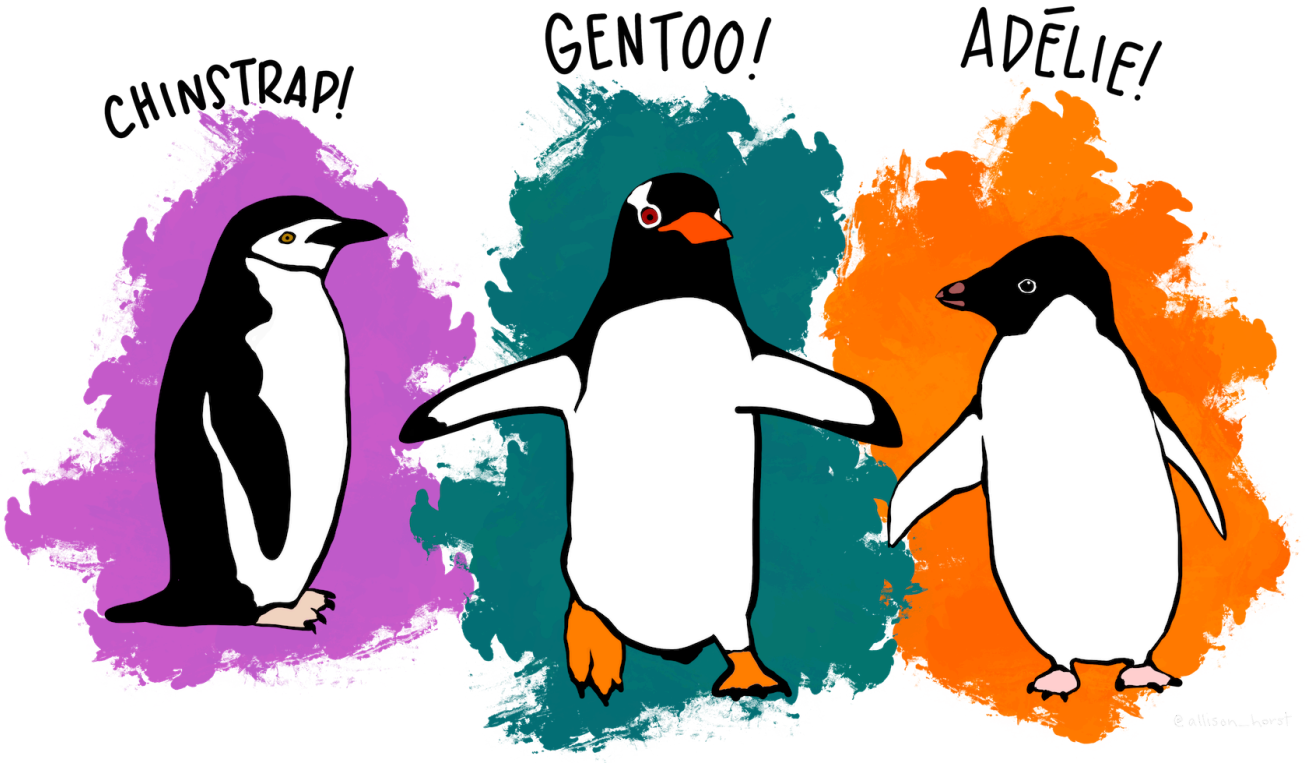
Data visualization

Agenda

- Today: data visualization with `ggplot`
  - Formalize what we've seen in class activities
- HW 1 due Friday
  - I'm available in office hours or by email if you have any questions

