Google Analytics_R for Data Visualization

Jiuru Lyu

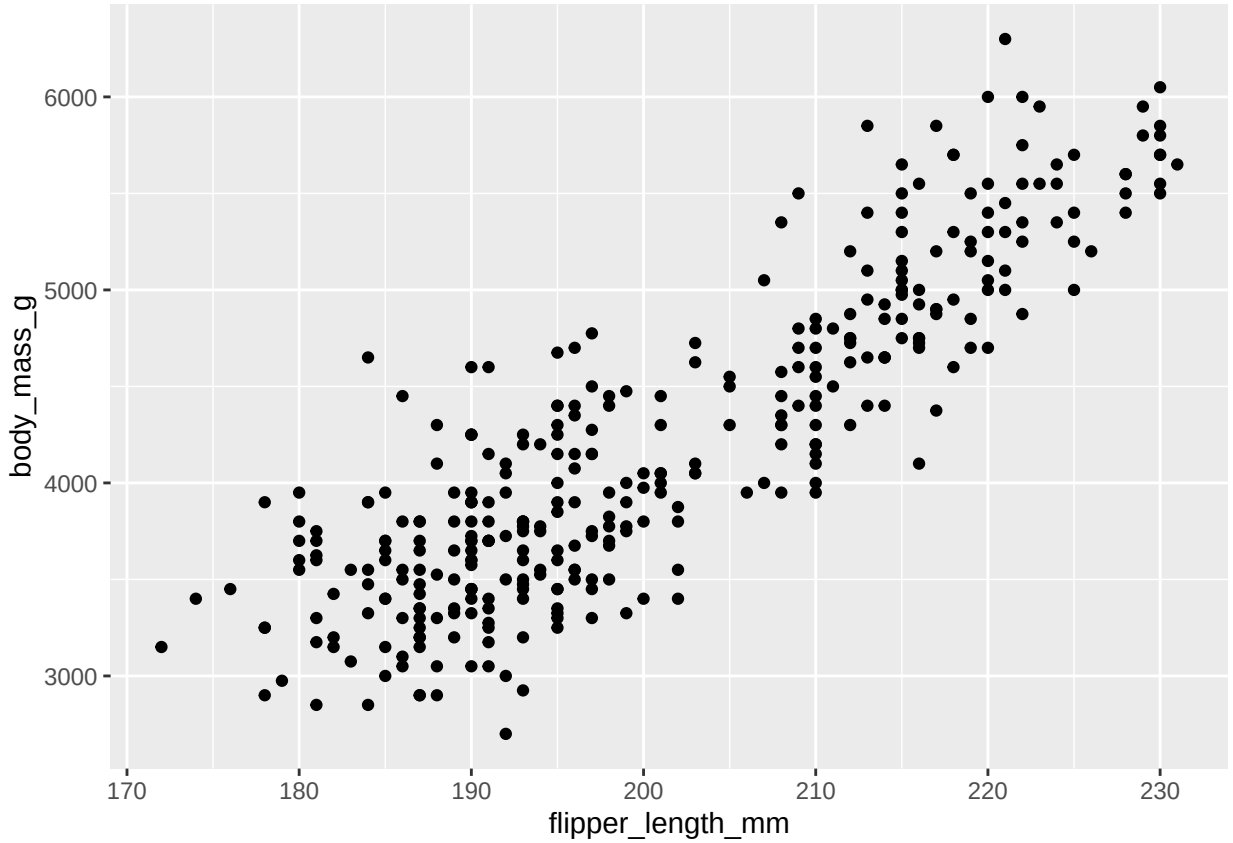
Jul 21, 2022

```
library("ggplot2")
library("palmerpenguins")
```

ggplot2 First Impression

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

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



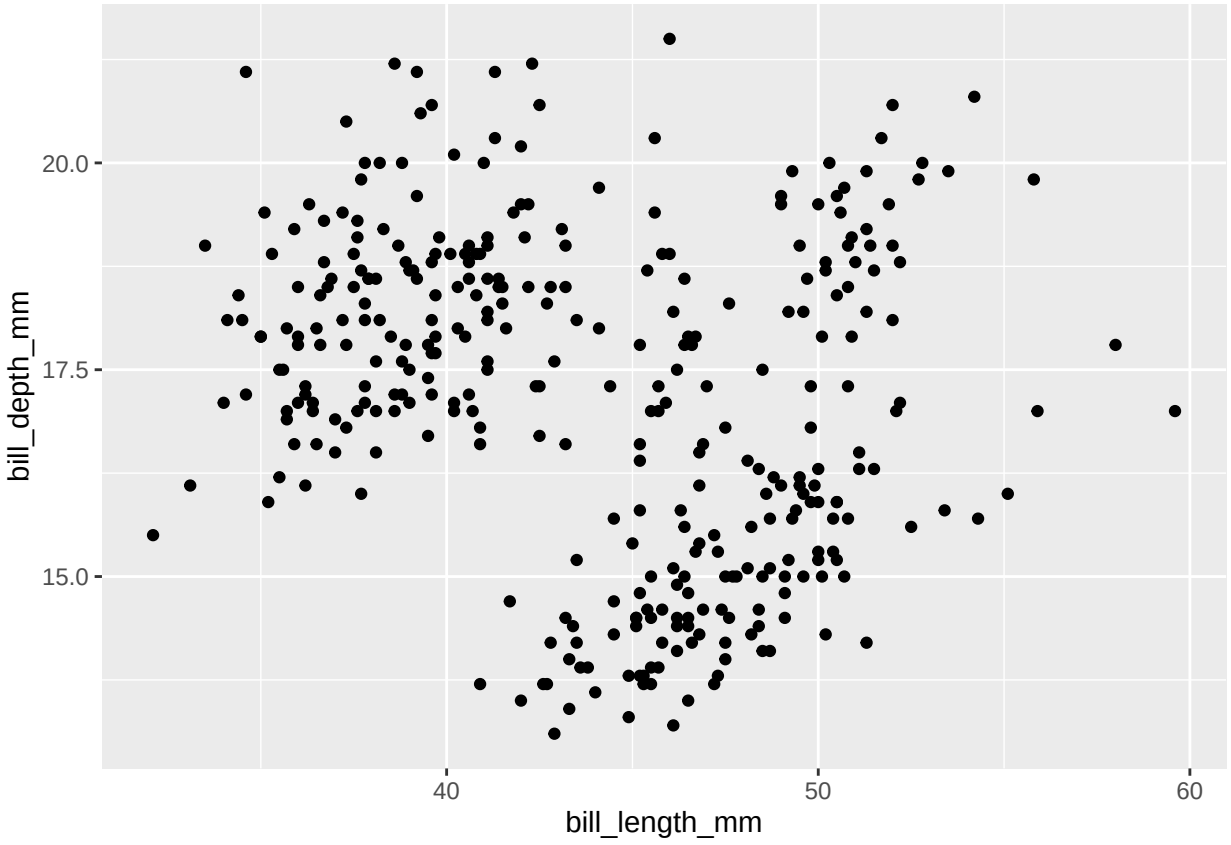
1. `ggplot(data=data_name)` tells what data frame ggplot2 should work with.
2. The plus sign (`+`) adds a layer to the plot. ggplot2 creates plots using multiple layers.
3. `geom_point` tells ggplot2 to use points to represent data points.
4. `mapping=aes()` tells ggplot2 the aesthetic.

```
ggplot(data=<DATA>)+<GEOM_FUNCTION>(mapping=aes(<AESTHETIC MAPPINGS>))
```

