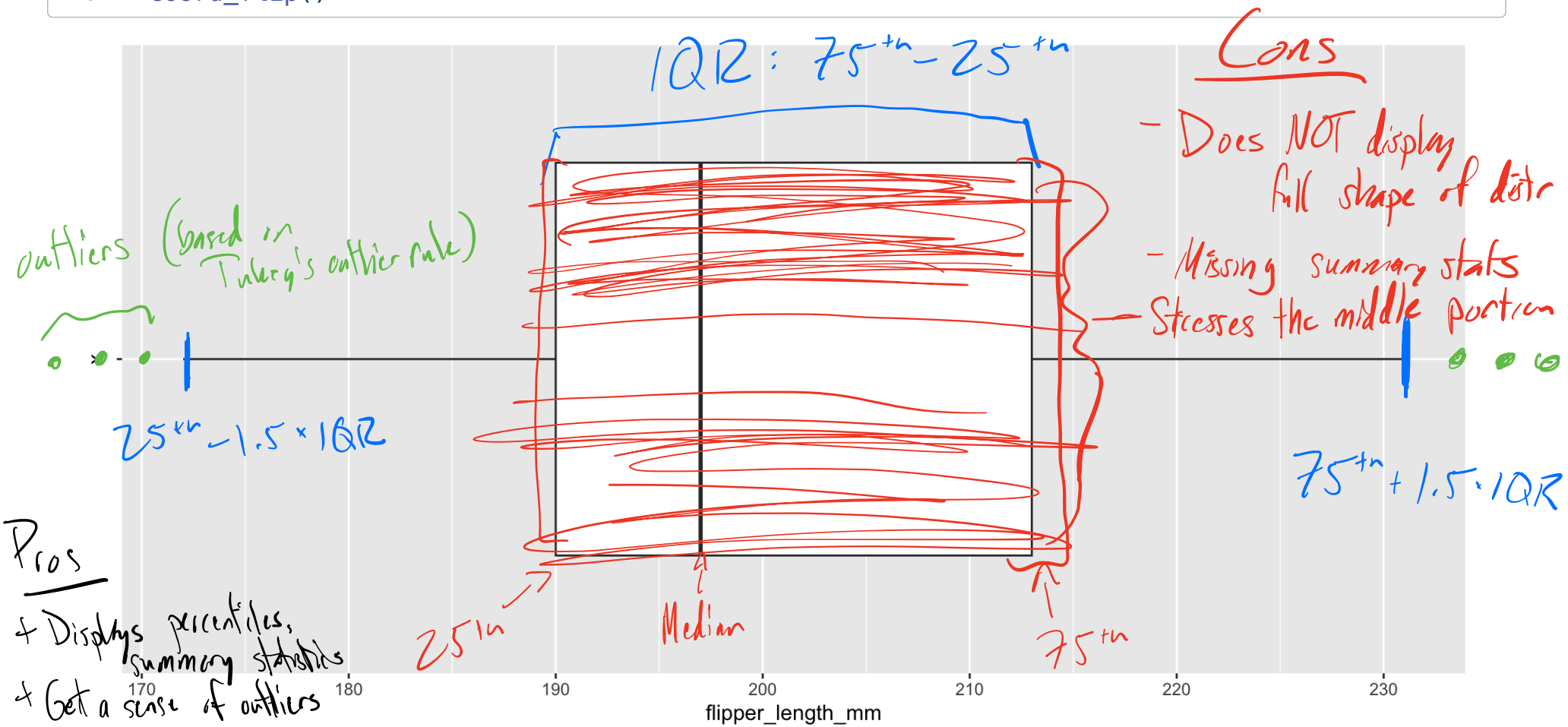
# Box plots visualize summary statistics

```
1 penguins |>
2   ggplot(aes(y = flipper_length_mm)) +
3   geom_boxplot(aes(x = "")) +
4   coord_flip()
```

