A Brief Intro to Plotting with ggplot2

Andreas Felderer

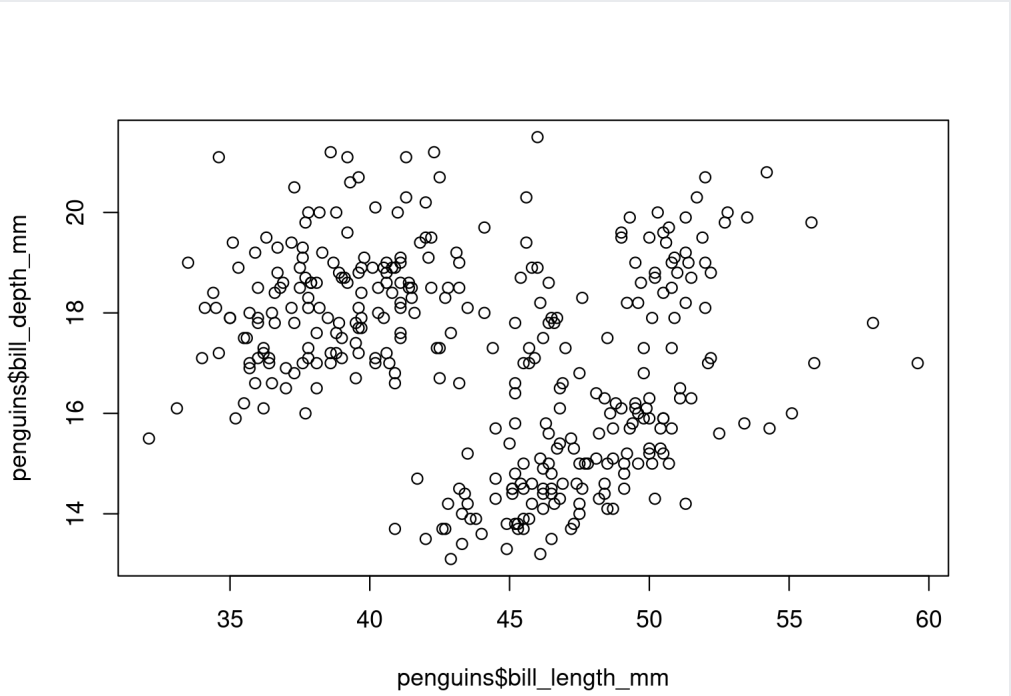
15.09.2021

Why ggplot(2)?

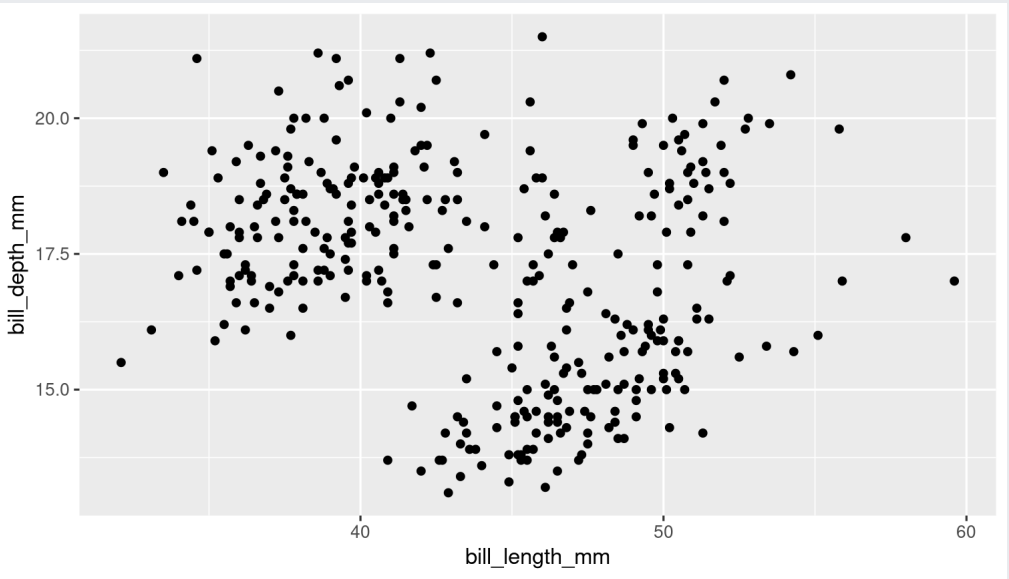
- Beautiful plots
- Part of the tidyverse
- Readability & re-usability
- Documentation

