Week 3: Data Visualization

{ggplot2}

Eunji Kong

(Adapted from Dr. Joe Nese's Lecture)

University of Oregon

Fall 2025

UNIVERSITY OF OREGON

Greetings!

- Eunji Kong
- 4th year SPED doctoral student
- Finished EDS specialization
- EDS project
  - Data Viz Project: https://eunjikong.shinyapps.io/edld652_shiny/#section-amber-1
  - Capstone: https://eunjikong.shinyapps.io/EDLD640_Capstone_Final/

Learning Objectives

- Understand the basic syntax requirements for {ggplot2}
- Recognize various options for displaying data
- Familiarity with various {ggplot2} options/layers

Basically, how to graph/visualize data

Lecture/Material Structure

- PDF Lecture Notes
  - Include hyperlinks that take you directly to the relevant topics
    - Hyperlinks: everything underlined

- .qmd file
  - Same information as the PDF but allows you to write notes directly in the file
  - You can also test out code interactively as you follow along
  - You can then render this document as an HTML file for later review
  - Visual mode

{tidyverse}

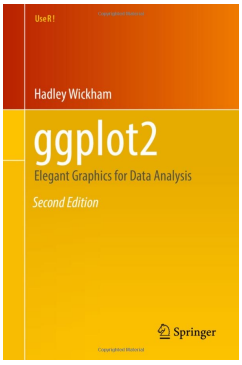
credit: Allison Horst

{tidyverse} is a meta-package that loads a set of core packages

- ✓ dplyr 1.1.4
- ✓ readr 2.1.5
- ✓ forcats 1.0.0
- ✓ stringr 1.5.1
- ✓ **ggplot2 3.5.2**
- ✓ tibble 3.3.0
- ✓ lubridate 1.9.4
- ✓ tidyr 1.3.1
- ✓ purrr 1.0.4

{ggplot2}

- gg stands for “grammar of graphics”
- **Resources**
  - ggplot2 book
    - email Dr. Nese for digital copy



  - Posit cheat sheet
    - *can be helpful, perhaps more so after a little experience*
  - R Graphics Cookbook
  - R Graph Gallery
    - *past students have really liked this one*

Components

Every ggplot has **three** components:

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

UNIVERSITY OF OREGON

Components

Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

data

Components

Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

```
data |>

ggplot(aes(mapping))
```

Components

Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

```
data |>

ggplot(a
```

I use the base R's version of pipe |> instead of %>% but they are essentially the same thing.

Components

Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

```
data |>

ggplot(aes(mapping)) +

geom_function() +

additional layers
```

Within ggplot, you use + instead of |>

Components

Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

```
data |>

  ggplot(aes(mapping)) +

  geom_function() +

  additional layers
```

UNIVERSITY OF OREGON

Components

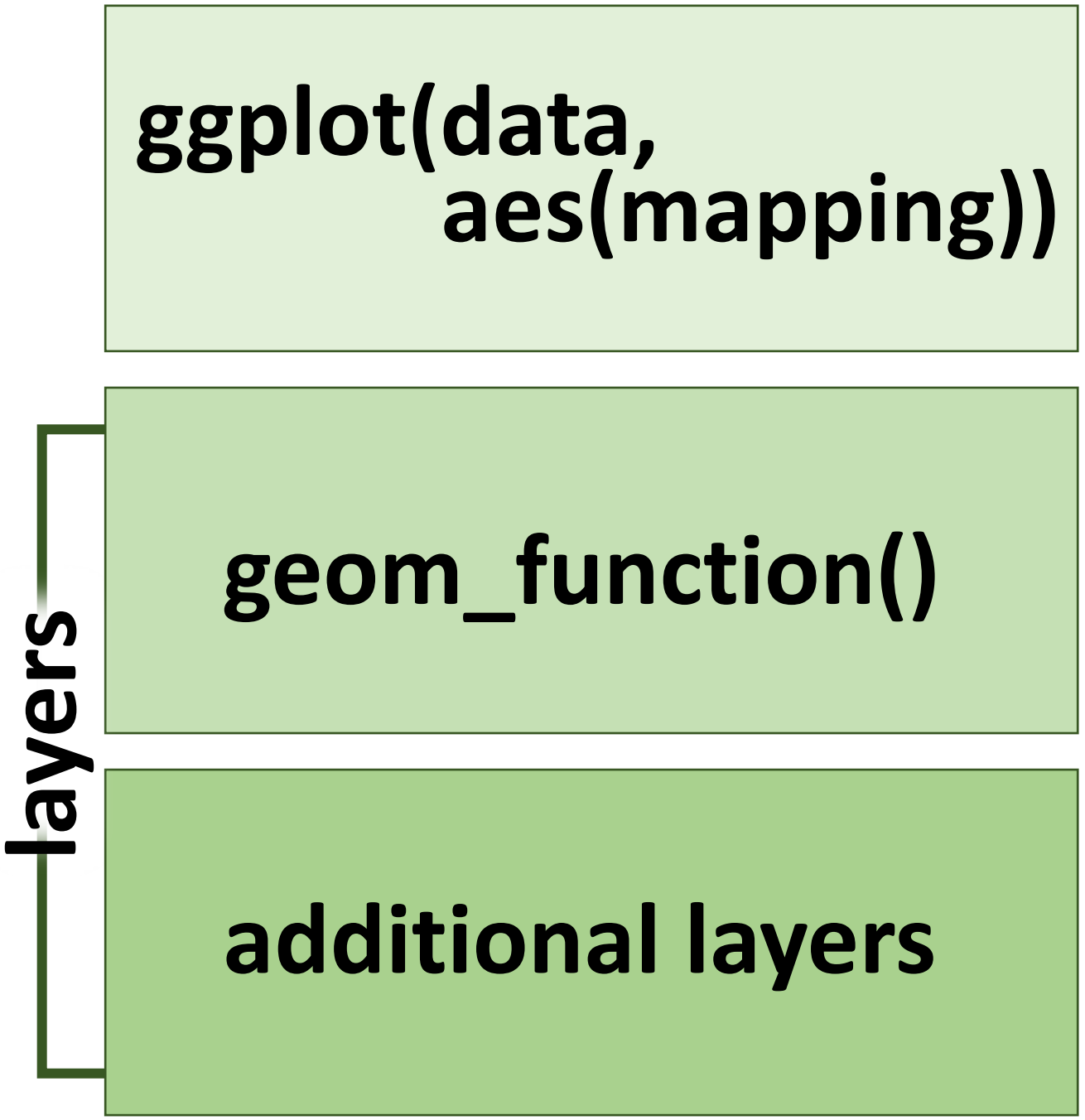
Every ggplot has **three** components

1. data
   - the data used to produce the plot
2. aesthetic mappings (aes)
   - between variables and visual properties
3. layers(s)
   - usually through the geom_*() function plus various other layers

Template

```
ggplot(data,
       aes(mapping)) +

geom_function() +

additional layers
```