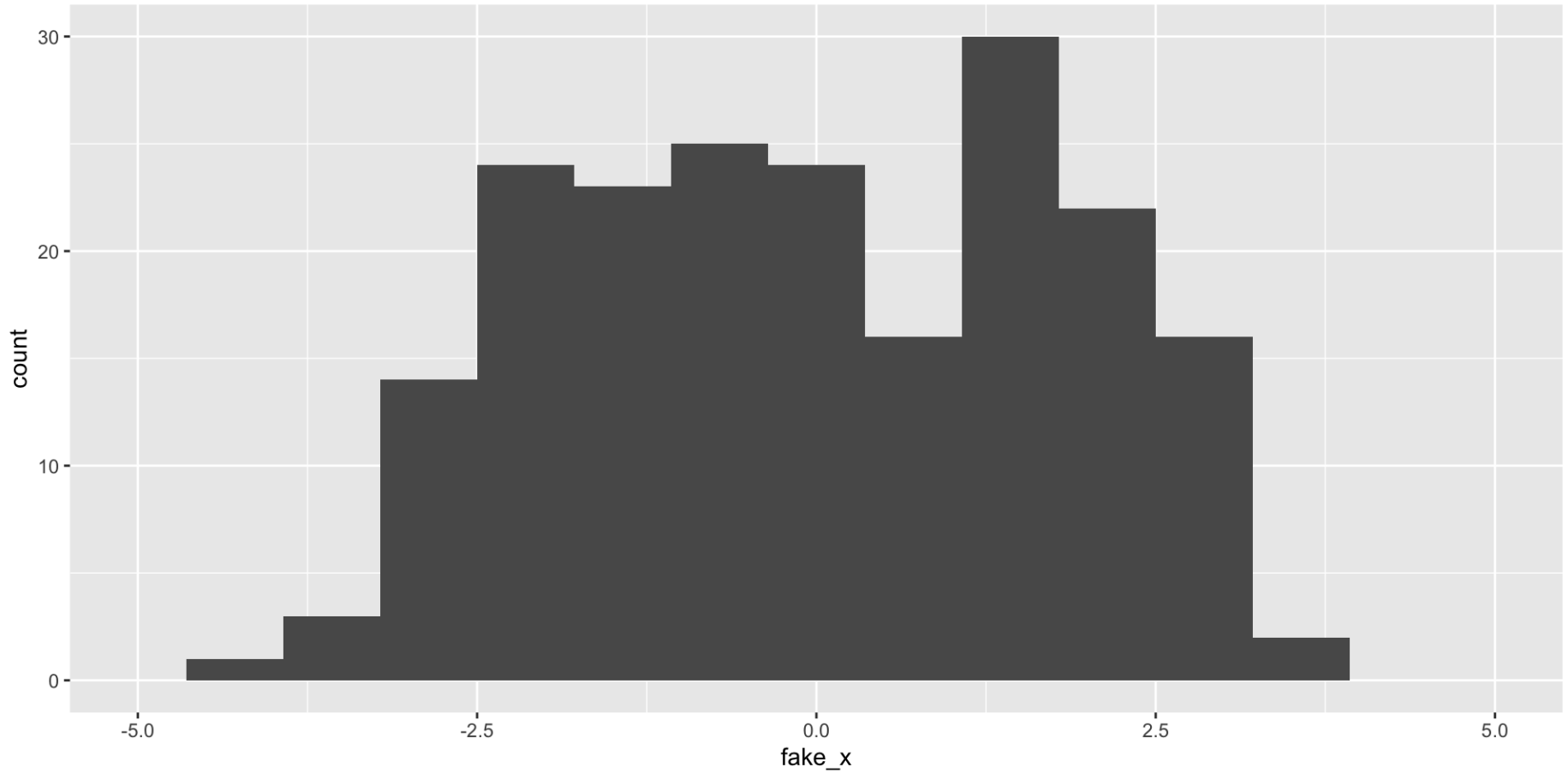



Pros

+ Displays percentiles, summary statistics
+ Get a sense of outliers
+ Useful for side-by-side