Plot example

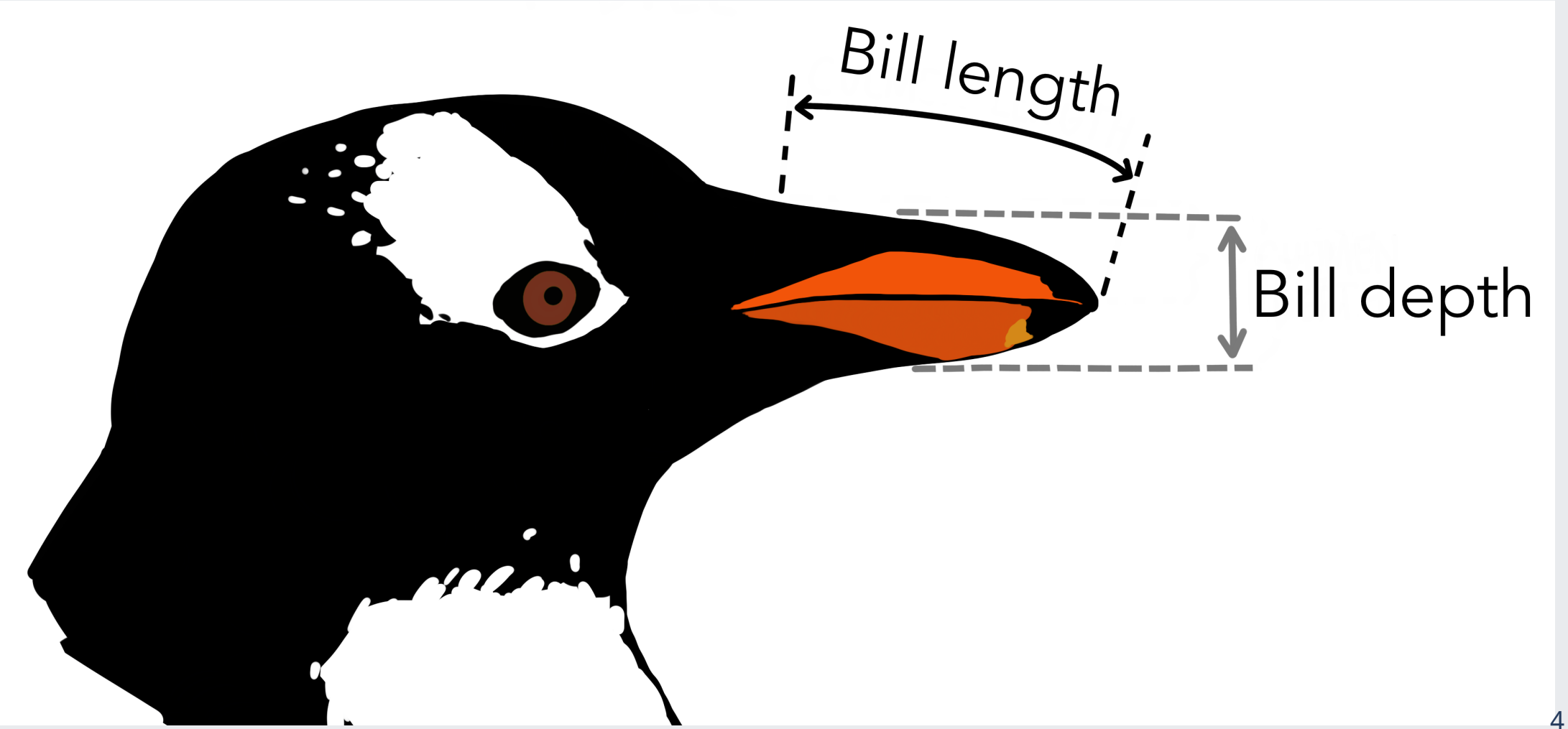
```
plot(penguins$bill_length_mm,
     penguins$bill_depth_mm)
```




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

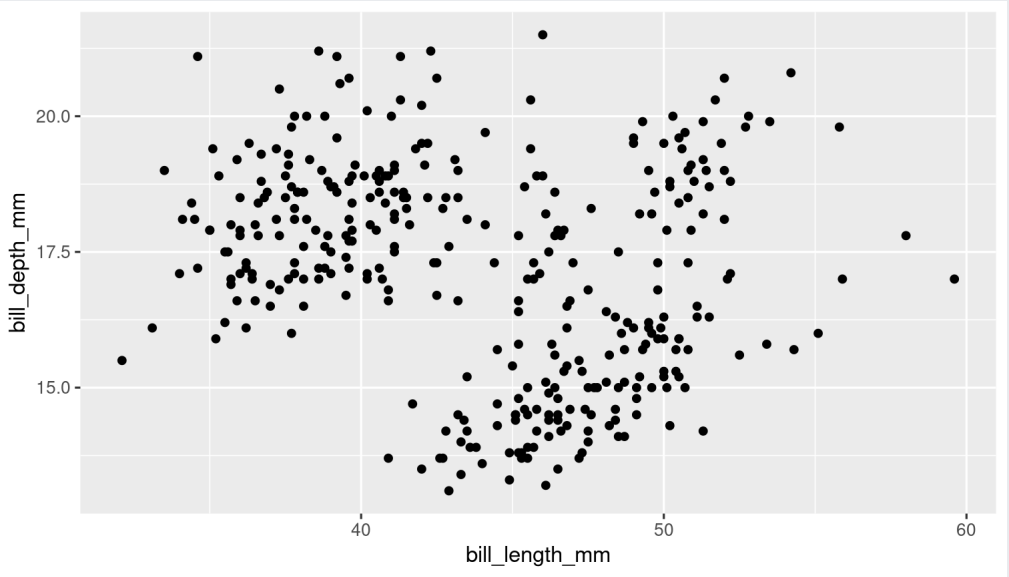


What did we just plot?



How does ggplot work ?

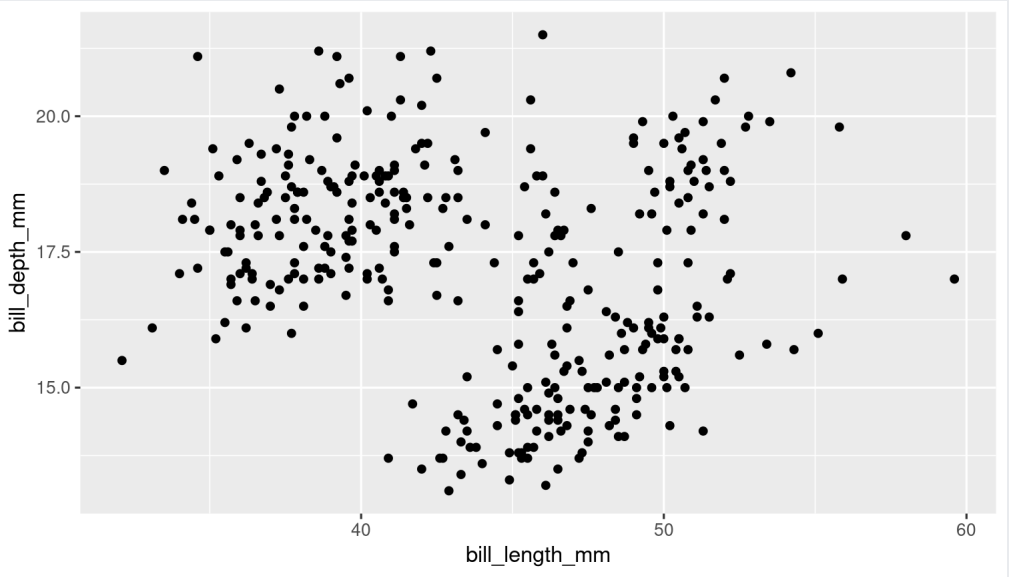
```
ggplot(penguins) +
  aes(x=bill_length_mm,
      y=bill_depth_mm) +
  geom_point()
```



1. **Data**
    - Provide ggplot with a dataframe

How does ggplot work ?