UNIVERSITY OF OREGON


ggplot(data,
aes(mapping))
geom_function()
additional layers
layers

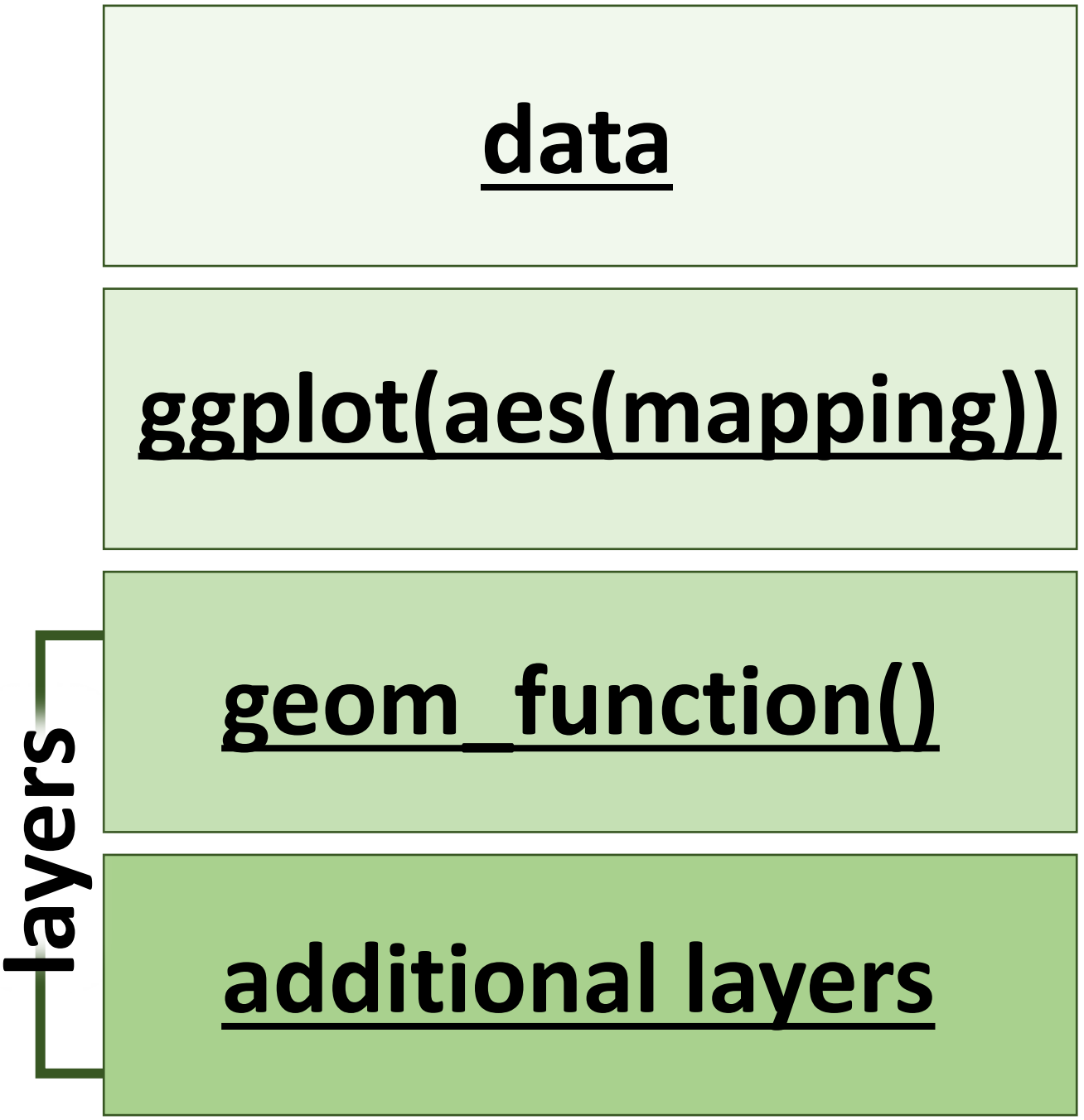
```
data |>
 ggplot(aes(mapping))
```

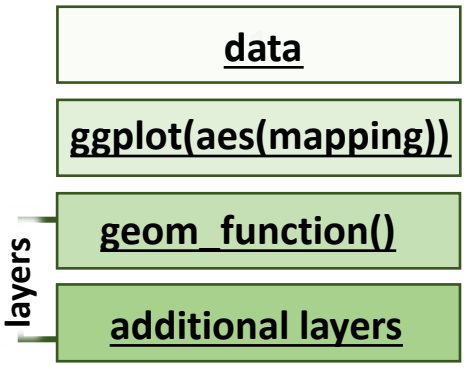
layers

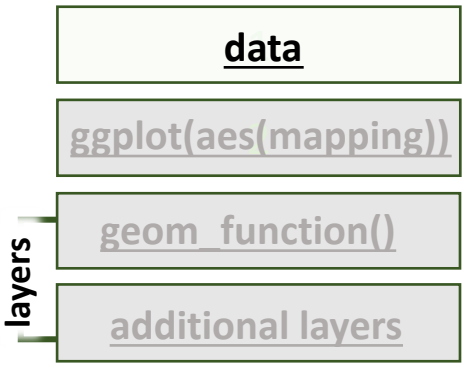
```
geom_function()
```

```
additional layers
```

UNIVERSITY OF OREGON


data
ggplot(aes(mapping))
geom_function()
additional layers
layers


data
ggplot(aes(mapping))
geom_function()
additional layers
layers

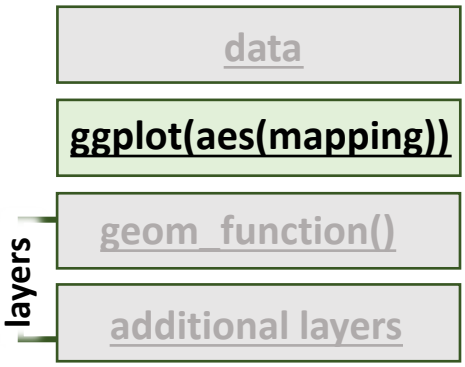


data

- install packages and load data

```
install.packages("palmerpenguins")

library(palmerpenguins)
```

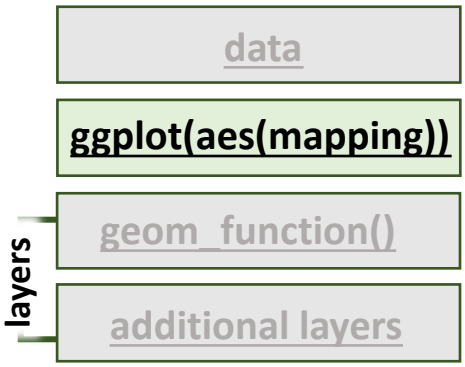
- examine the data
  - head()
  - str()
  - glimpse()
  - and more



ggplot(aes(mapping))

- aesthetic mappings describe how **variables** in the data are **mapped to visual properties**