Example:

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

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

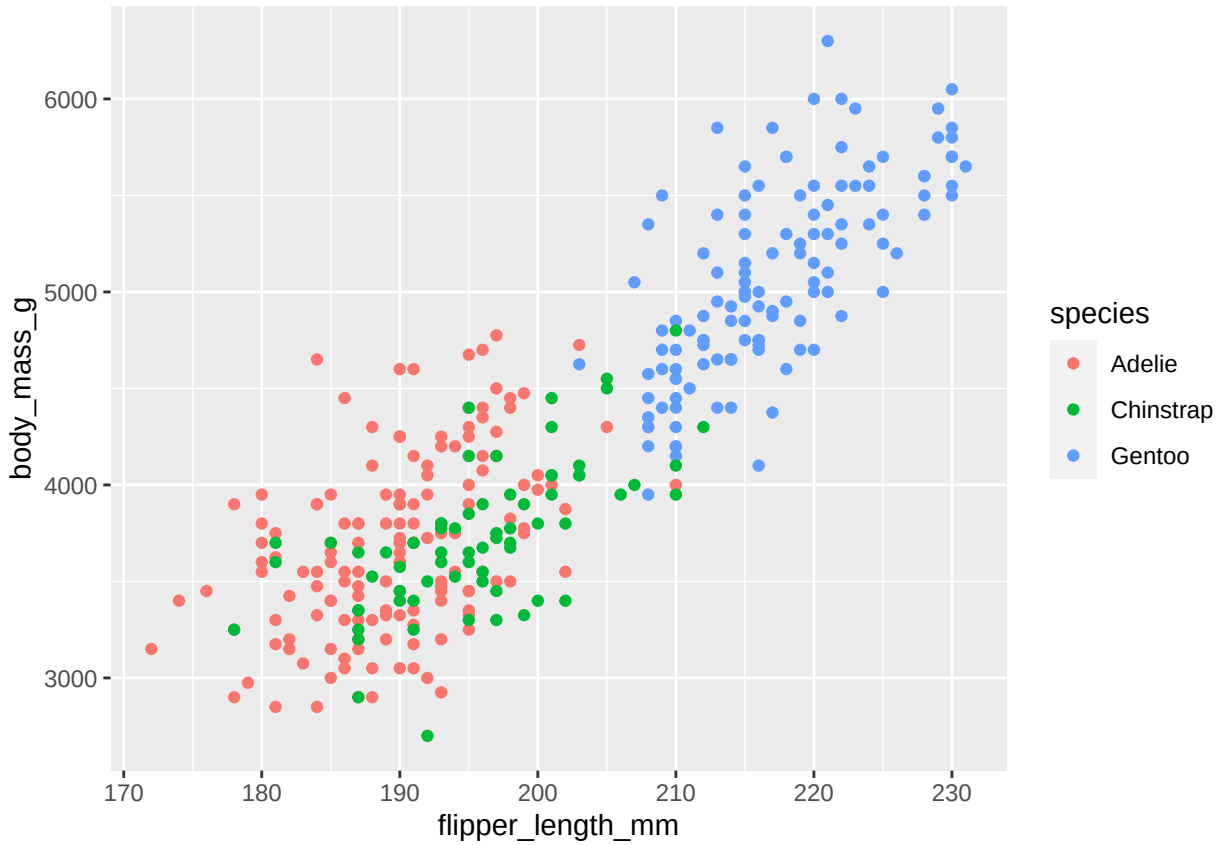


Explore Aesthetics in Analysis

Adding a new aesthetic to our previous code:

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

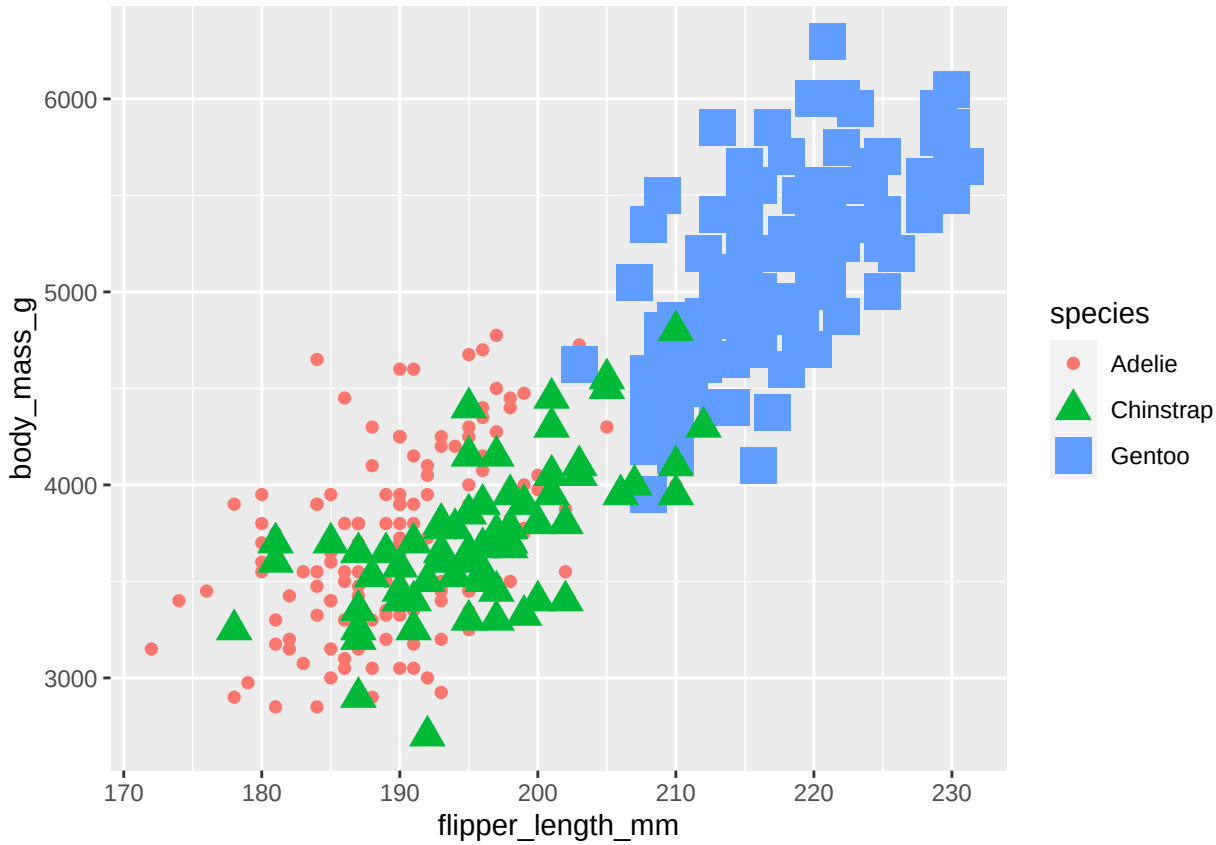
```
## Warning: Removed 2 rows containing missing values (geom_point).
```