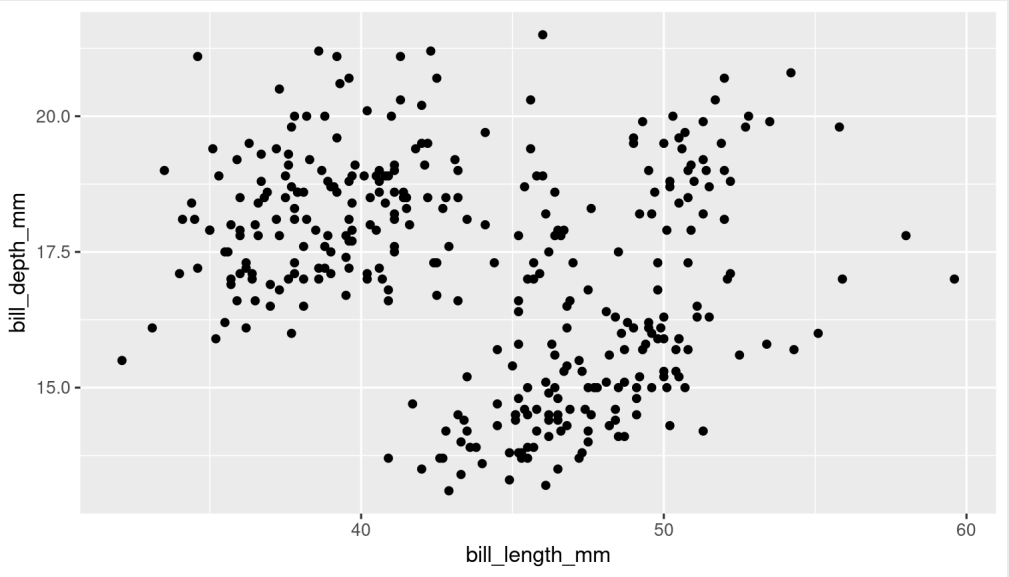
```
ggplot(penguins) +
  aes(x=bill_length_mm,
      y=bill_depth_mm) +
  geom_point()
```



1. Data
    - Provide ggplot with a dataframe
2. **Aesthetics ("aes")**
    - Tell ggplot how to style the data
    - e.g. what belongs on the x and y axis or how to points

How does ggplot work ?

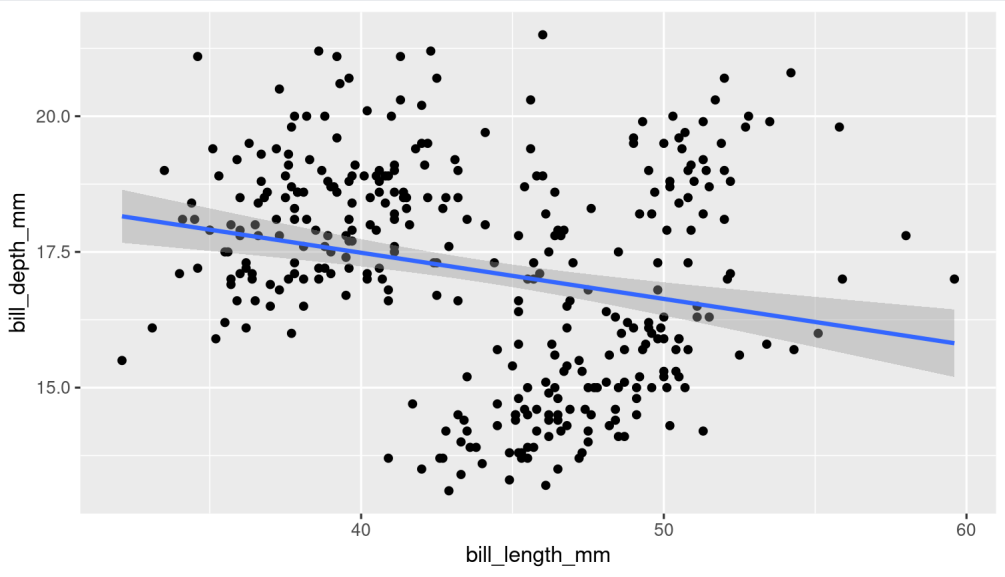
```
ggplot(penguins) +
  aes(x=bill_length_mm,
      y=bill_depth_mm) +
  geom_point()
```



1. Data
    - Provide ggplot with a dataframe
2. Aesthetics ("aes")
    - Tell ggplot how to style the data
    - e.g. what belongs on the x and y axis or how to points
3. **Geometries ("geoms")**
    - Tell ggplot the geometrical form to be used
    - e.g.: line, point, bar, box, ...

Whats the matter with geom?