- Some visual properties:
  - ✓ x
  - ✓ y
  - ✓ color *(will come back to it)*
  - ✓ fill *(will come back to it)*
  - ✓ alpha *(will come back to it)*
  - ✓ others *(linetype, shape, linewidth, size, group)*



ggplot(aes(mapping))

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)
```

data

ggplot(aes(mapping))

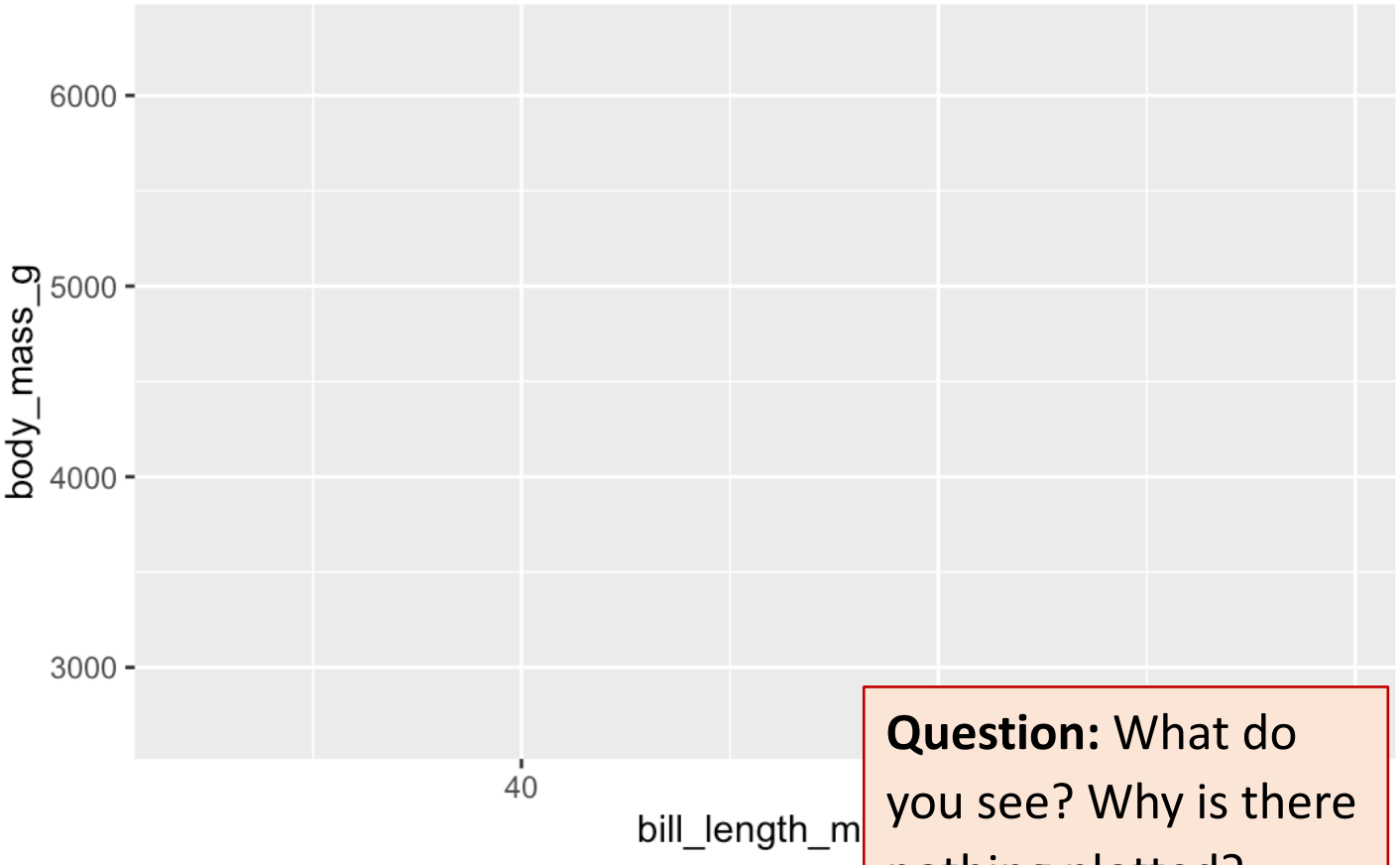
geom_function()

additional layers

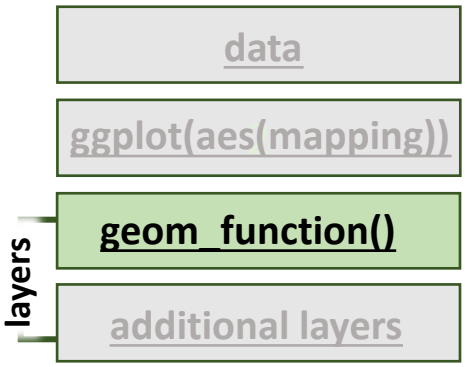
layers

ggplot(aes(mapping))

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)
```

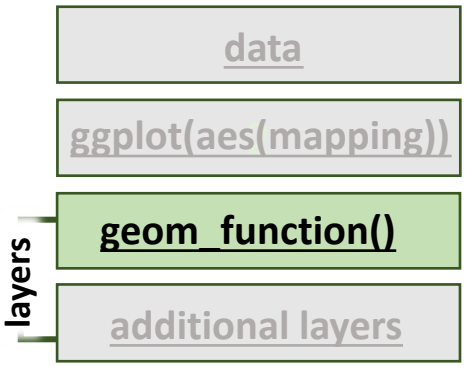


Question: What do you see? Why is there nothing plotted?



geom_function()

- Use a geom_function to represent **data points**



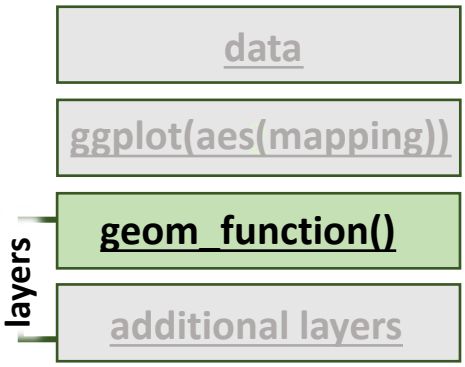
geom_function()

- Use a geom_function to represent **data points**

		Only 1 Variable	Variable 2	
			Continuous	Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	x	geom_count

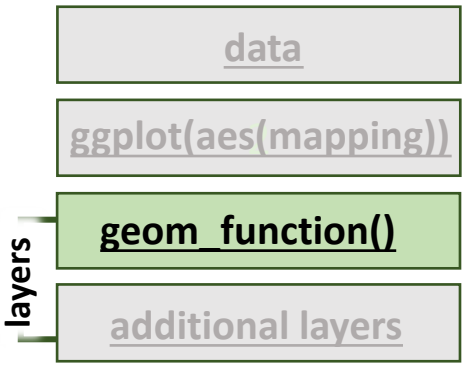
Heatmaps: geom_tile

UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

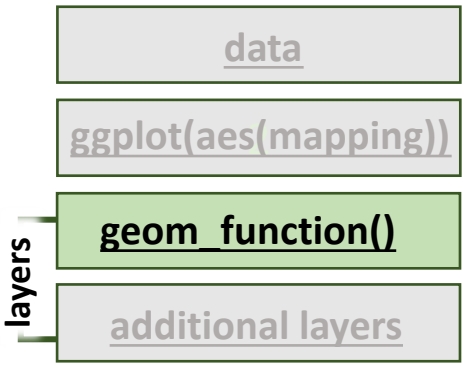
geom_histogram

General research question:

- How does the values of my continuous variable vary across its range?