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

```
## Warning: Using size for a discrete variable is not advised.
```

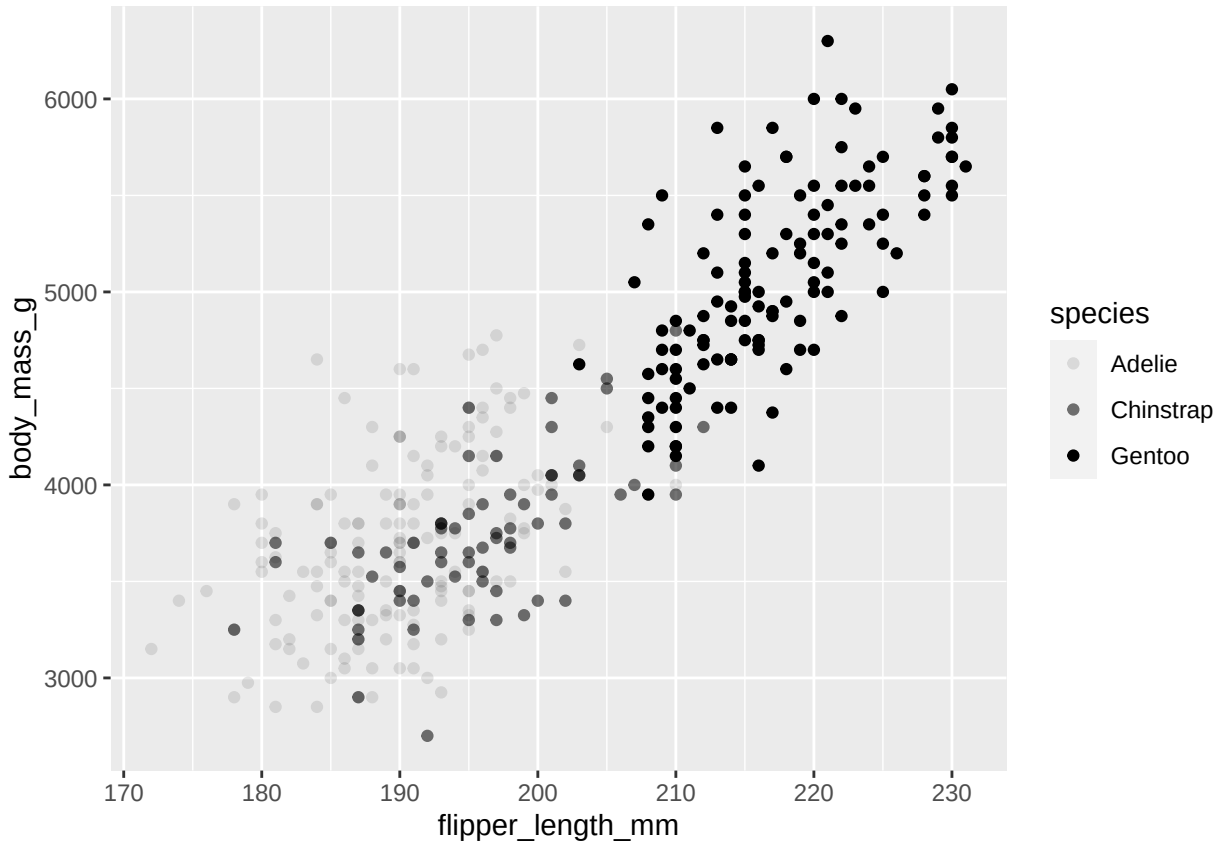
```
## Warning: Removed 2 rows containing missing values (geom_point).
```




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

```
## Warning: Using alpha for a discrete variable is not advised.
```

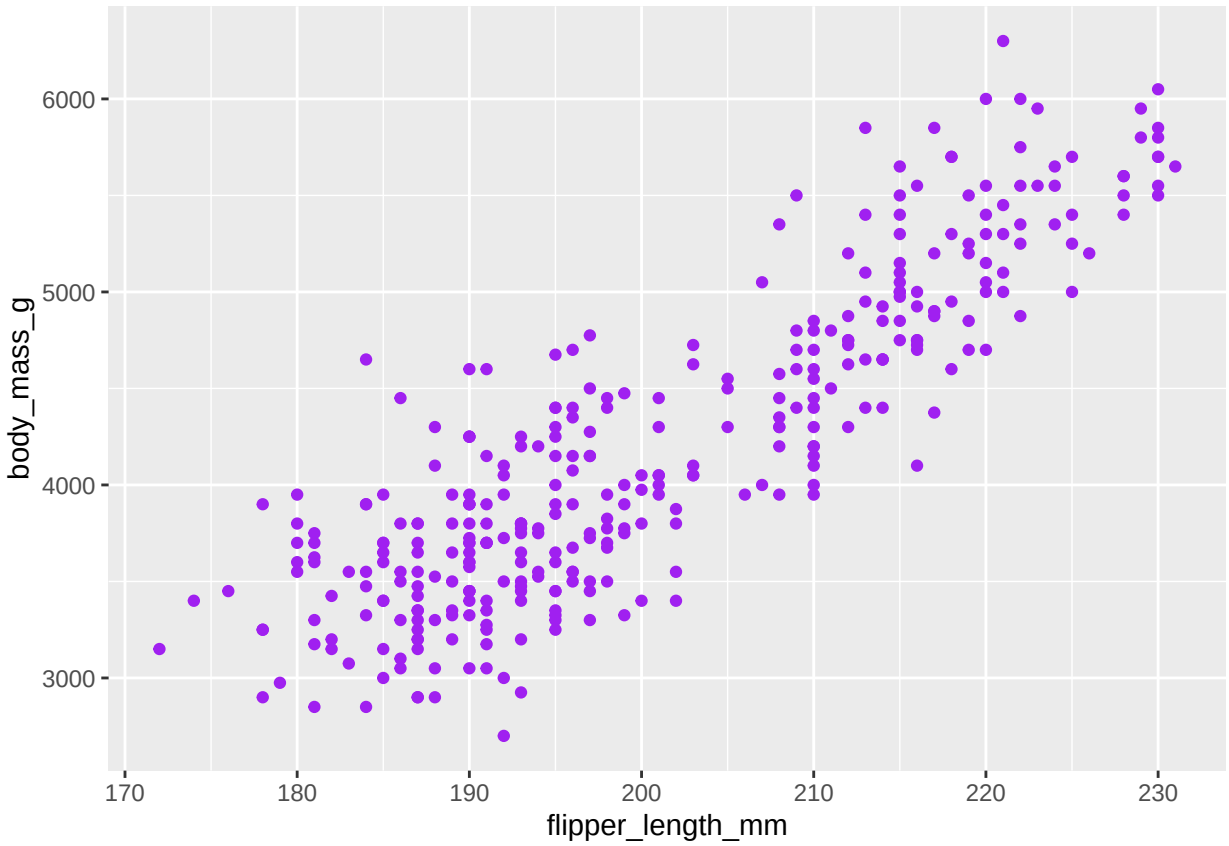
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



You can also add code outside the aes function. In this case, you are not mapping to the variable, but you are changing the overall plot.