# What happens as we change the number of bins?

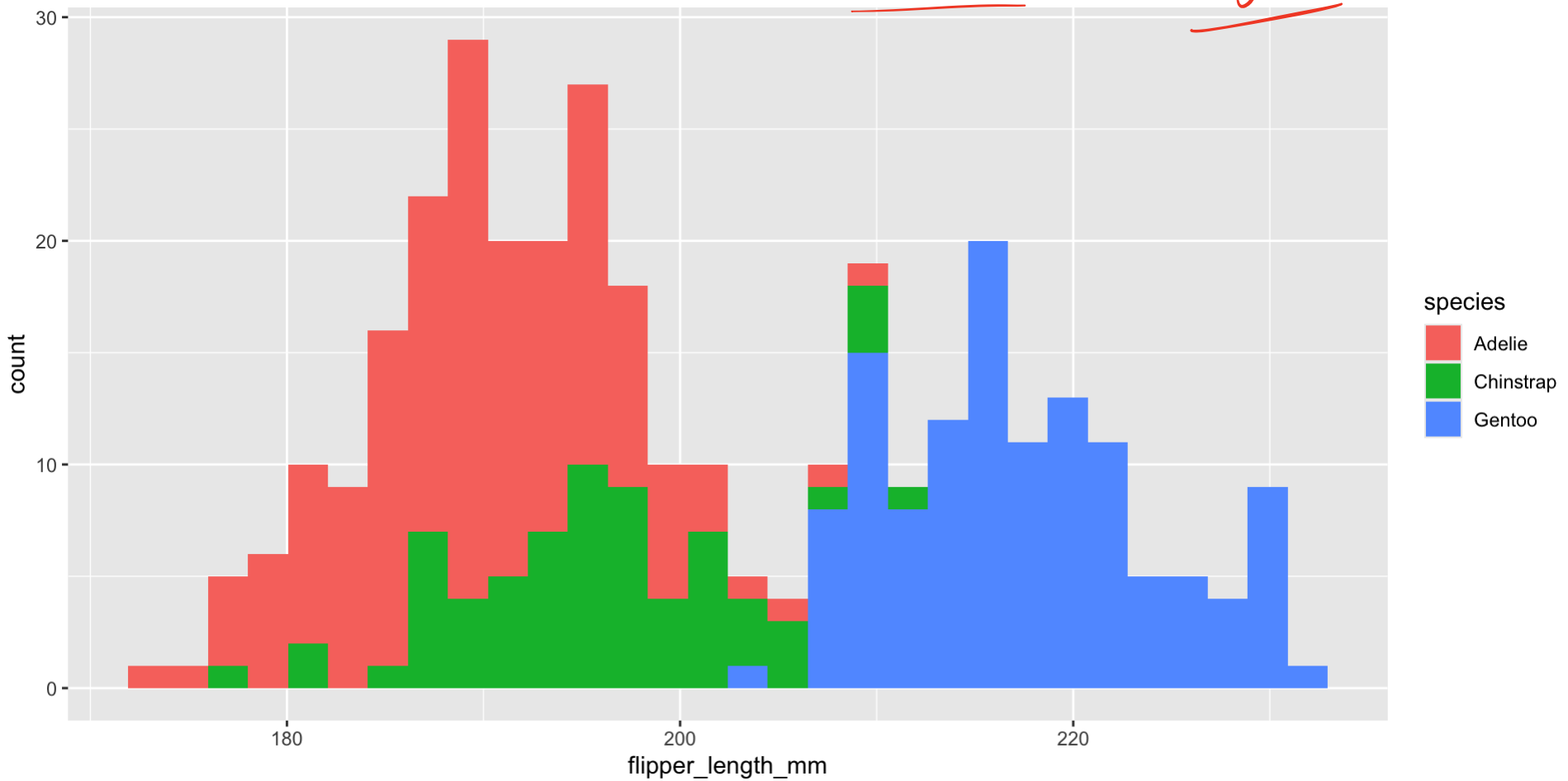
```
fake_data |>
  ggplot(aes(x = fake_x)) +
  geom_histogram(bins = 15) +
  scale_x_continuous(limits = c(-5, 5))
```