Our data specific question:

- How is the distribution of penguin bill lengths (mm) in this sample? Any outliers? Unimodal?



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_histogram

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram()
```

Remember to use "+" instead of "|>" or "%>%"

data

ggplot(aes(mapping))

layers

geom_function()

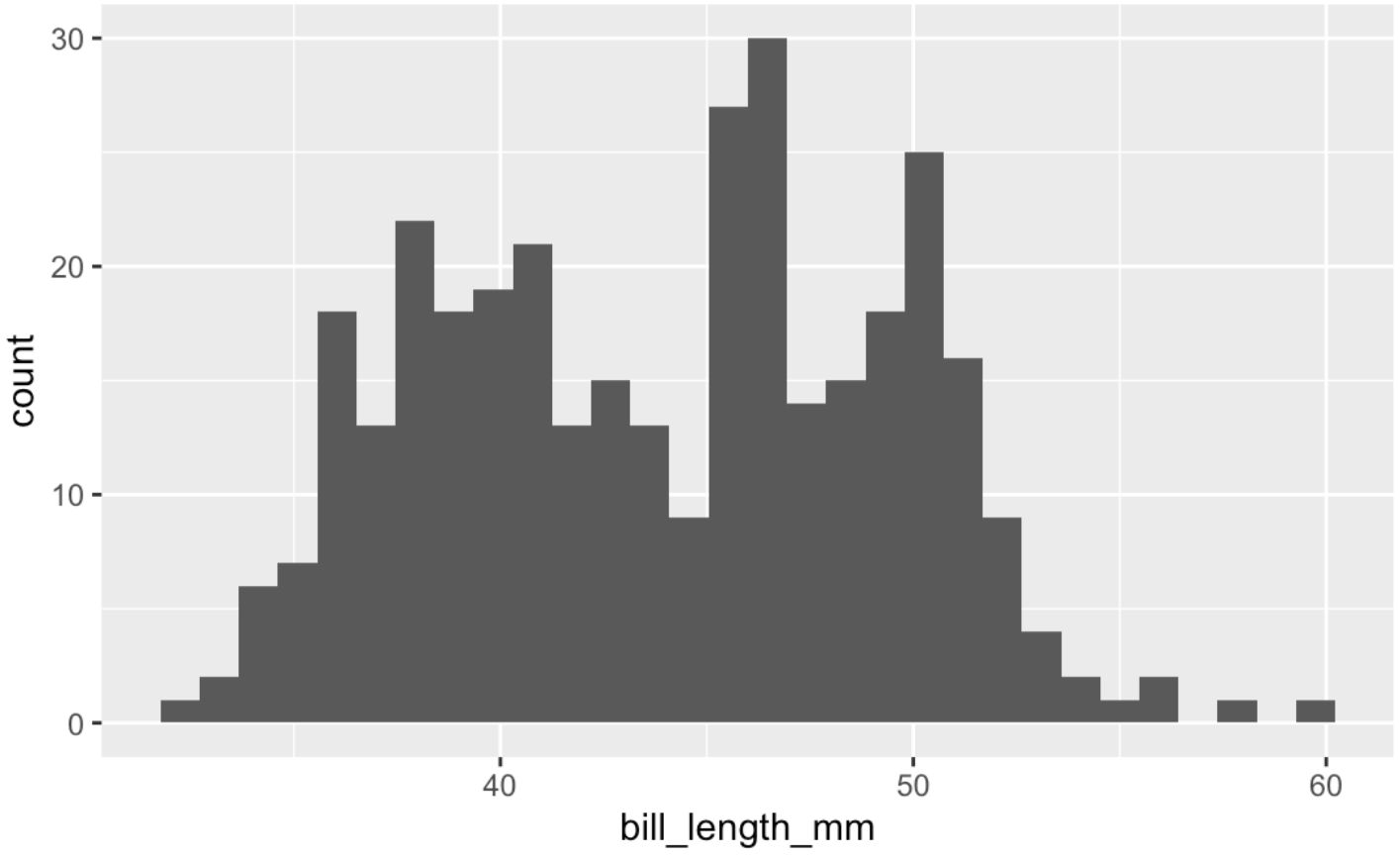
additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_histogram

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram()
```



layers

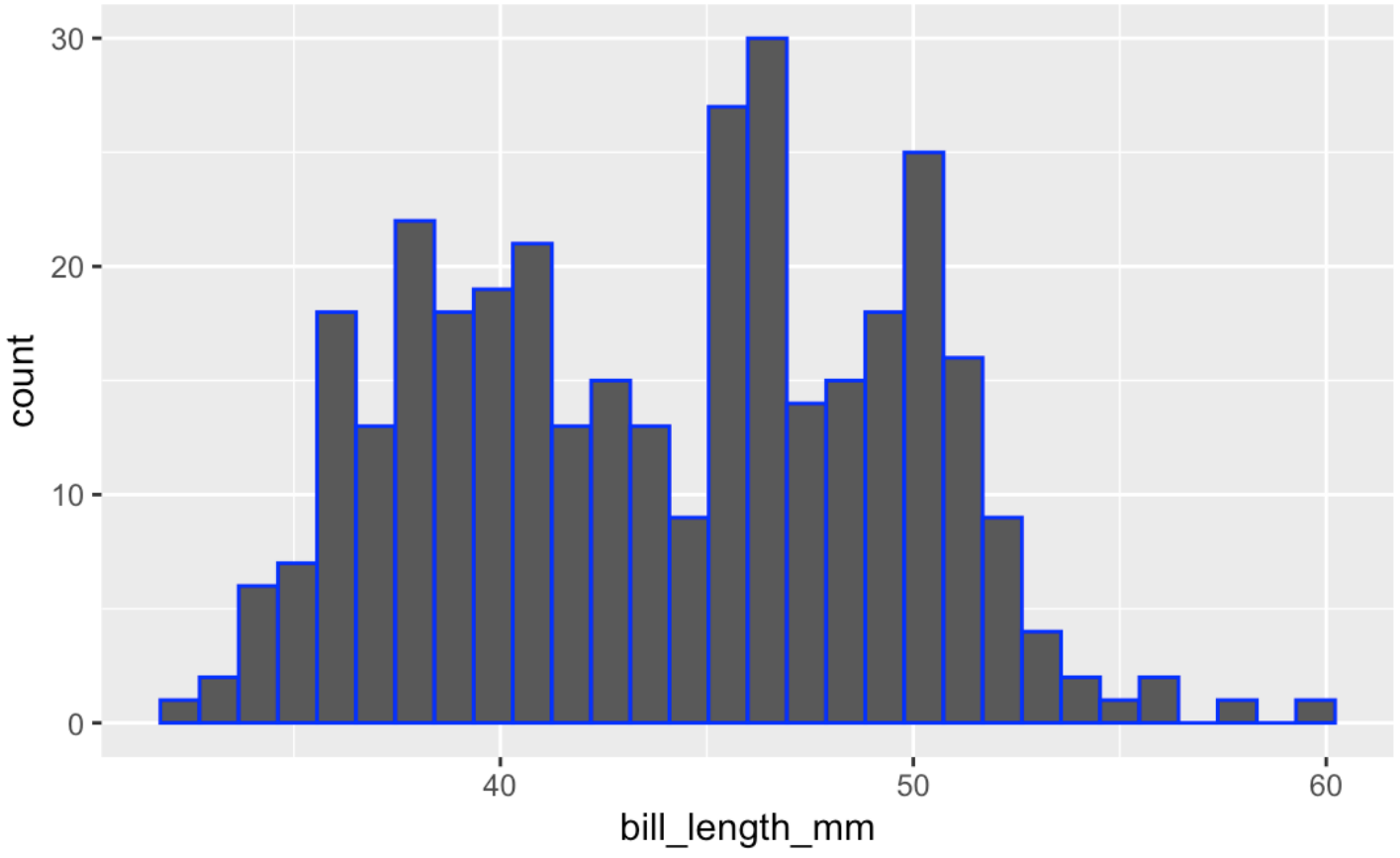
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Color vs. Fill