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

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



Different Types of Geom

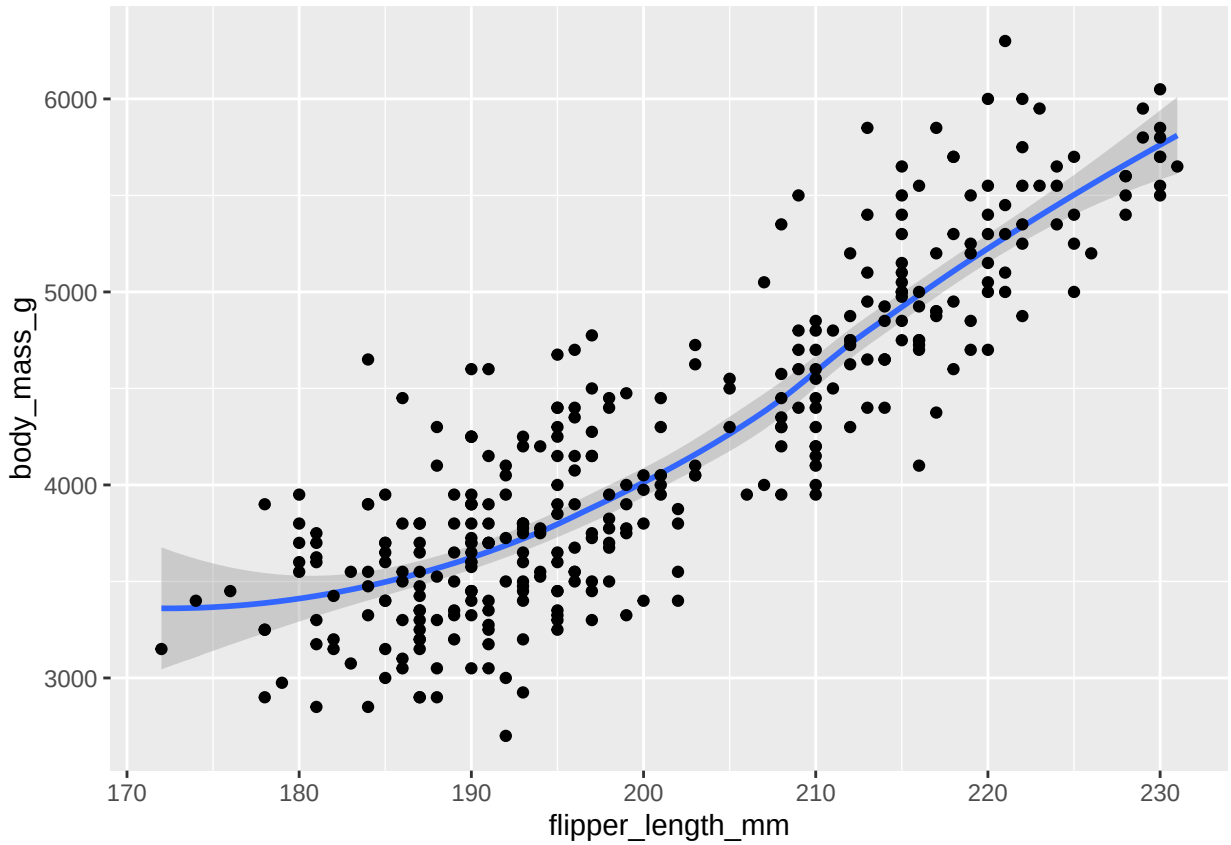
1. geom_point: scatter plot
2. geom_bar: bar charts
3. geom_line
4. geom_smooth: a smooth line with the existing data points
5. geom_jitter: when the data is hard to recognize, use it for better vizulization

```
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))
```

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

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

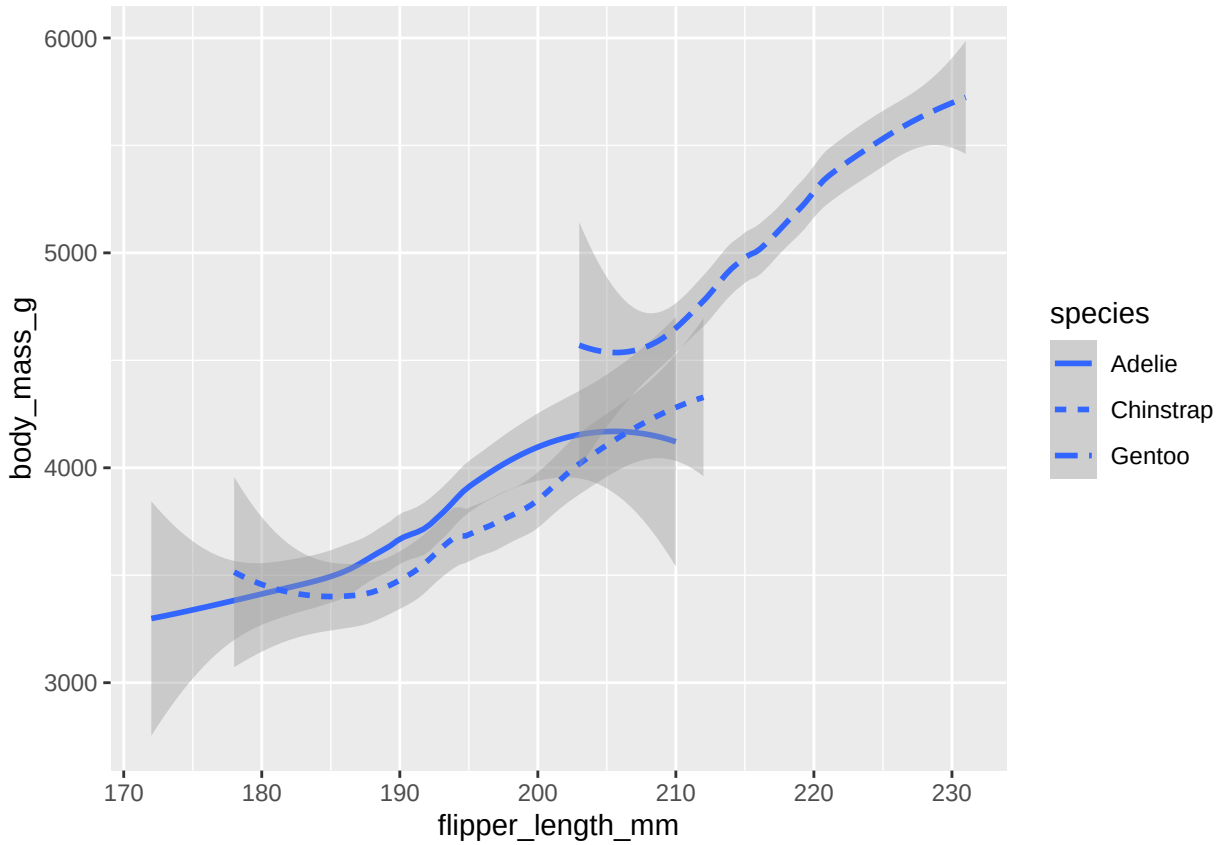
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