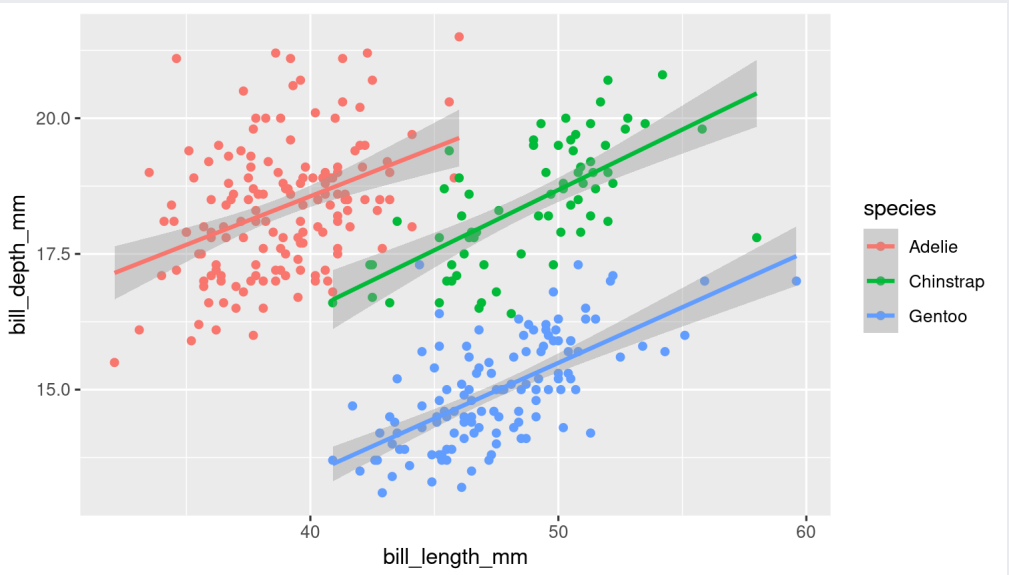
```
ggplot(penguins) +
  aes(x=bill_length_mm,
      y=bill_depth_mm) +
  geom_point() +
  geom_smooth(method="lm")
```



- We can easily add another representation of the same data
- In this case the best linear approximation of the data
- But wait why is the slope negative?

Whats the matter with aes?

```
ggplot(penguins) +
  aes(x=bill_length_mm,
      y=bill_depth_mm,
      color=species) +
  geom_point() +
  geom_smooth(method="lm")
```



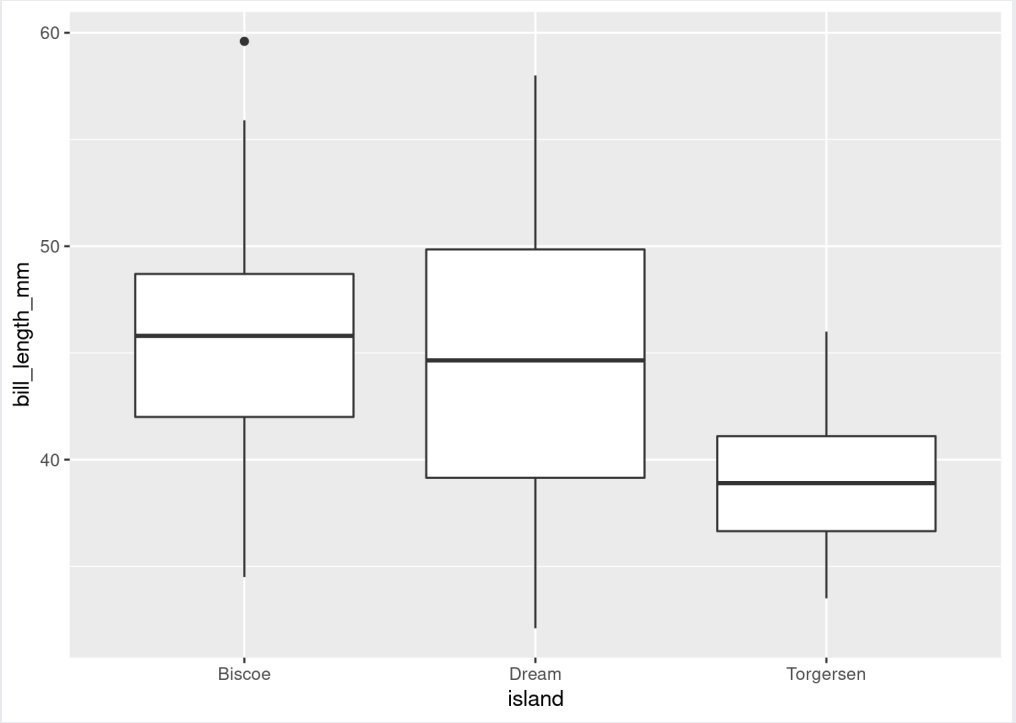
- We can easily add another aesthetic
- In this case color each species in a different color
- Now the slopes are positive
- A nice example of Simpsons paradox

Example:

Plot distribution of bill length per islands:

Let's start with a box plot

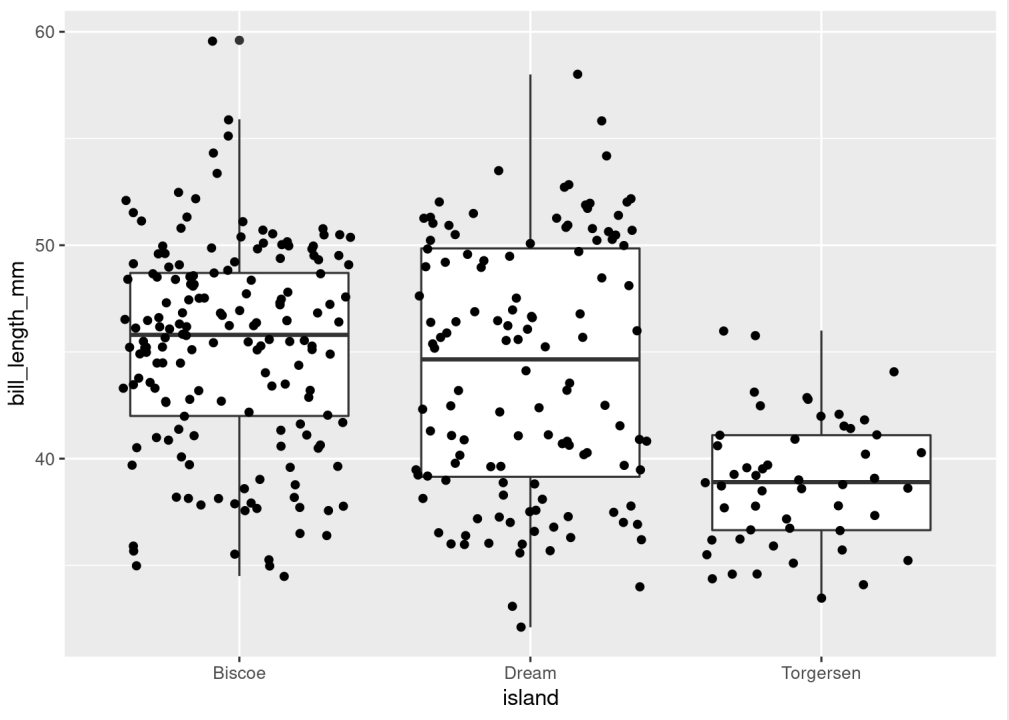
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island) +
  geom_boxplot()
```



Example:

A boxplot may hide information which can be made accessible with jitter

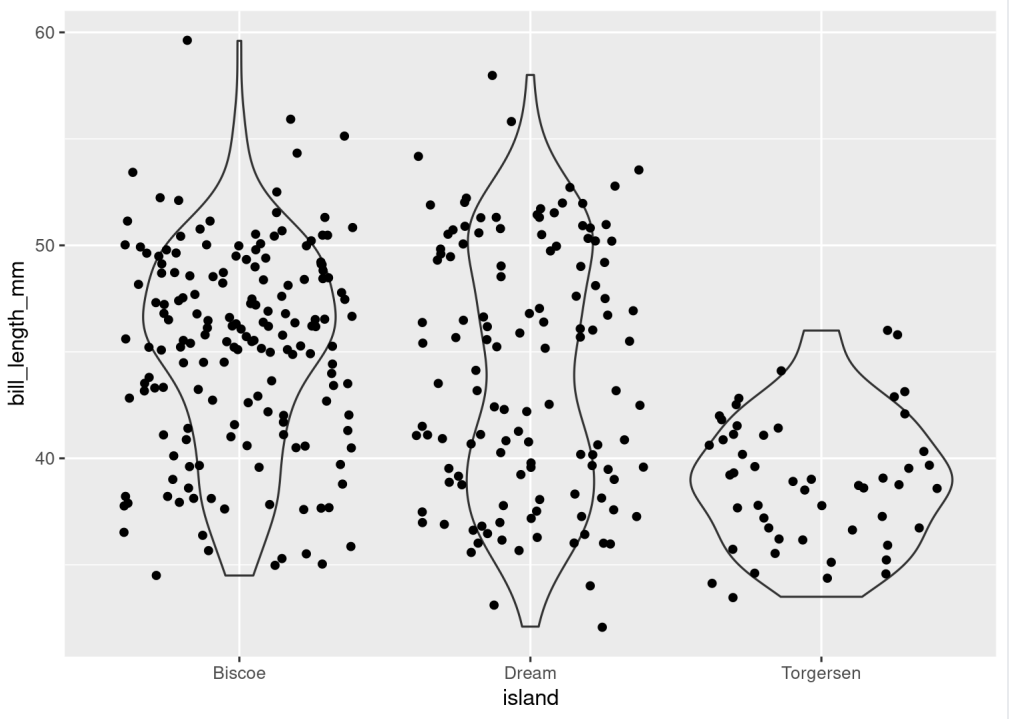
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island) +
  geom_boxplot() +
  geom_jitter()
```



Example:

Another possibility is a violin plot (works best with lots of data)

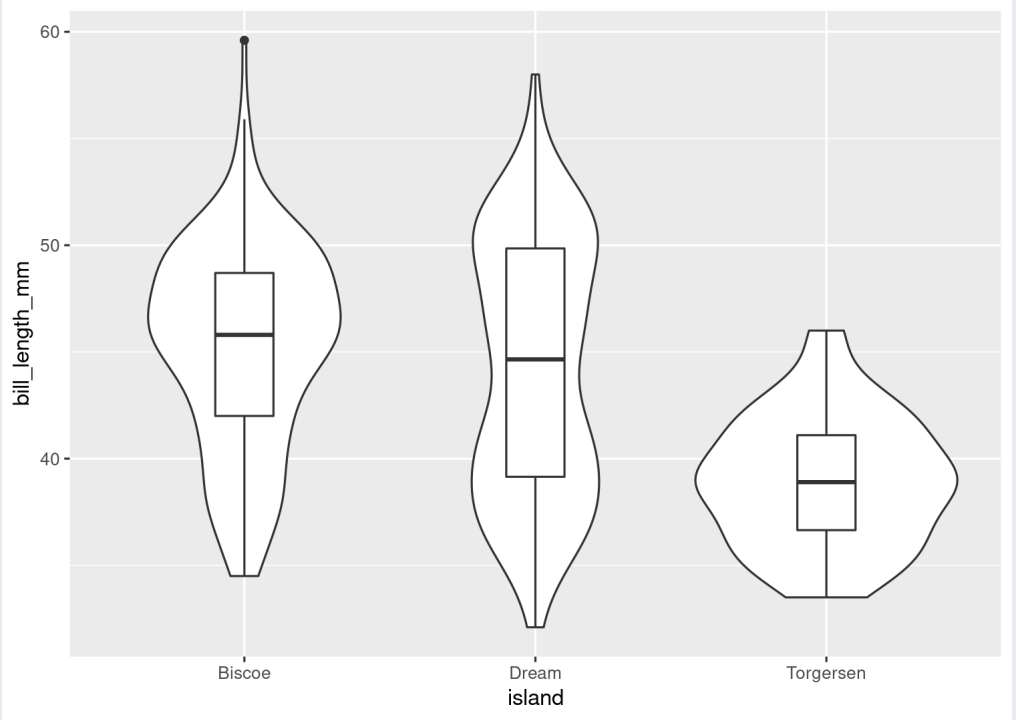
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island) +
  # geom_boxplot() +
  geom_violin(fill=NA) +
  geom_jitter()
```



Example:

Combining violin with boxplot is a powerful option