Data for today



Artwork by @allison_horst

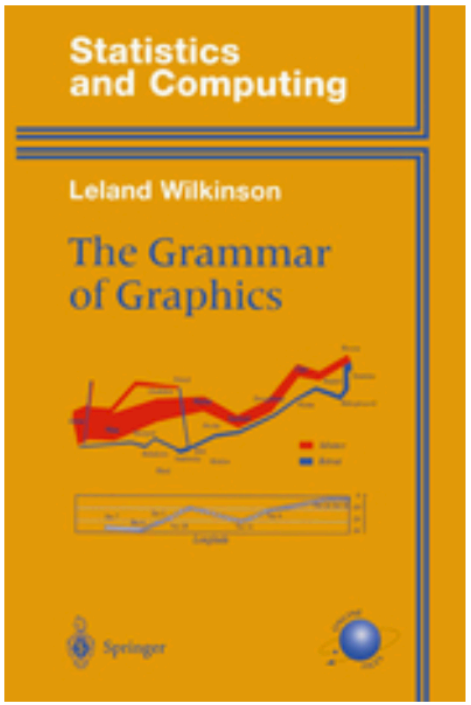
Data visualization with ggplot2

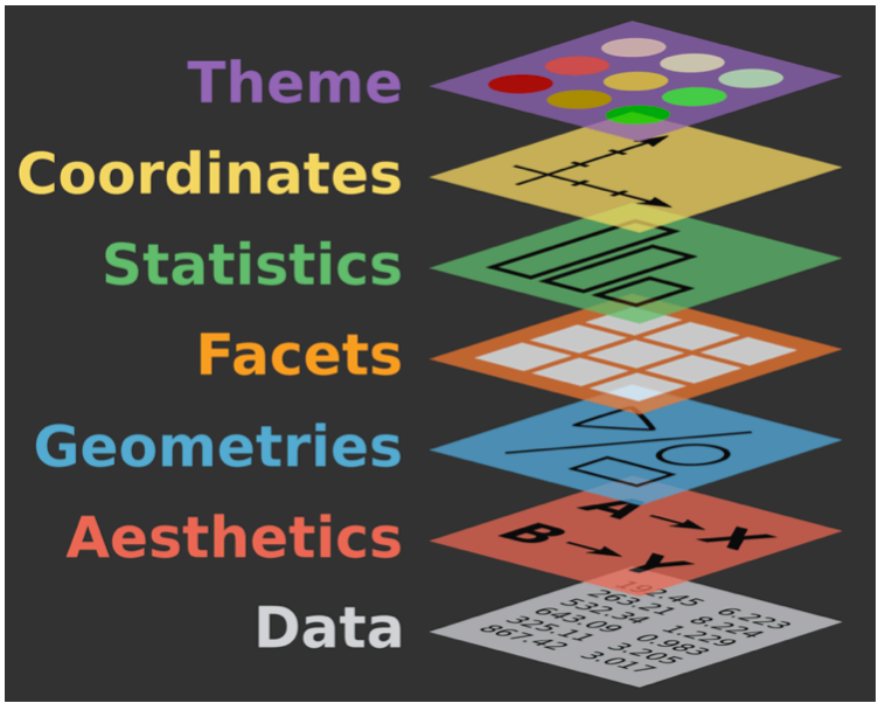
- `ggplot2`: the R package
- `ggplot`: the function (from `ggplot2`) used to make plots
- `gg` stands for "Grammar of Graphics"

Artwork by @allison_horst

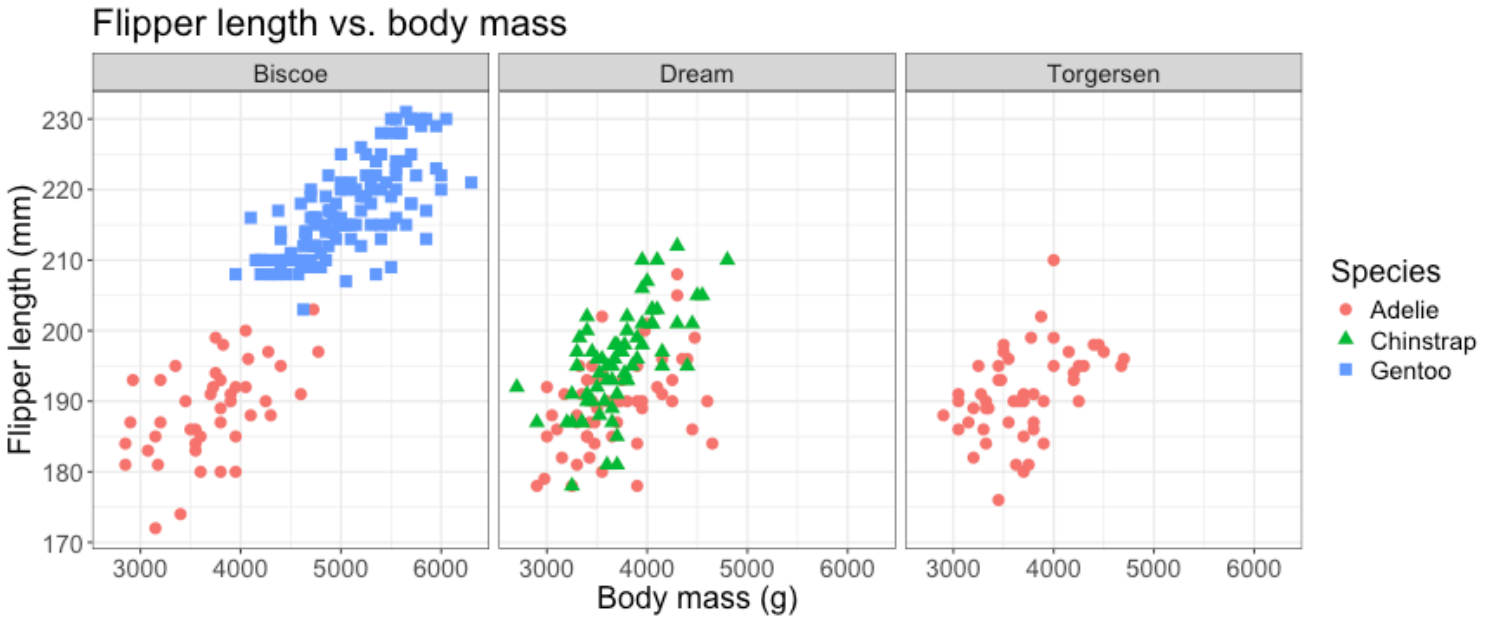
Grammar of Graphics

Build visualizations in layers:

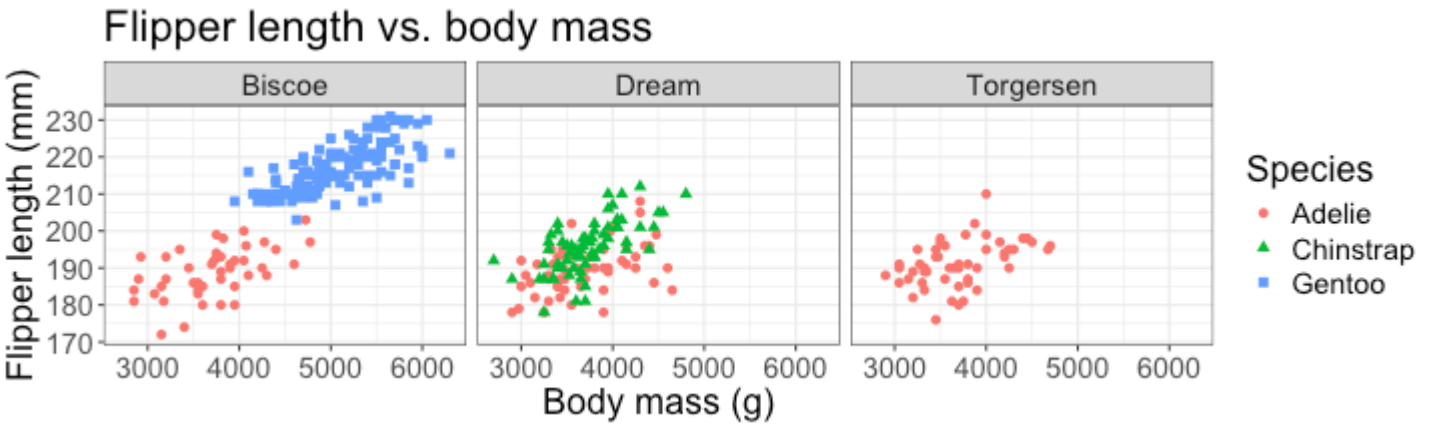




Plot to make

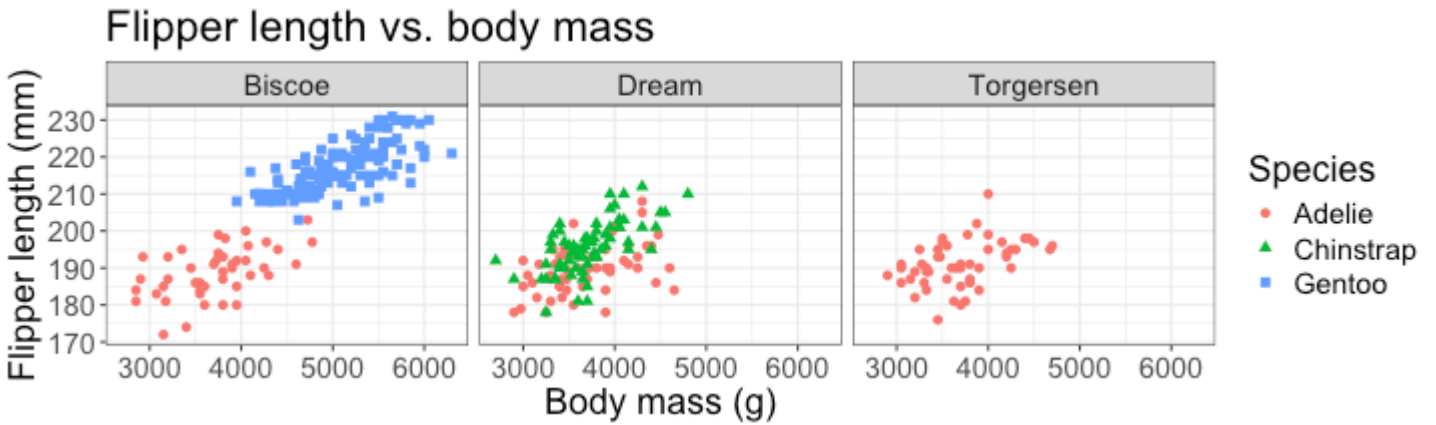

Flipper length vs. body mass
Biscoe
Dream
Torgersen
Flipper length (mm)
230
220
210
200
190
180
170
3000
4000
5000
6000
Body mass (g)
Species
Adelie
Chinstrap
Gentoo

Layer 1: the data



Data: Which data do I want to plot?