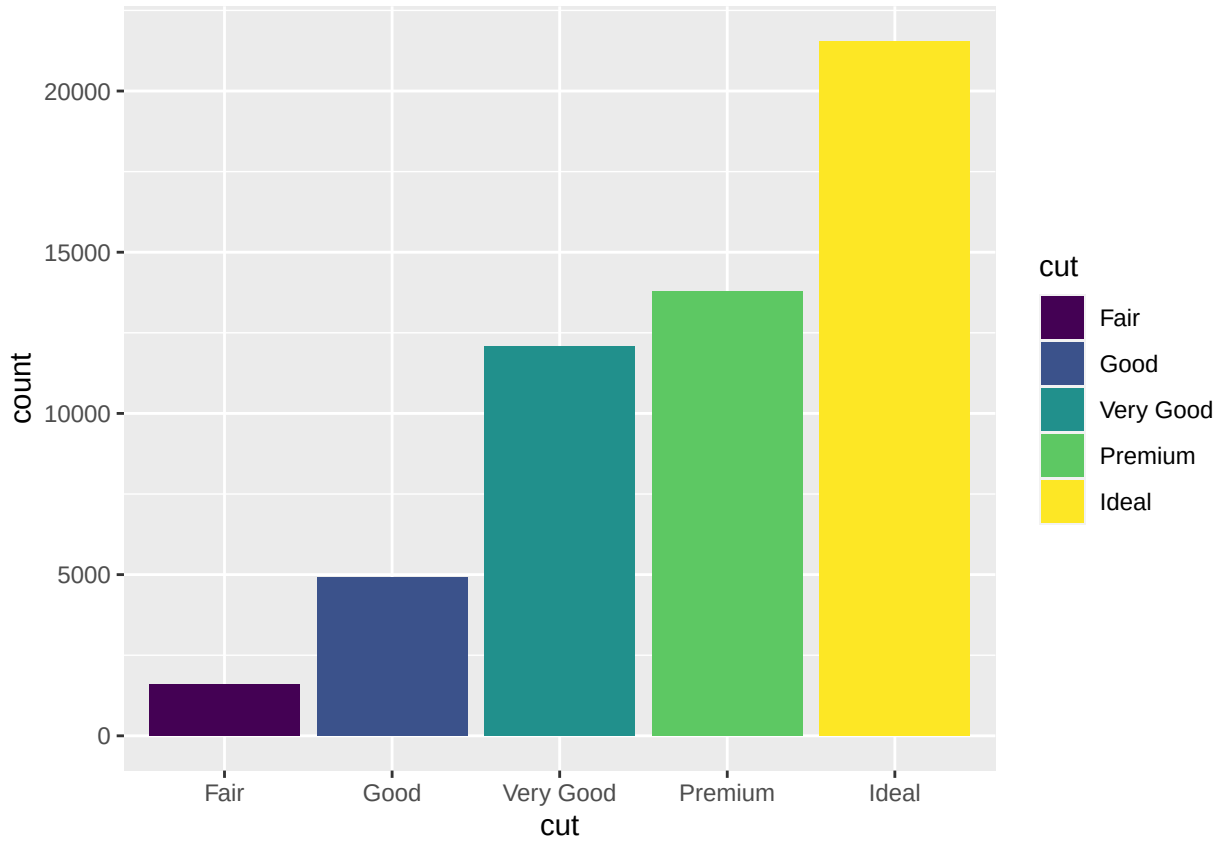

```
ggplot(data=penguins) +
  geom_smooth(mapping=aes(x=flipper_length_mm,y=body_mass_g,linetype=species))
```

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

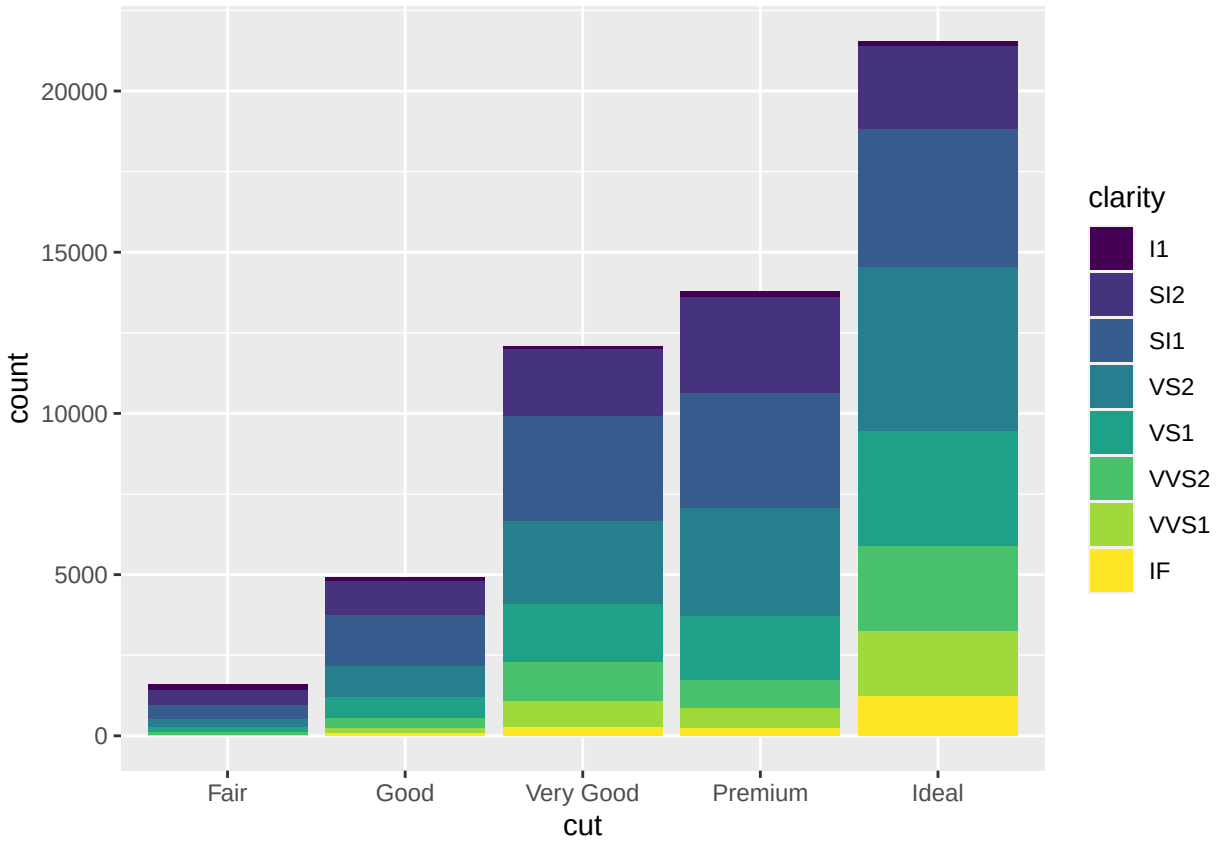
```
## Warning: Removed 2 rows containing non-finite values (stat_smooth).
```




```
ggplot(data=diamonds) +
  geom_bar(mapping=aes(x=cut,fill=cut))
```




```
ggplot(data=diamonds) +
  geom_bar(mapping=aes(x=cut,fill=clarity))
```