# What about displaying conditional distributions?

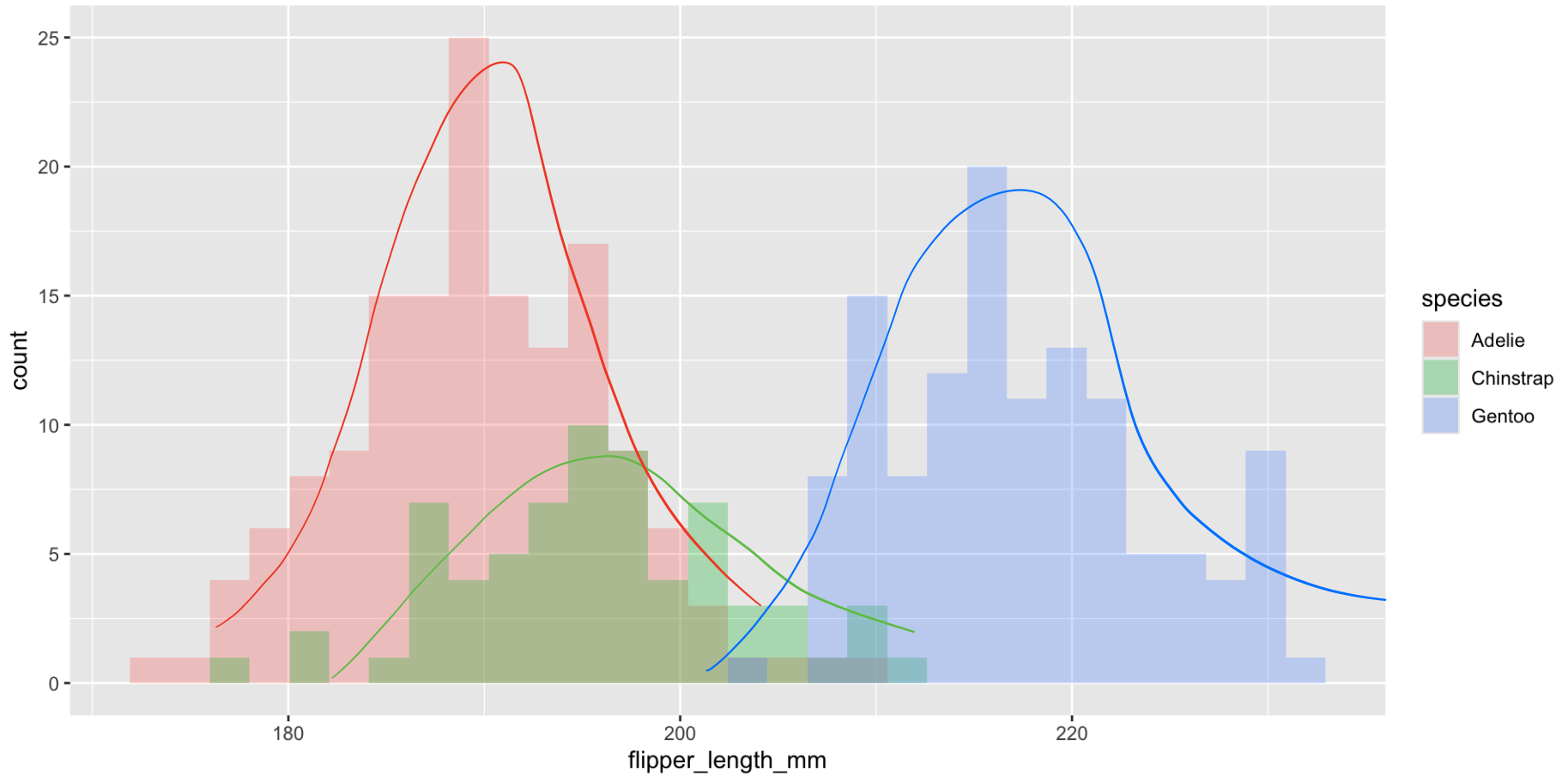
```
1 penguins |>
2   ggplot(aes(x = flipper_length_mm)) +
3   geom_histogram(aes(fill = species))
```