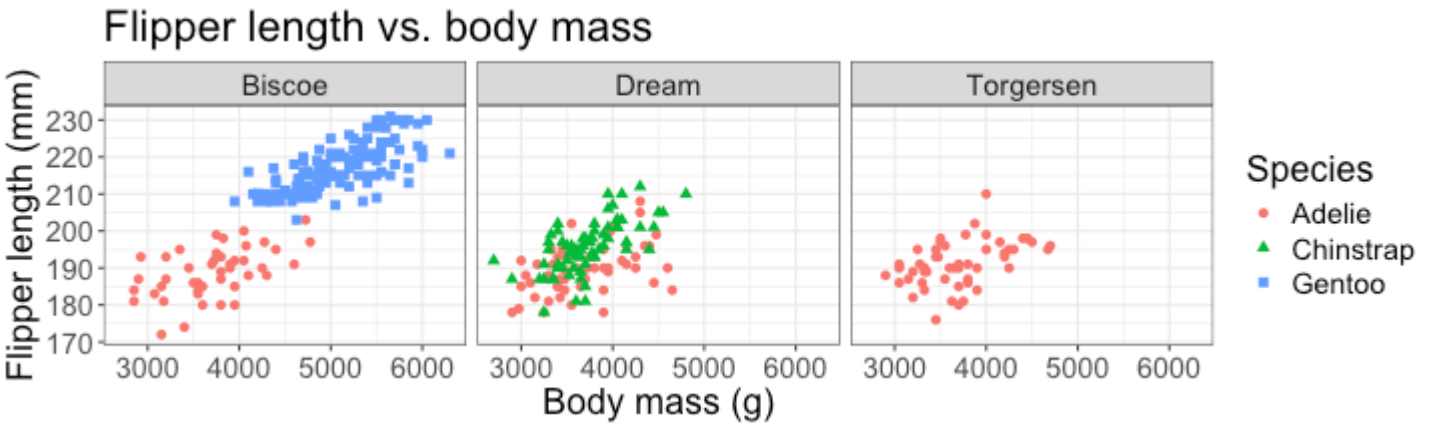
Layer 1: the data



Data: Which data do I want to plot?

```
penguins
```

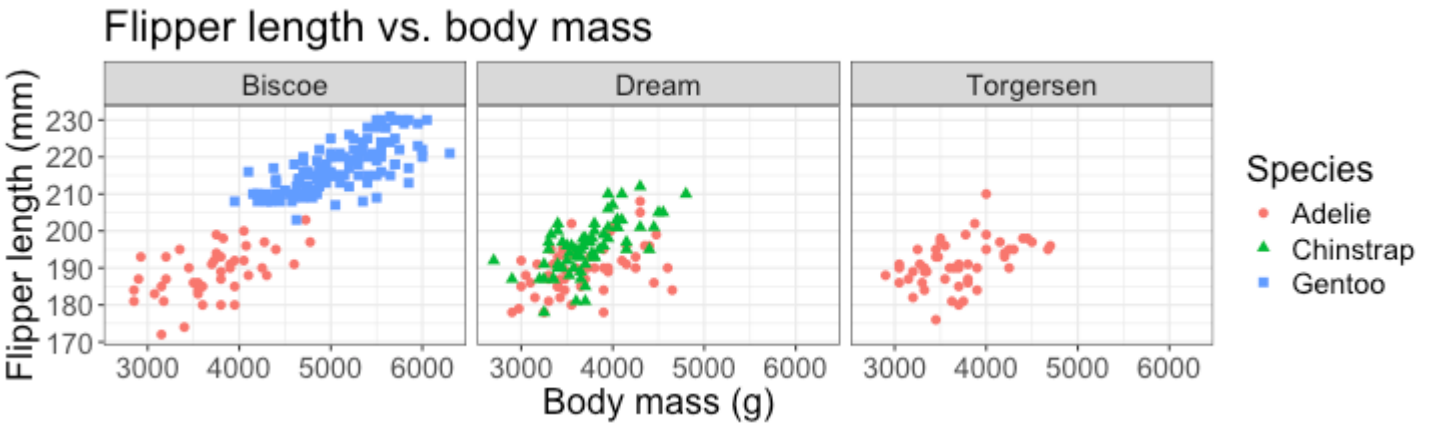
Layer 2: Aesthetics



Aesthetics: mapping features of the plot to variables in the data

> Which variables from the data do I want to plot?

