided information on how these visualizations were created and used to support the analysis.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.