→ Stacking → Marginal



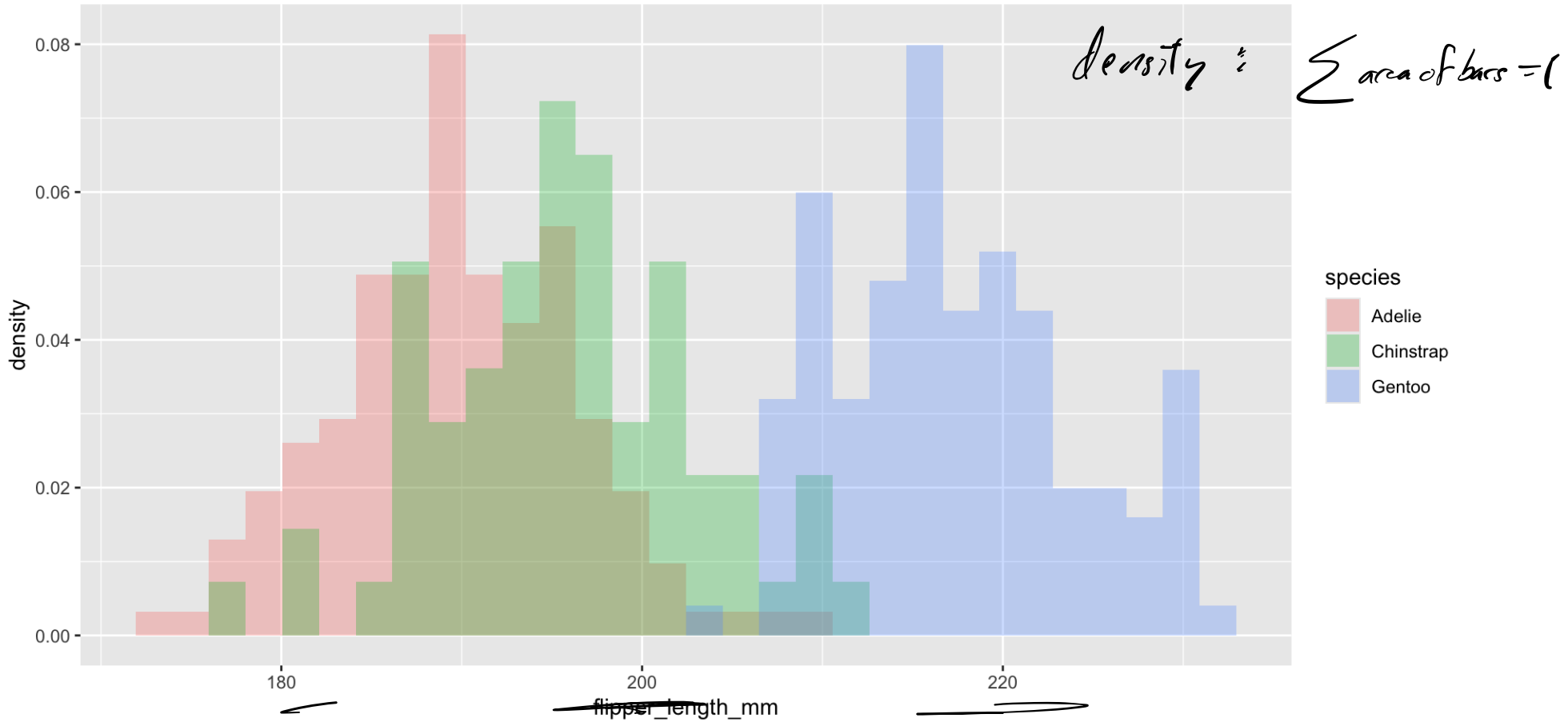
# What about displaying conditional distributions?

```
penguins |>
  ggplot(aes(x = flipper_length_mm)) +
  geom_histogram(aes(fill = species),
                 position = "identity", alpha = 0.3)
```