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram(color = "blue")
```



layers

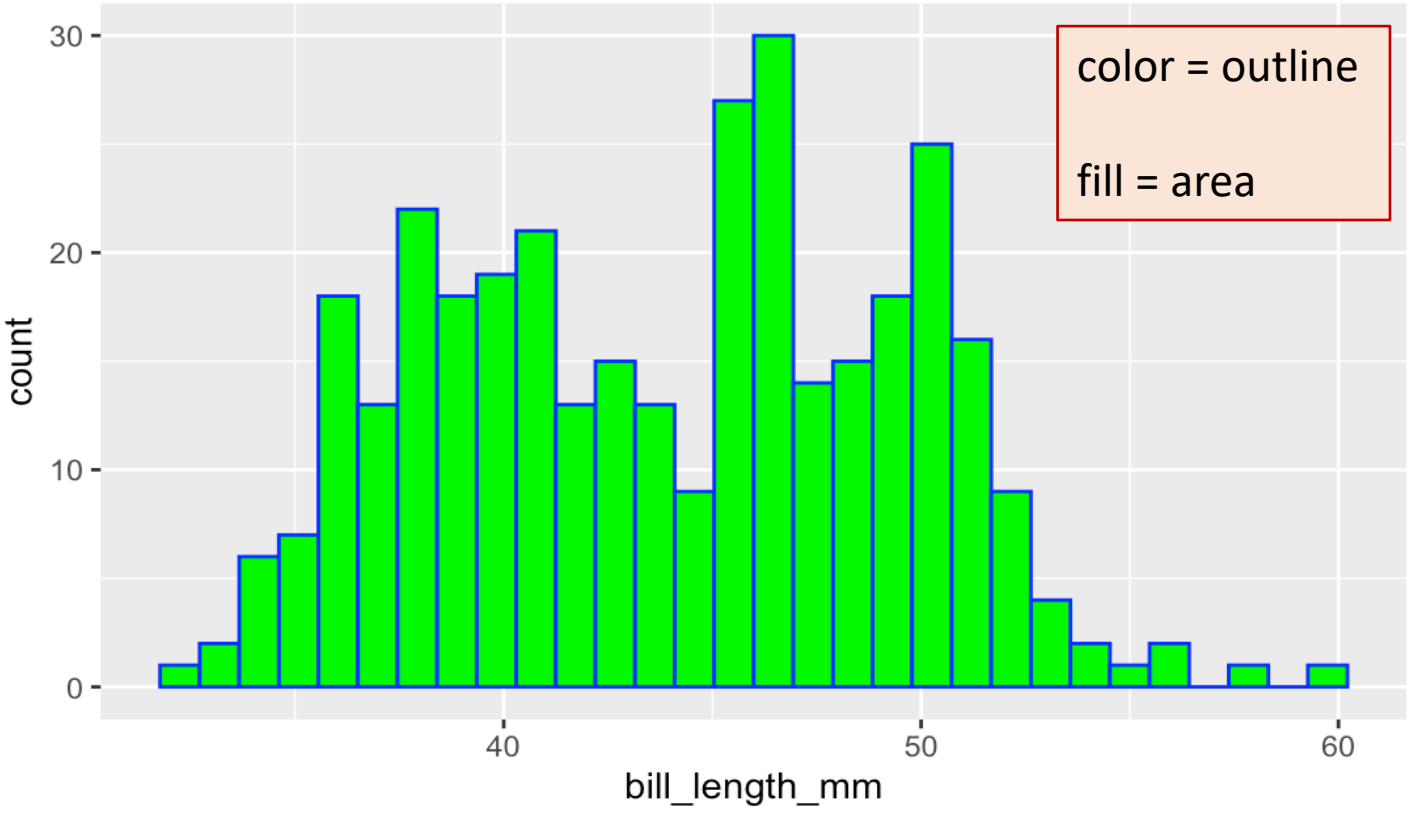
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Color vs. Fill

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram(color="blue",
                 fill="green")
```