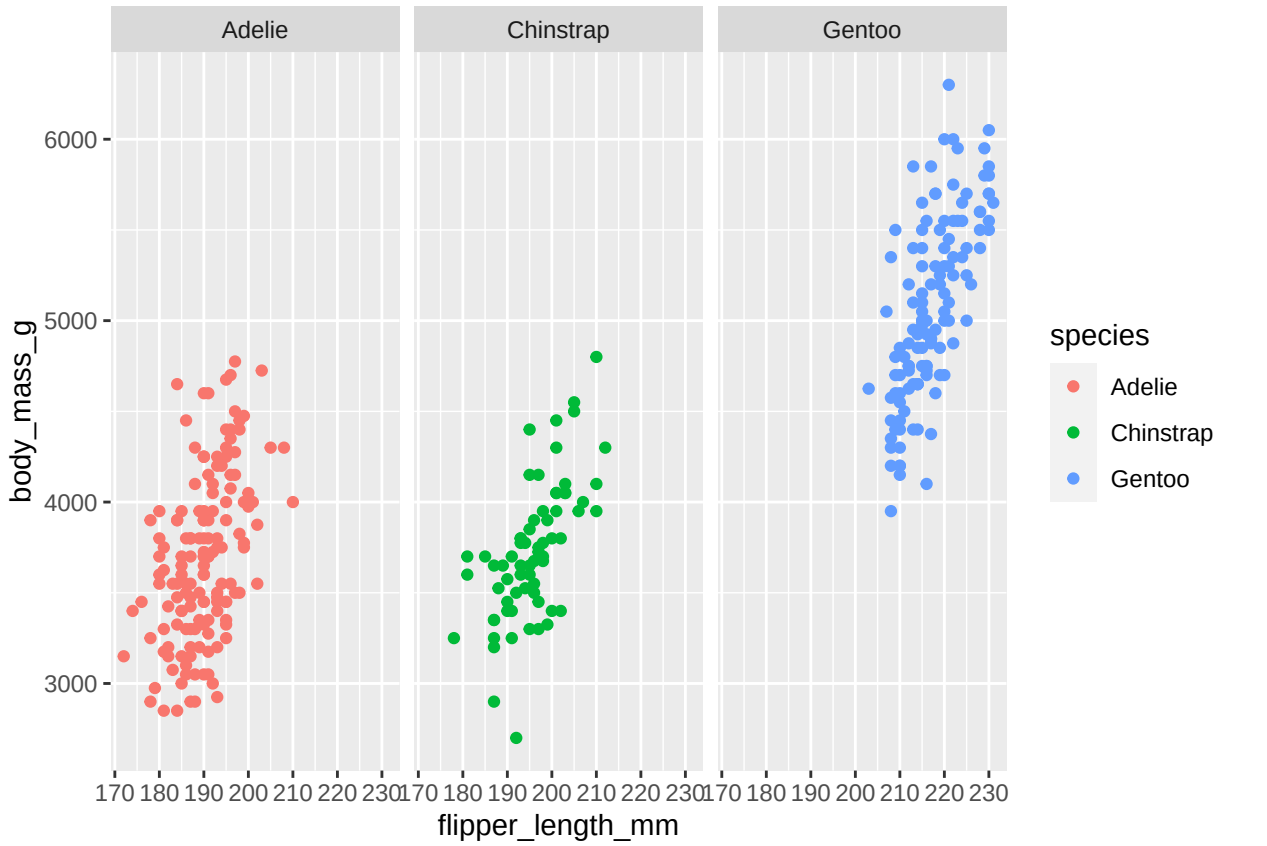


Aesthetics and Facets

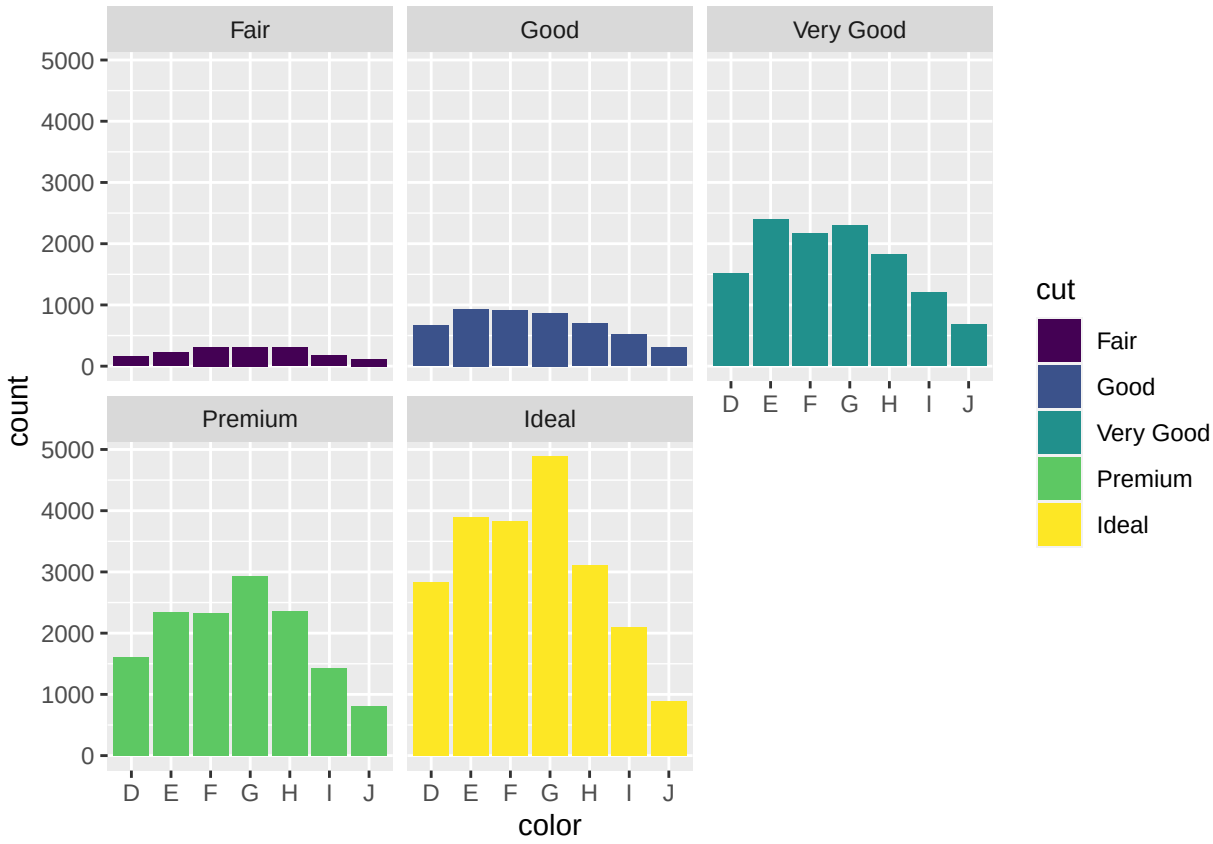
1. facet_wrap: facet the plot with a single variable
2. facet_grid: facet the plot with two variables

```
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).
```