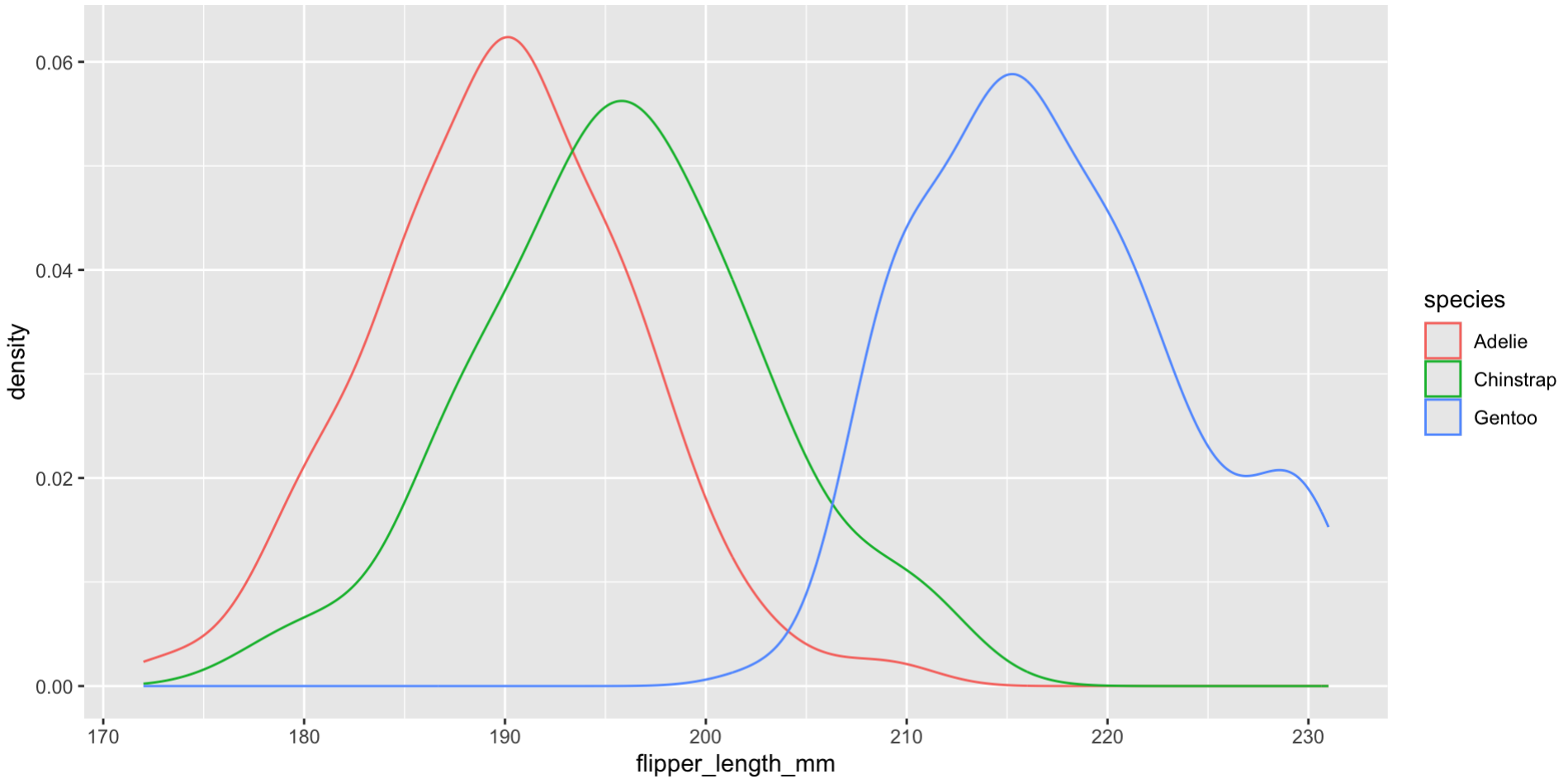
# Normalize histogram frequencies with density

```
penguins |>
  ggplot(aes(x = flipper_length_mm)) +
  geom_histogram(aes(y = after_stat(density), fill = species),
                 position = "identity", alpha = 0.3)
```



# Can use density curves instead

```
1 penguins |>
2   ggplot(aes(x = flipper_length_mm)) +
3   geom_density(aes(color = species))
```



# We should NOT fill the density curves

```
1 penguins |>
2   ggplot(aes(x = flipper_length_mm)) +
3   geom_density(aes(fill = species), alpha = .3)
```