UNIVERSITY OF OREGON

layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

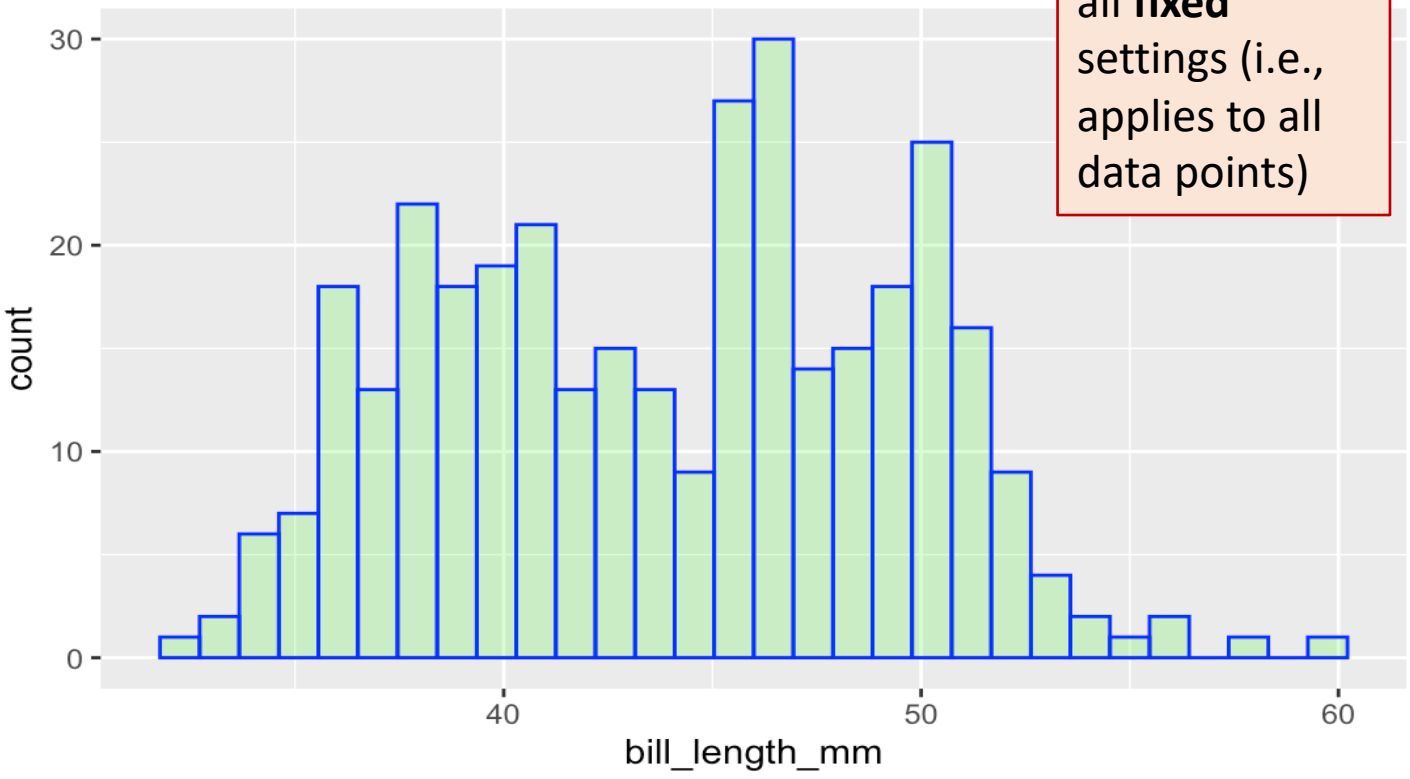
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Transparency

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram(color="blue",
                 fill="green",
                 alpha=0.2)
```

Color, fill & alpha in this example are all **fixed** settings (i.e., applies to all data points)



data

ggplot(aes(mapping))

layers

geom_function()

additional layers

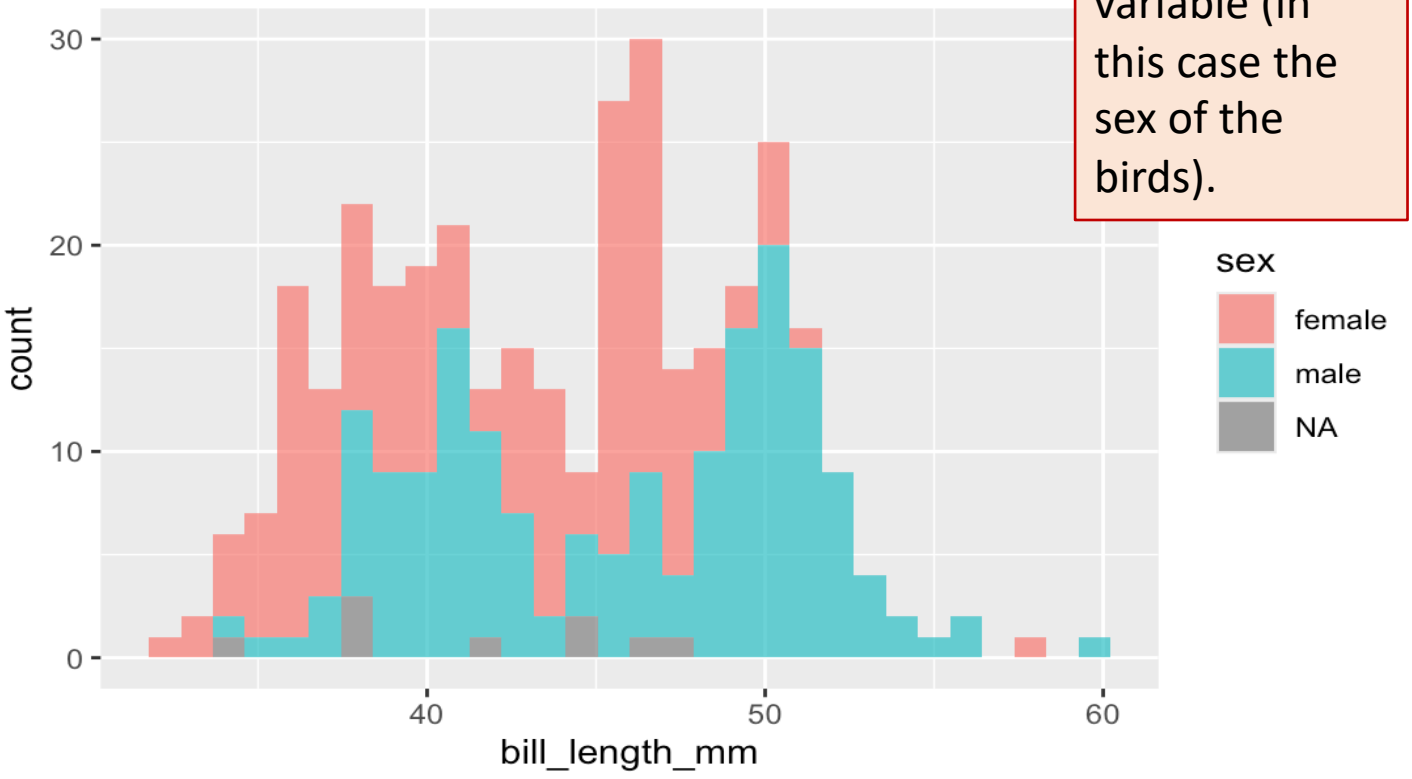
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

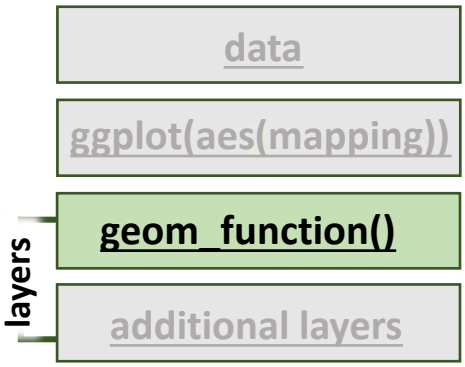
More aes mapping

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_histogram(aes(fill=sex),
                 alpha=0.7)
```

Fill here is a **conditional** mapping, meaning that the fill color is different based on the variable (in this case the sex of the birds).



UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Fixed vs Conditional

```
penguins |>
  ggplot(aes(bill_length_mm))+
  geom_histogram(fill="green")
```

Fill is not within aes(), fill is a fixed setting. Also notice that color is in quotes.

```
penguins |>
  ggplot(aes(bill_length_mm))+
  geom_histogram(aes(fill=sex))
```

aes() is used to access variables and make changes according to a specific variable. Also notice that variables are not in quotes.

layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

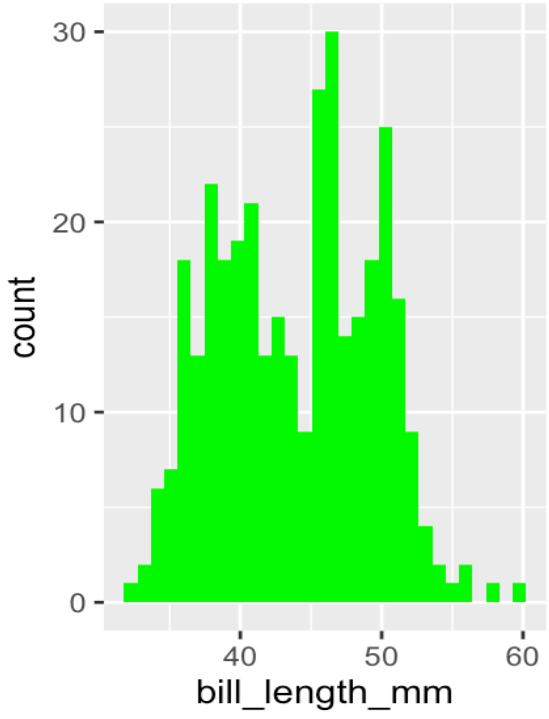
Fixed vs Conditional

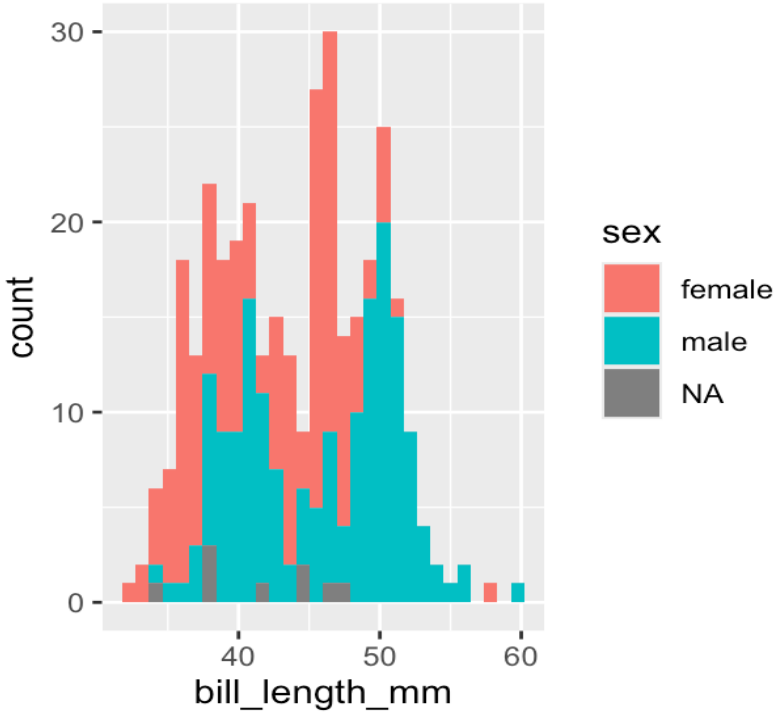
```
penguins |>
  ggplot(aes(bill_length_mm))+
  geom_histogram(fill="green")
```

fill is not within aes(), fill is a fixed setting. Also notice that color is in quotes.

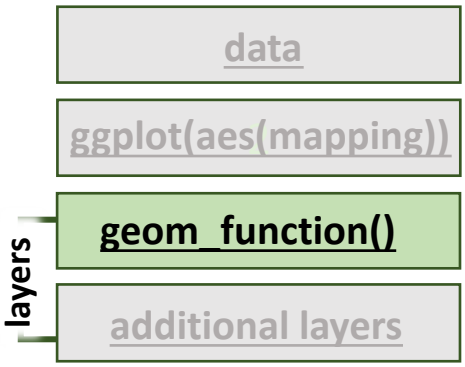
```
penguins |>
  ggplot(aes(bill_length_mm))+
  geom_histogram(aes(fill=sex))
```

aes() is used to access variables and make changes according to a specific variable. Also notice that variables are not in quotes.





UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Be mindful of aes()

```
penguins |>
 ggplot(aes(x=bill_length_mm))+
 geom_histogram(fill="green")
```

```
penguins |>
 ggplot(aes(x=bill_length_mm))+
 geom_histogram(aes(fill="green"))
```

Question: what is wrong with the bottom code? How do you think the plot will look like?