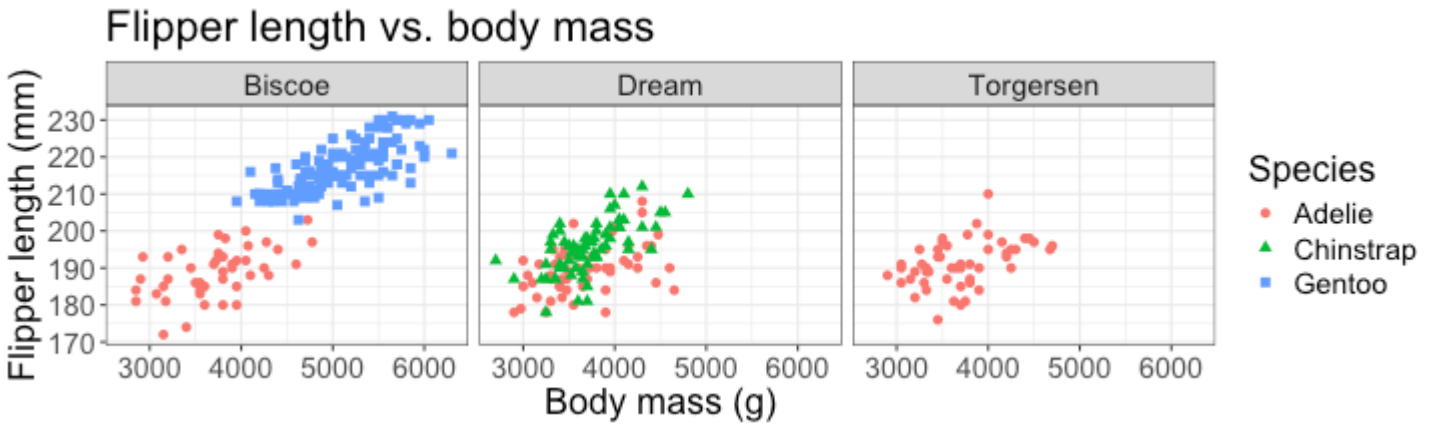
Layer 2: Aesthetics



Aesthetics: mapping features of the plot to variables in the data

```
penguins |>
  ggplot(aes(x = body_mass_g,
             y = flipper_length_mm,
             color = species,
             shape = species))
```

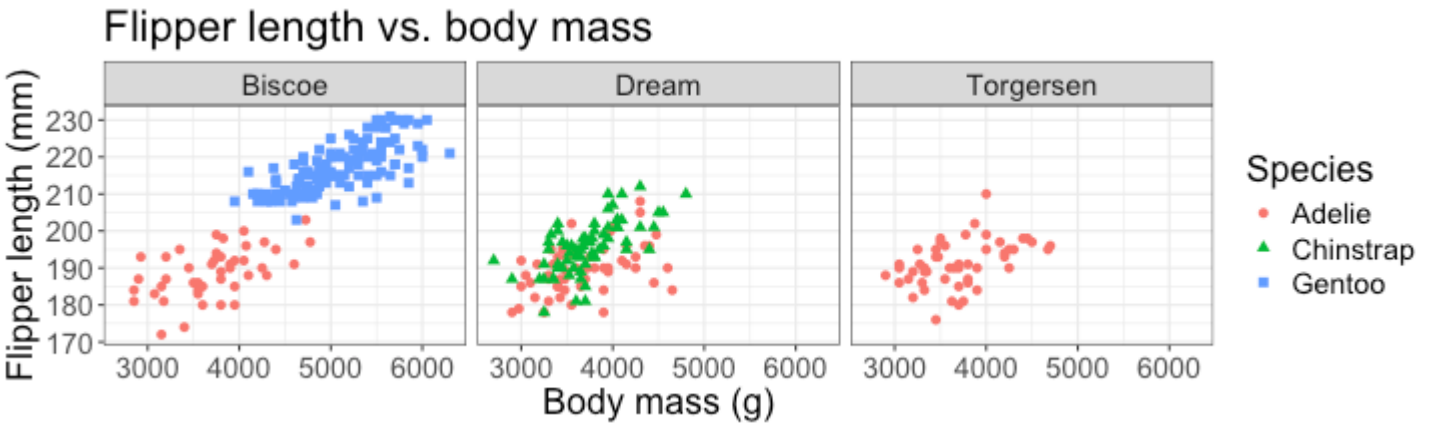
Layer 3: Geometric objects



Geometric objects: objects we use to visualize the data

> What objects do we use to represent the penguins on this plot?