```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island) +
  geom_violin() +
  # geom_jitter()
  geom_boxplot(width=0.2)
```



Example:

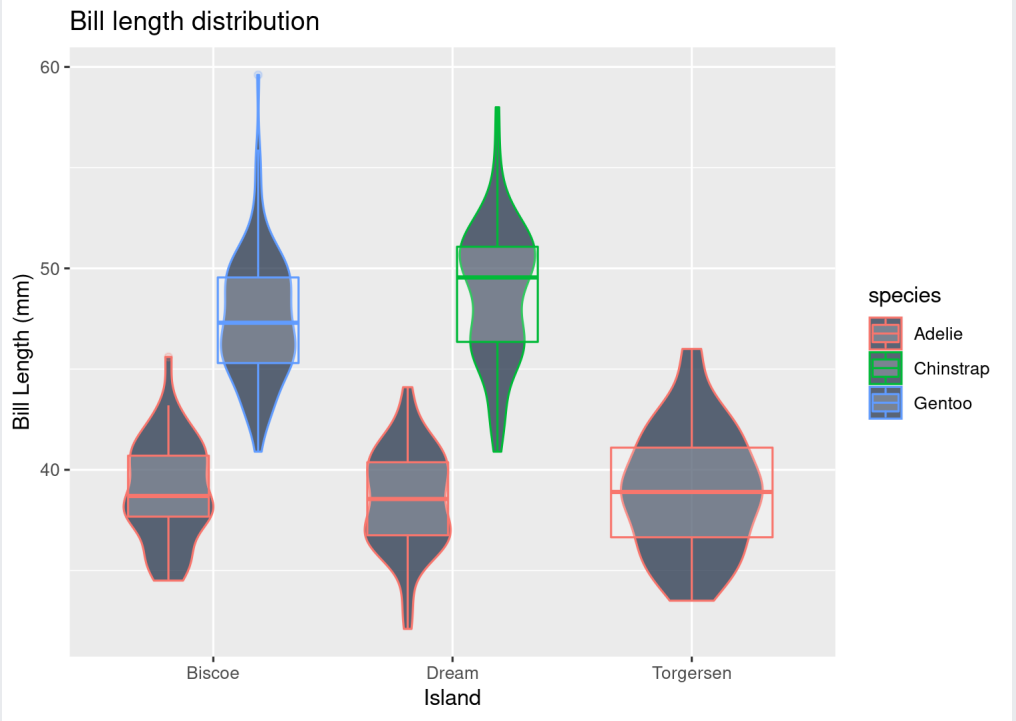
Let's add title, labels, and a theme

```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island) +
  geom_violin() +
  geom_boxplot(width=0.2,
               alpha=0.2,
               color = "white") +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)") +
  theme_xaringan()
```

Example:

Let's distinguish between species

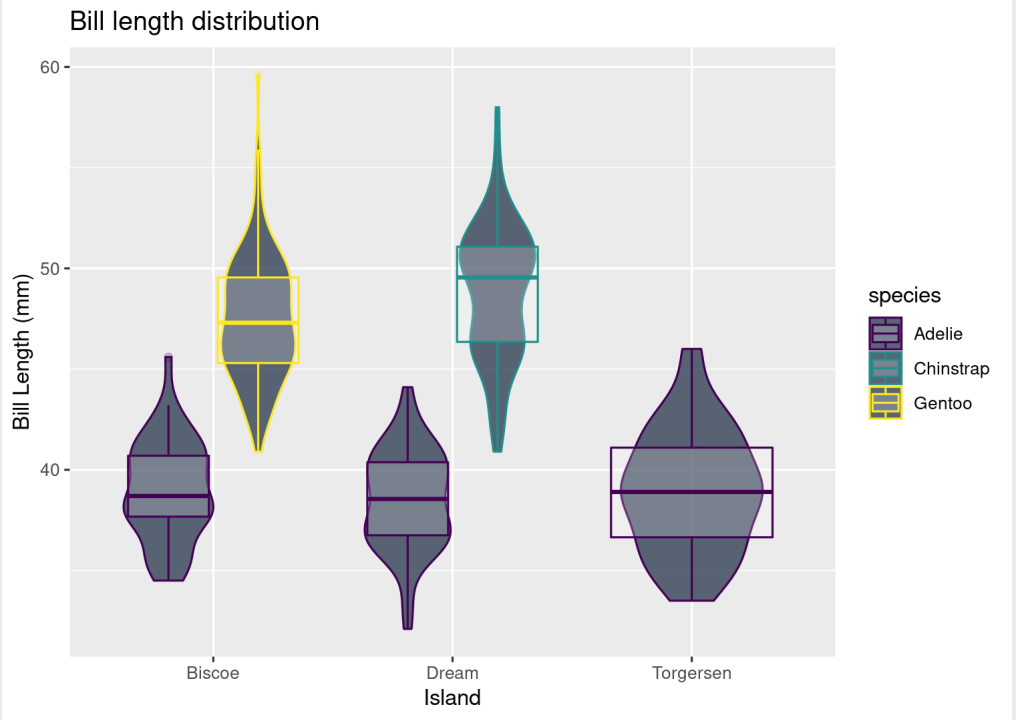
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island,
      color=species) +
  geom_violin(width=0.75, alpha=0.7) +
  geom_boxplot(alpha=0.2) +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)")
```



Example:

And change colors to the viridis scale (optimized for colorblind, and grayscale printing)