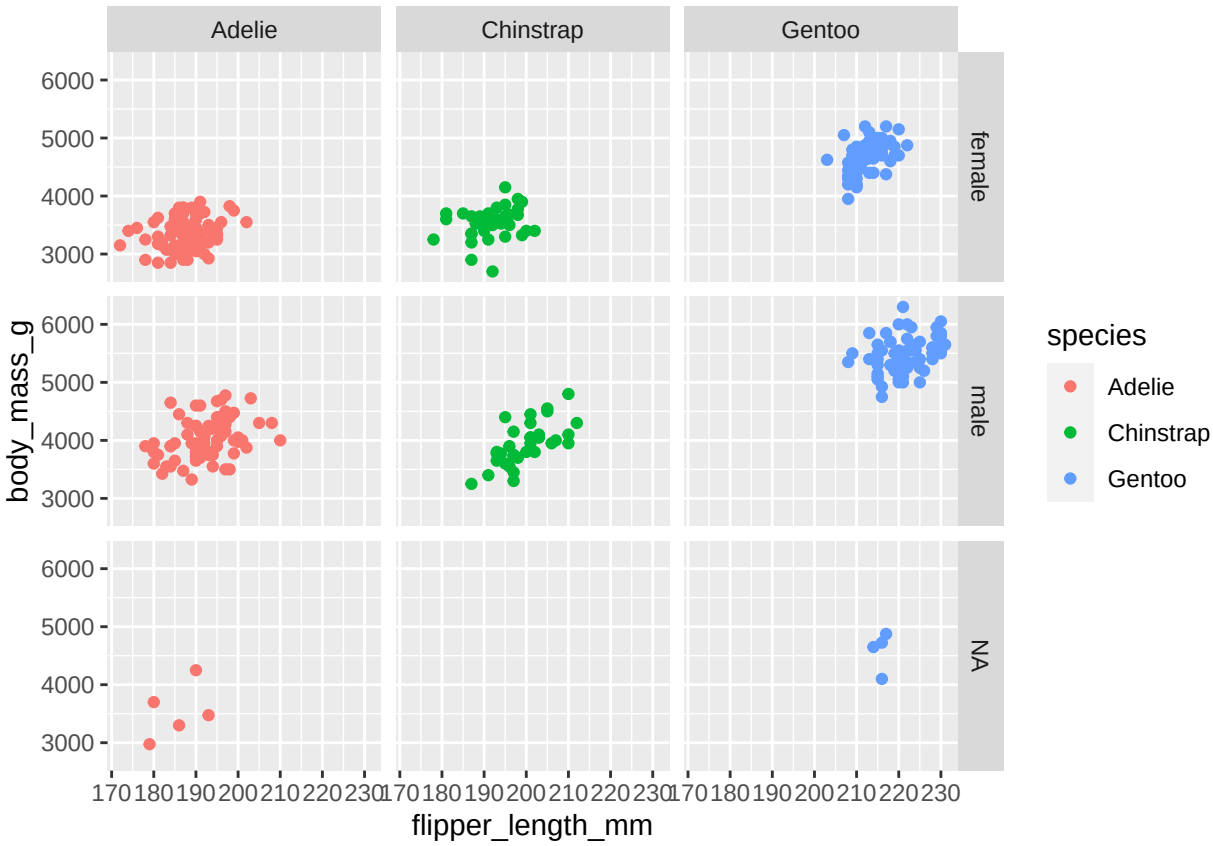



```
ggplot(data=diamonds) +
  geom_bar(mapping=aes(x=color,fill=cut)) +
  facet_wrap(~cut)
```




```
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).
```