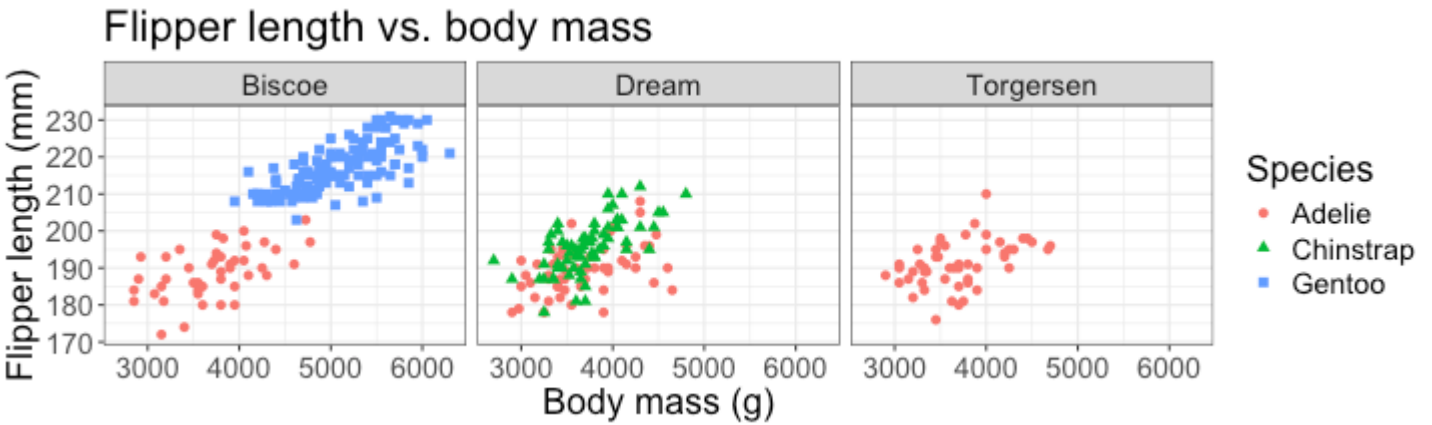
Layer 3: Geometric objects



Geometric objects: objects we use to visualize the data

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

Layer 3: Geometric objects

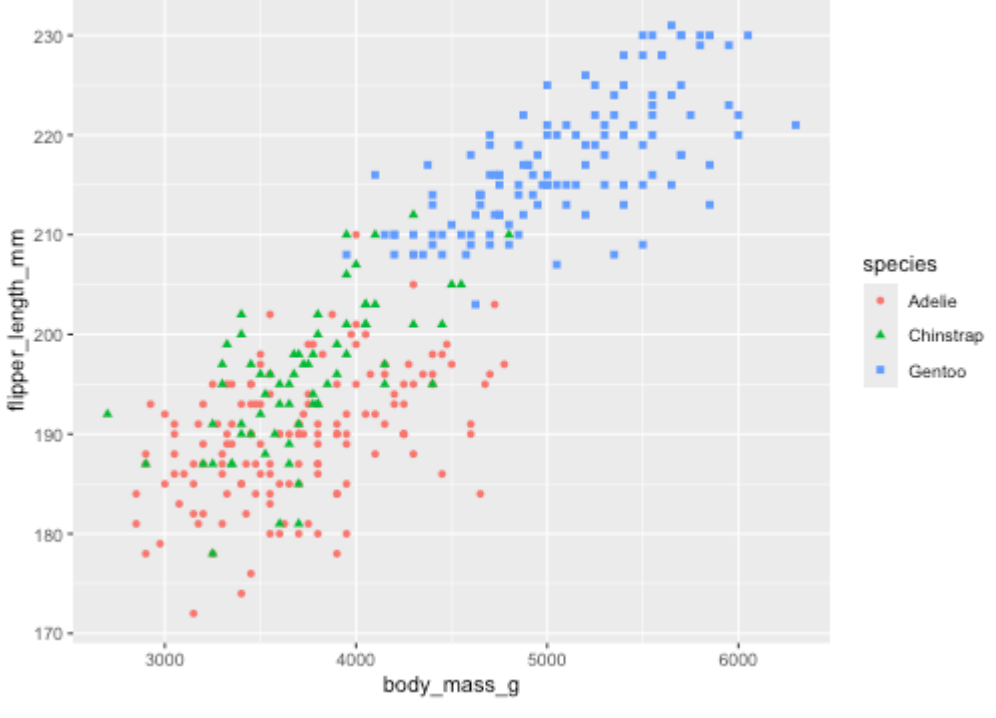



```
penguins |>
  ggplot(aes(x = body_mass_g,
             y = flipper_length_mm,
             color = species,
             shape = species)) +
  geom_point()
```

Other geometric objects include `geom_histogram`, `geom_boxplot`, `geom_bar`, `geom_smooth`, and `geom_line`

What we have so far

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

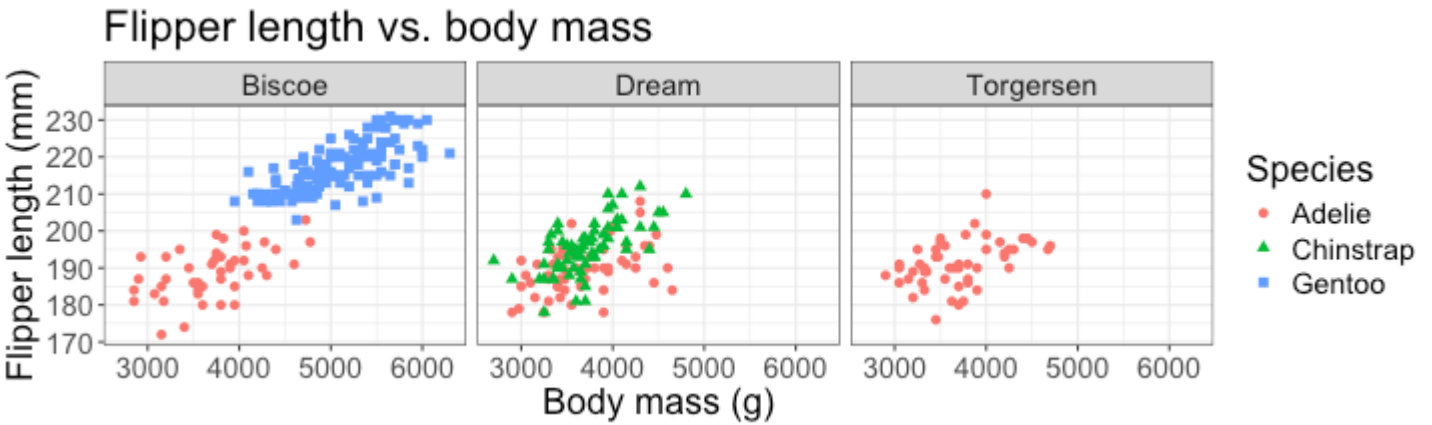


What we have so far

Still need to:

- Divide the plot by island (Biscoe, Dream, and Torgersen)
- Add labels and a title
- Change the background

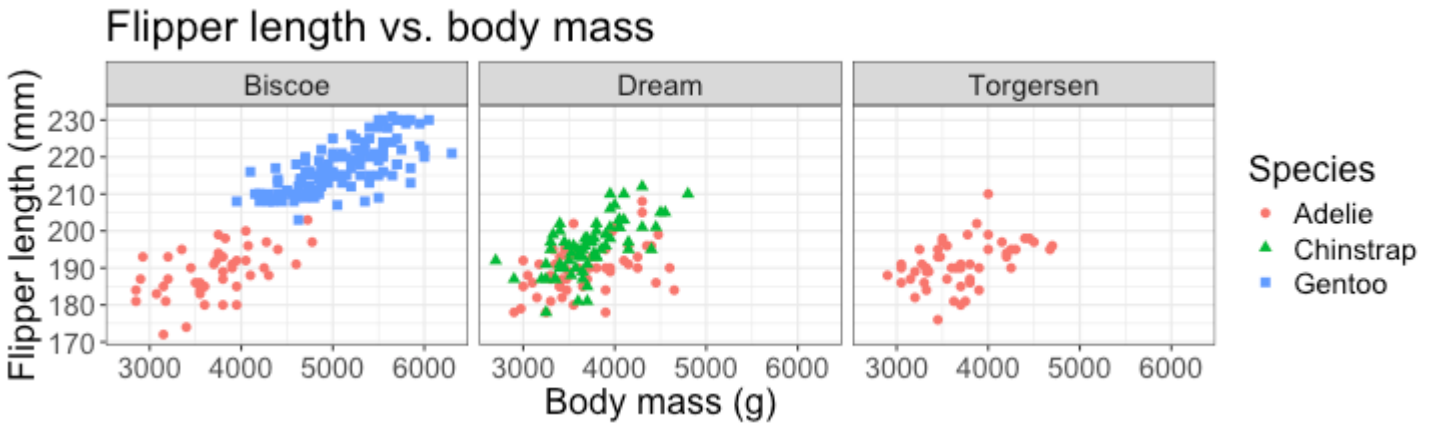
Layer 4: Faceting



Facets: split visualization by the value of another variable

```
penguins |>
  ggplot(aes(x = body_mass_g,
             y = flipper_length_mm,
             color = species,
             shape = species)) +
  geom_point() +
  facet_wrap(~island)
```

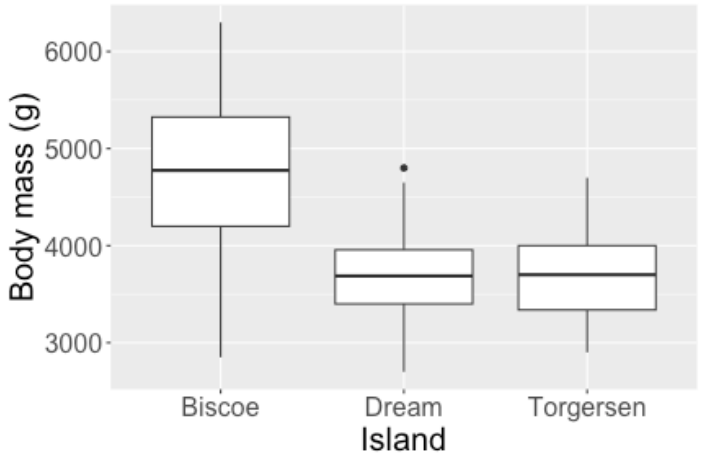
Making the plot look nice: labels and theme




```
penguins |>
  ggplot(aes(x = body_mass_g,
             y = flipper_length_mm,
             color = species,
             shape = species)) +
  geom_point() +
  facet_wrap(~island) +
  labs(x = "Body mass (g)",
       y = "Flipper length (mm)",
       color = "Species",
       shape = "Species",
       title = "Flipper length vs. body mass") +
  theme_bw()
```

Concept check

> Which code created this plot?