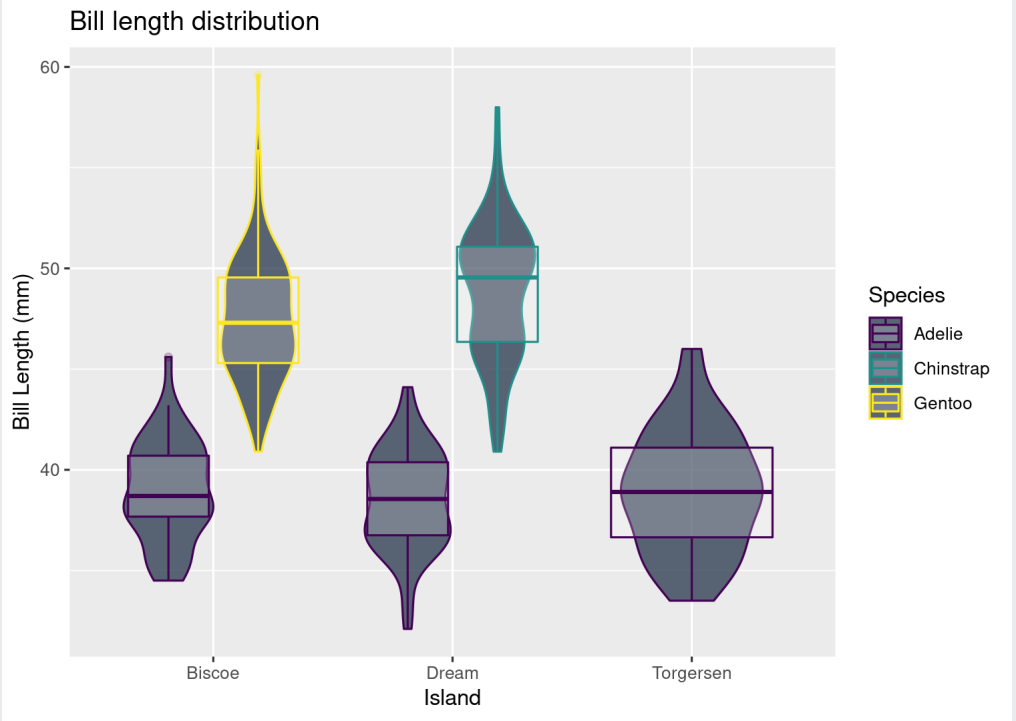
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island,
      color=species) +
  geom_violin(width=0.75, alpha=0.7) +
  geom_boxplot(alpha=0.2) +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)") +
  scale_color_viridis(discrete = TRUE)
```



Example:

Rename the legend

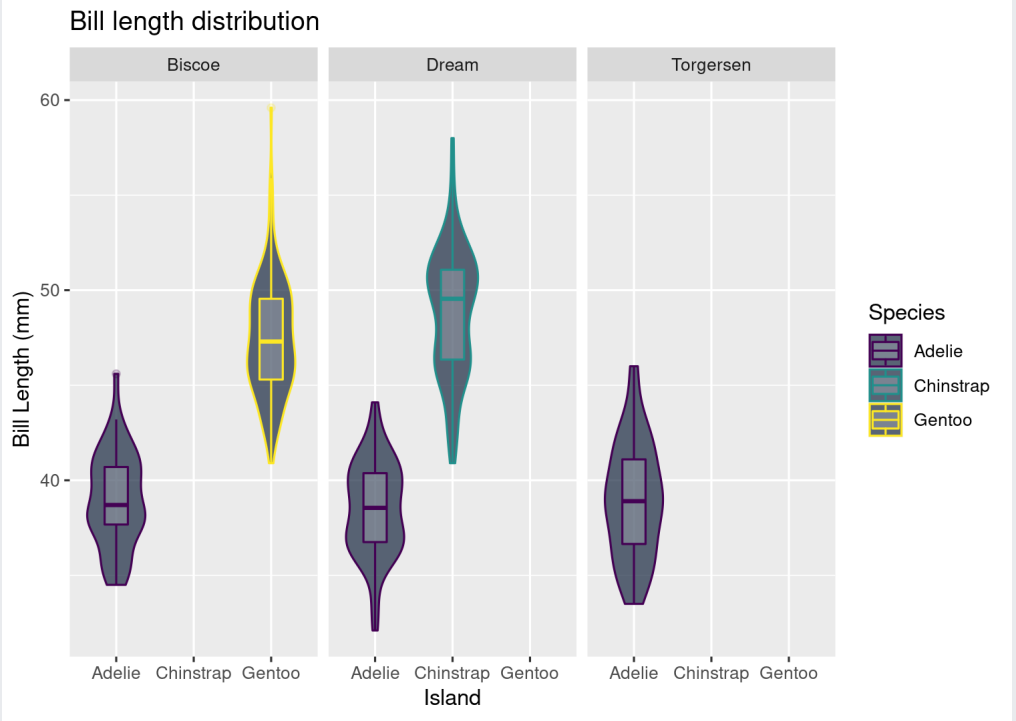
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=island,
      color=species) +
  geom_violin(width=0.75, alpha=0.7) +
  geom_boxplot(alpha=0.2) +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)",
       color = "Species") +
  scale_color_viridis(discrete = TRUE)
```



Example:

Use the power of faceting

```
ggplot(penguins) +
  aes(y=bill_length_mm, x=species,
      color=species) +
  facet_grid(cols = vars(island)) +
  geom_violin(width=0.75, alpha=0.7) +
  geom_boxplot(width=0.3, alpha=0.2) +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)",
       color = "Species") +
  scale_color_viridis(discrete = TRUE)
```