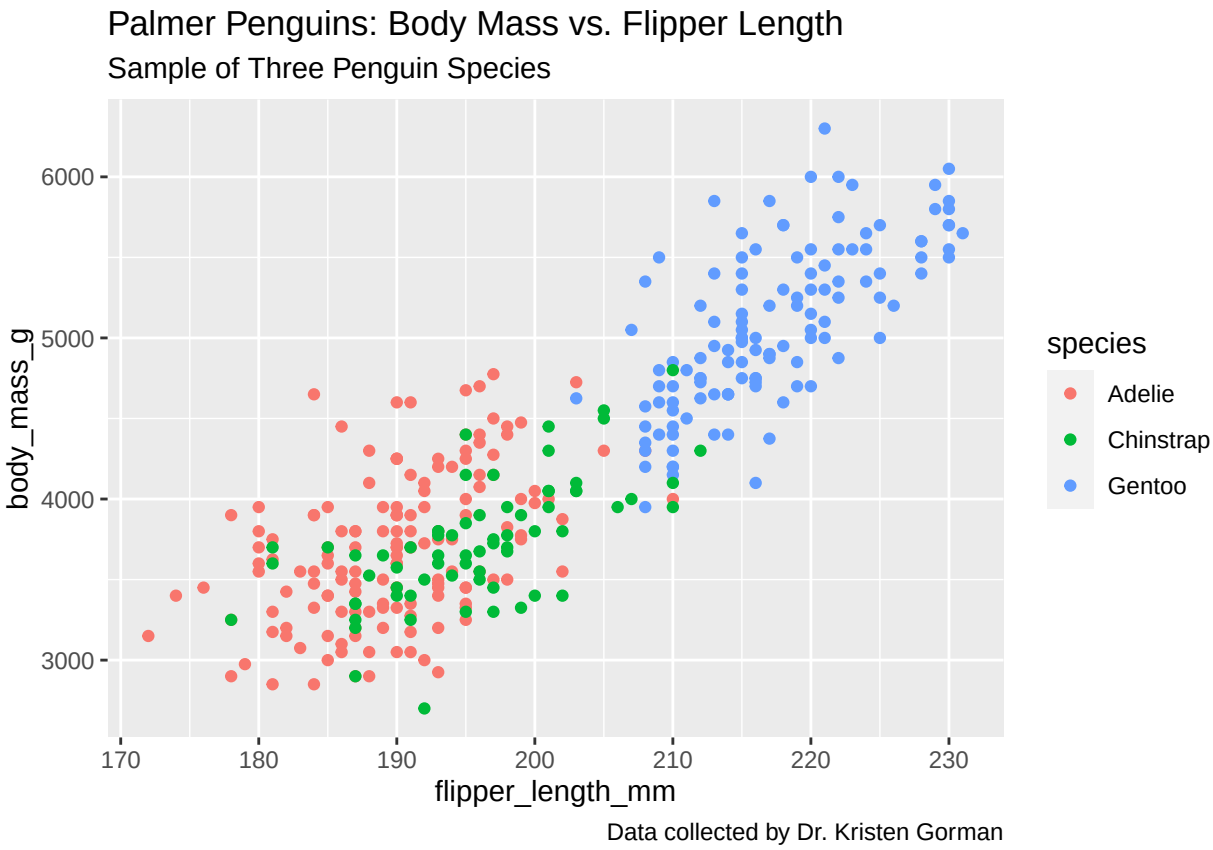


Annotation on the Plot

Adding titles, subtitles, and captions to 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"
```

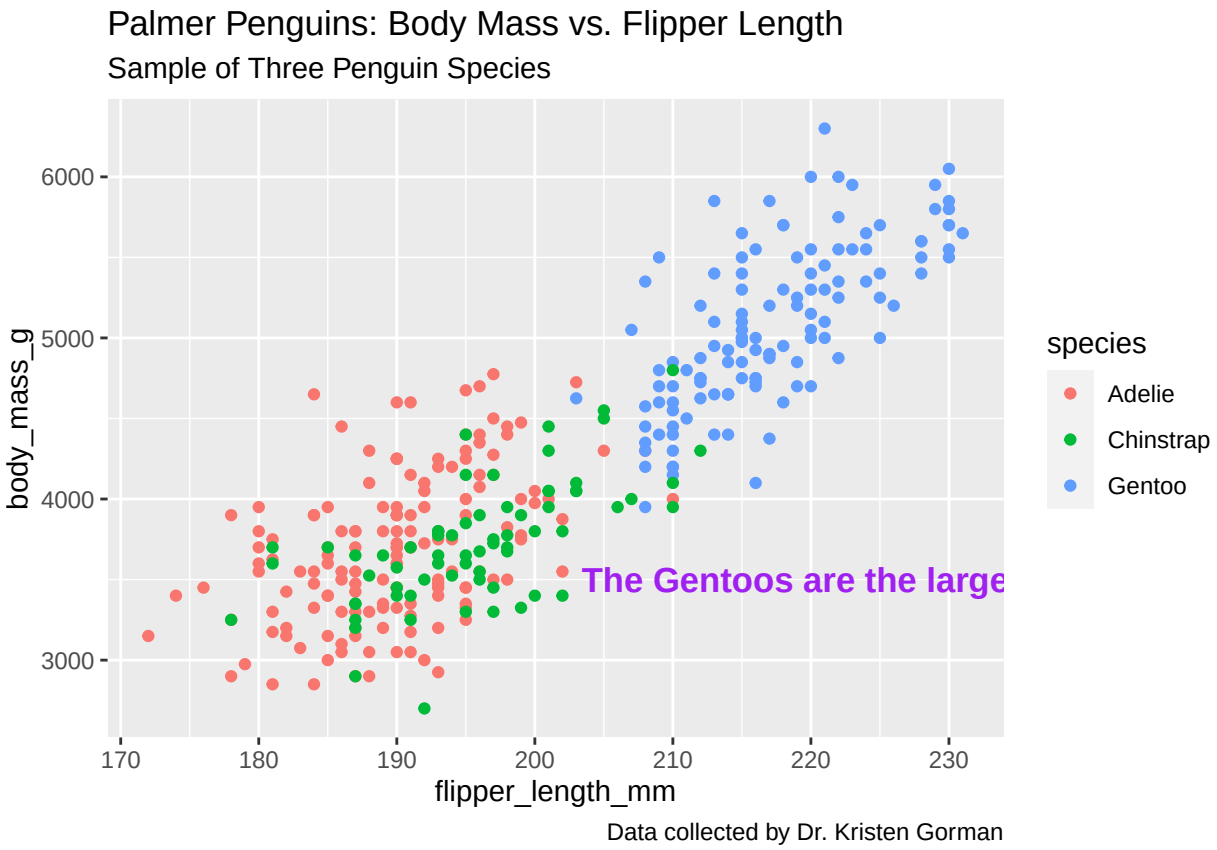
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



Adding annotation to 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"
  annotate("text",x=220,y=3500,label="The Gentoos are the largest",color="purple",fontface="bold",size=
```

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

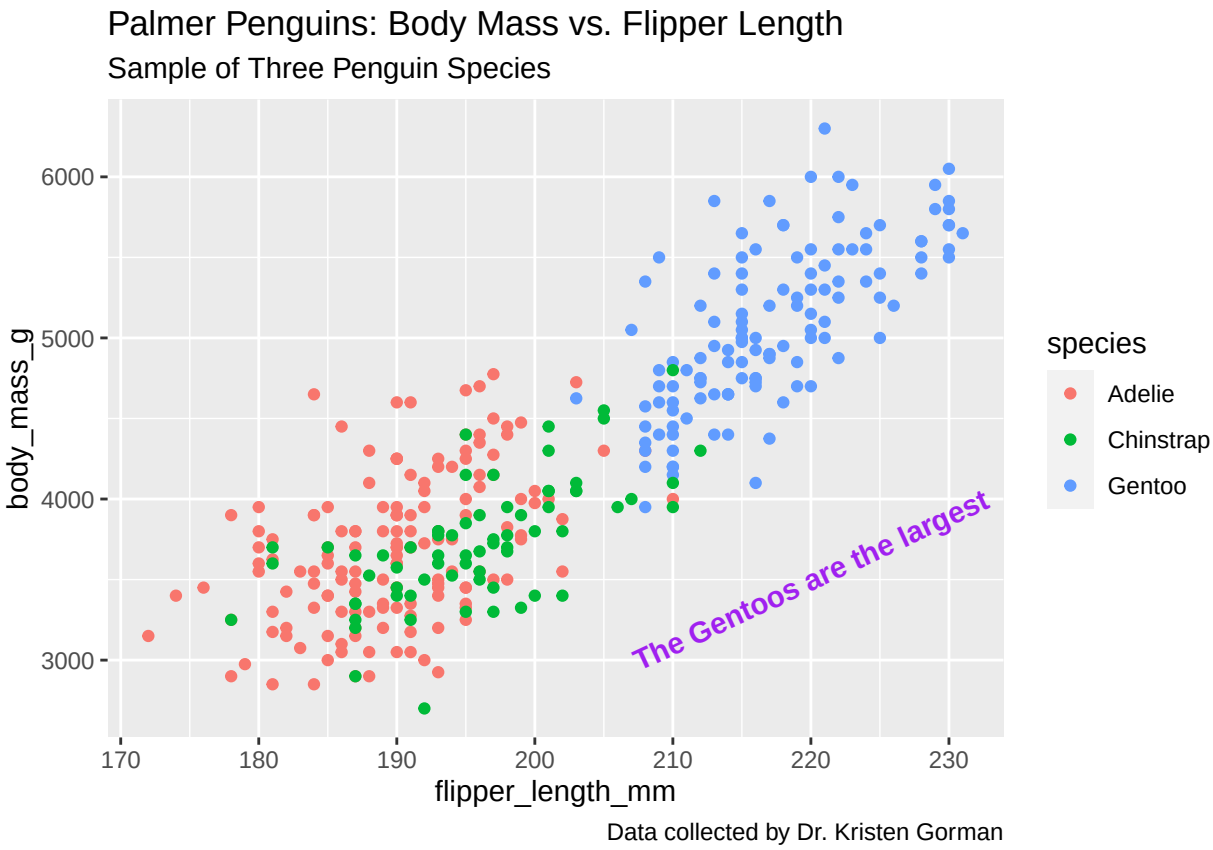


You can assign your code to variables, so you do not need to type super long codes.

```
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"

p + annotate("text",x=220,y=3500,label="The Gentoos are the largest",color="purple",fontface="bold",ang
```

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



Save the Plots

Use ggsave("name") to save plots.