(A)
```
penguins |>
  ggplot(aes(x = island,
             y = body_mass_g)) +
  geom_boxplot()
```

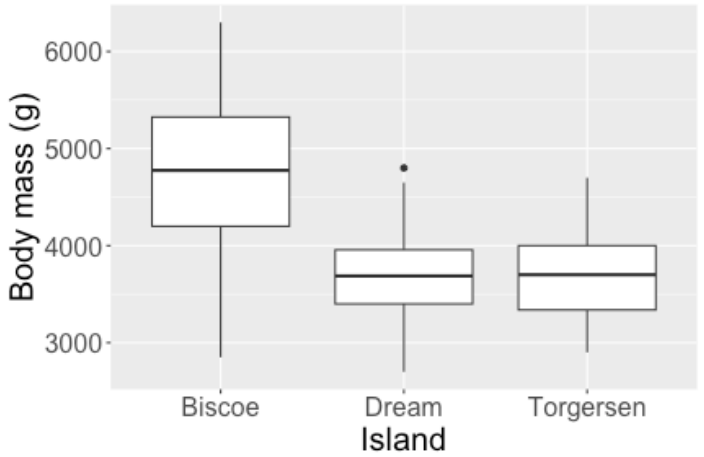
(B)
```
penguins |>
  ggplot(aes(x = island,
             y = body_mass_g)) +
  geom_boxplot() +
  labs(x = "Island",
       y = "Body mass (g)")
```

(C)
```
penguins |>
  geom_boxplot()
```

(D)
```
penguins |>
  ggplot(aes(x = island,
             y = body_mass_g)) +
  geom_histogram() +
```

Concept check

> Which code created this plot?



Answer: (B)

```
(A)  penguins |>
       ggplot(aes(x = island,
                  y = body_mass_g)) +
       geom_boxplot()
```

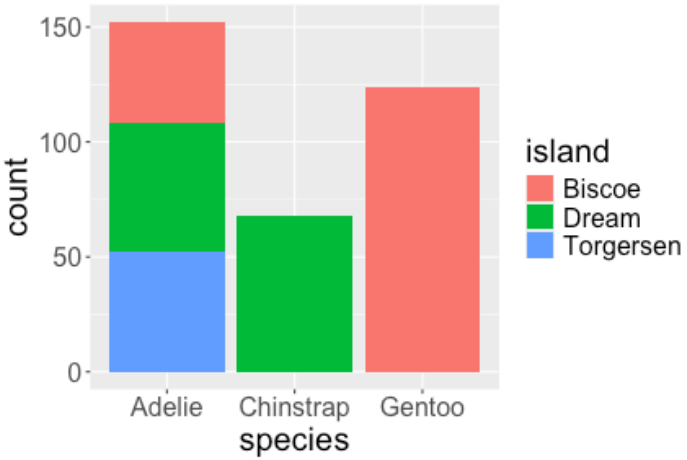
```
(B)  penguins |>
       ggplot(aes(x = island,
                  y = body_mass_g)) +
       geom_boxplot() +
       labs(x = "Island",
            y = "Body mass (g)")
```

```
(C) penguins |>
      geom_boxplot()
```

```
(D) penguins |>
      ggplot(aes(x = island,
                 y = body_mass_g)) +
      geom_histogram() +
```

Concept check

> Which code created this plot?



(A)
```
penguins |>
  ggplot(aes(x = species))
```

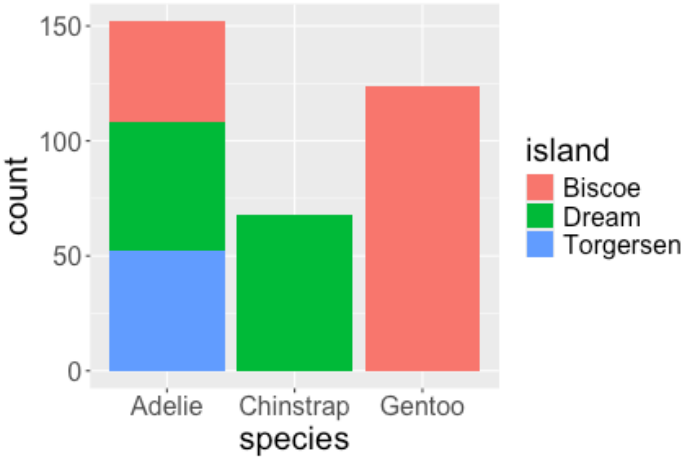
(B)
```
penguins |>
  ggplot(aes(x = species,
             fill = island)) +
  geom_point()
```

(C)
```
penguins |>
  ggplot(aes(x = species,
             fill = island)) +
  geom_bar()
```

(D)
```
penguins |>
  ggplot(aes(x = species)) +
  geom_bar()
```

Concept check

> Which code created this plot?




```
(A)  penguins |>
       ggplot(aes(x = species))
```

```
(B)  penguins |>
       ggplot(aes(x = species,
                  fill = island)) +
       geom_point()
```

```
(C)  penguins |>
       ggplot(aes(x = species,
                  fill = island)) +
       geom_bar()
```

```
(D)  penguins |>
       ggplot(aes(x = species)) +
       geom_bar()
```

Answer: (C)

Class Activity

https://sta112-s26.github.io/class_activities/ca_04.html

- Welcome to work in groups
- Submit HTML on Canvas when finished
- Can work on HW 1 if done early