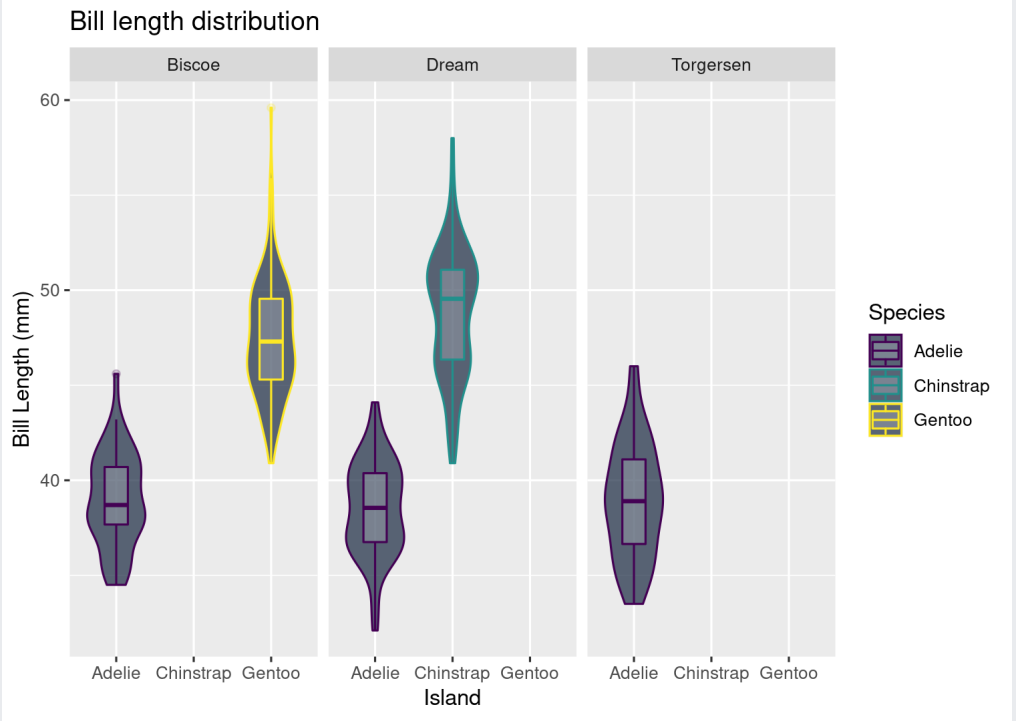


Example:

And save the plot as .png with ggsave.

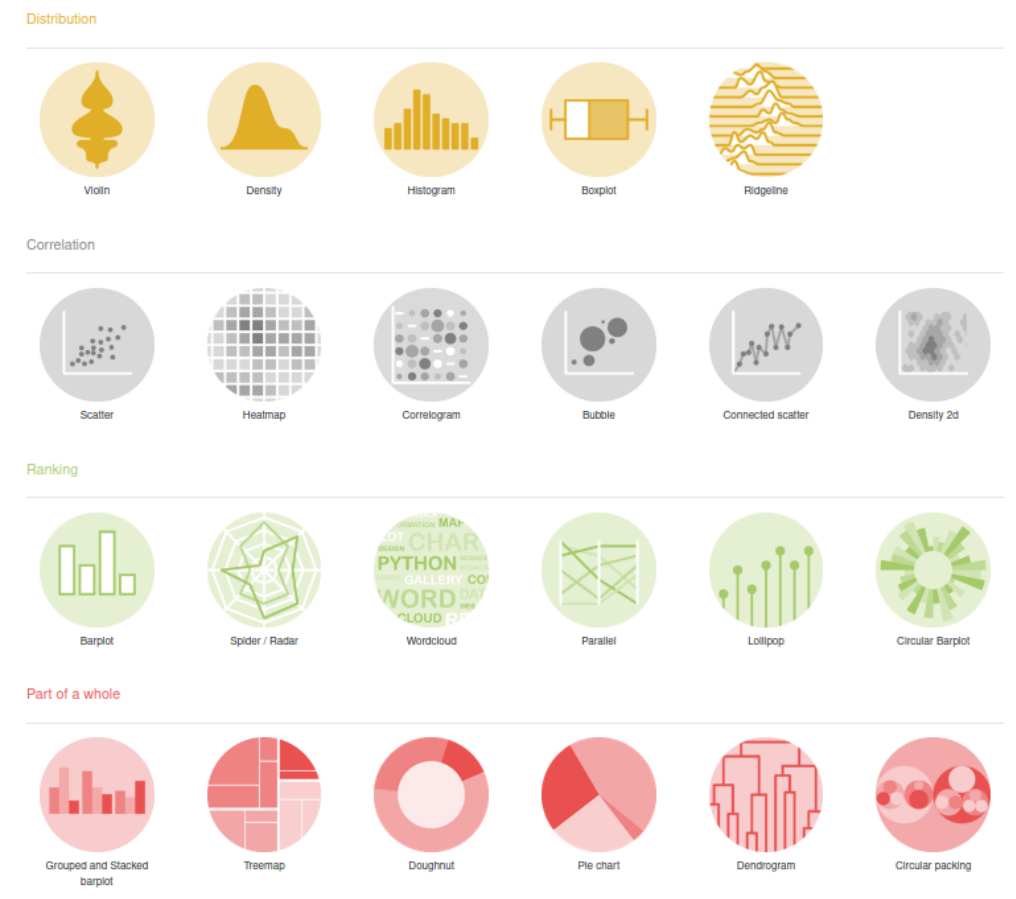
```
ggplot(penguins) +
  aes(y=bill_length_mm, x=species,
      color=species) +
  facet_grid(cols = vars(island)) +
  geom_violin(width=0.75, alpha=0.7) +
  geom_boxplot(width=0.3, alpha=0.2) +
  labs(title = "Bill length distribution",
       x = "Island",
       y = "Bill Length (mm)",
       color = "Species") +
  scale_color_viridis(discrete = TRUE)
```

```
ggsave("our_first_plot.png")
```



ggsave() can save plots as: "eps", "ps", "tex" (pictex), "pdf", "jpeg", "tiff", "png", "bmp", "svg" or "wmf" (windows only).

There are many, many ways to represent data



1. Ridgline
2. Histogram
3. Spider plot
4. Treemap
5. Bubble
6. etc. ...

Helpful resources

- **The Documaentation:**

  https://ggplot2.tidyverse.org/index.html

- **R Cheatsheets (ggplot2):**

  https://github.com/rstudio/cheatsheets/blob/master/data-visualization-2.1.pdf

- **Extensive intro to ggplot:** The R-Graphics Cookbook:

  https://r-graphics.org/

- **Concise intro to ggplot:** The *Graphics for Communication* chapter from R4DS:

  https://r4ds.had.co.nz/graphics-for-communication.html

- **Inspiration:** The R Graph Gallery:

  https://www.r-graph-gallery.com/

Questions?