layers

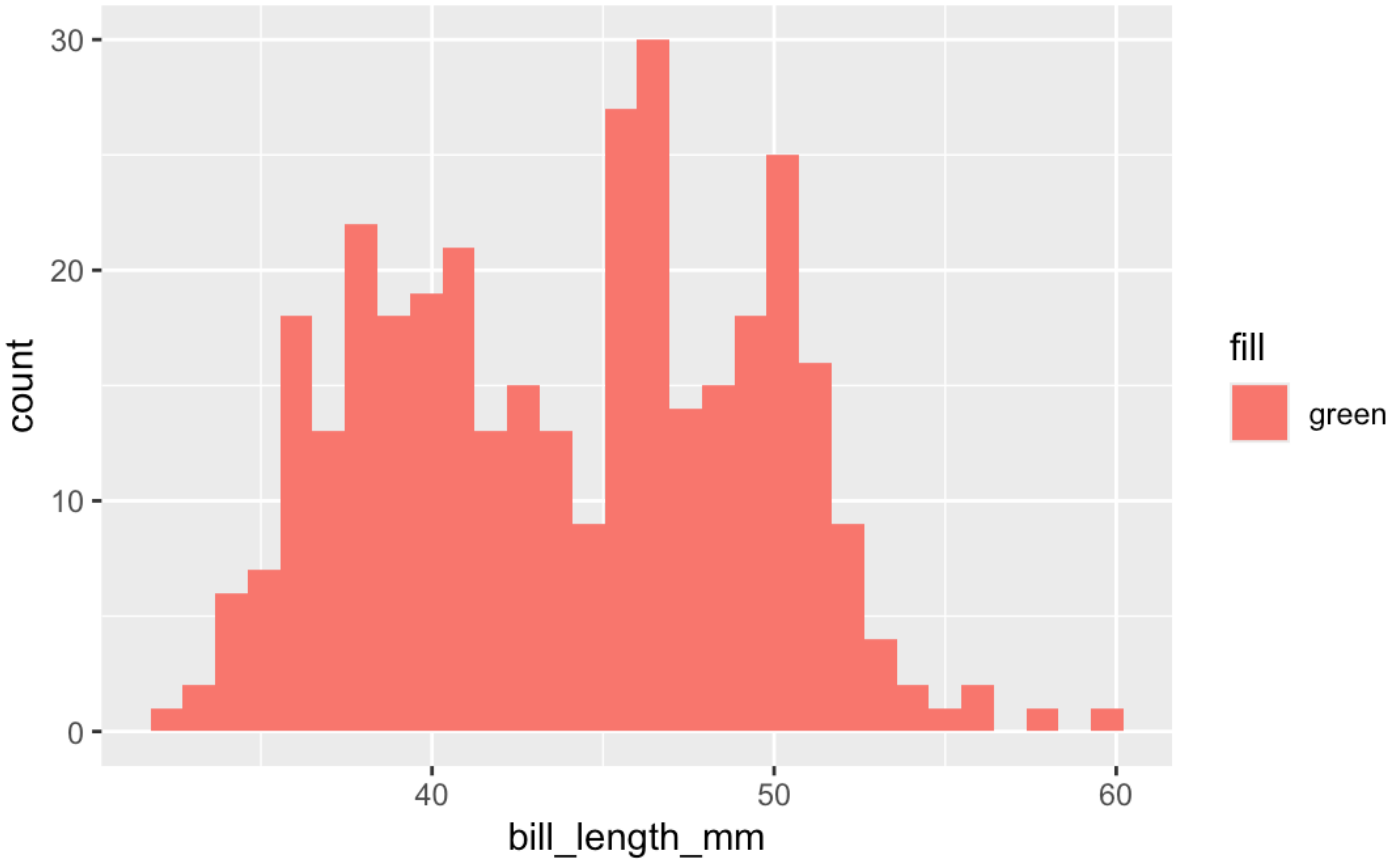
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

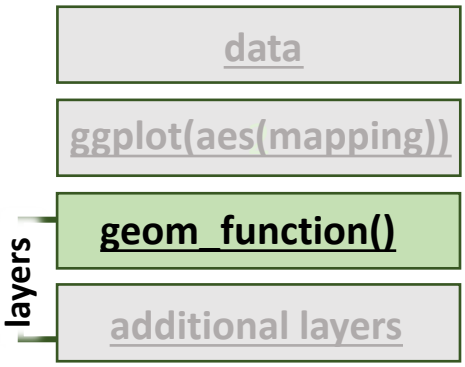
Other - Heatmaps: geom_tile

Be mindful of aes()

```
penguins |>
 ggplot(aes(x=bill_length_mm))+
 geom_histogram(aes(fill="green"))
```



UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_density()

General research question:

- How does the probability density of my continuous variable vary across its range?

- Think of it as a smoothed histogram

  - **Difference**: not use bins; not use count but relative frequency per unit of x

layers

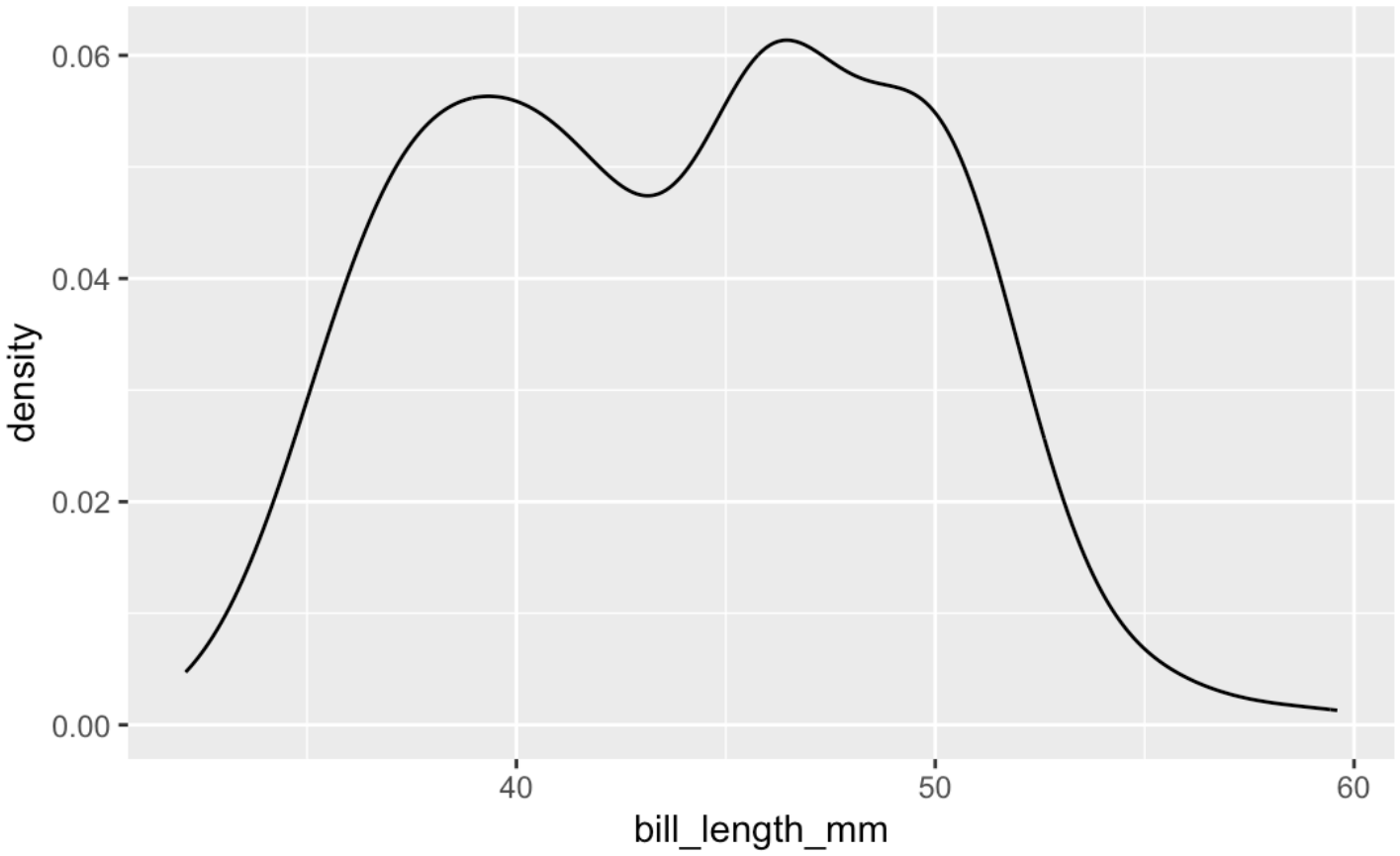
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_density()

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density()
```



layers

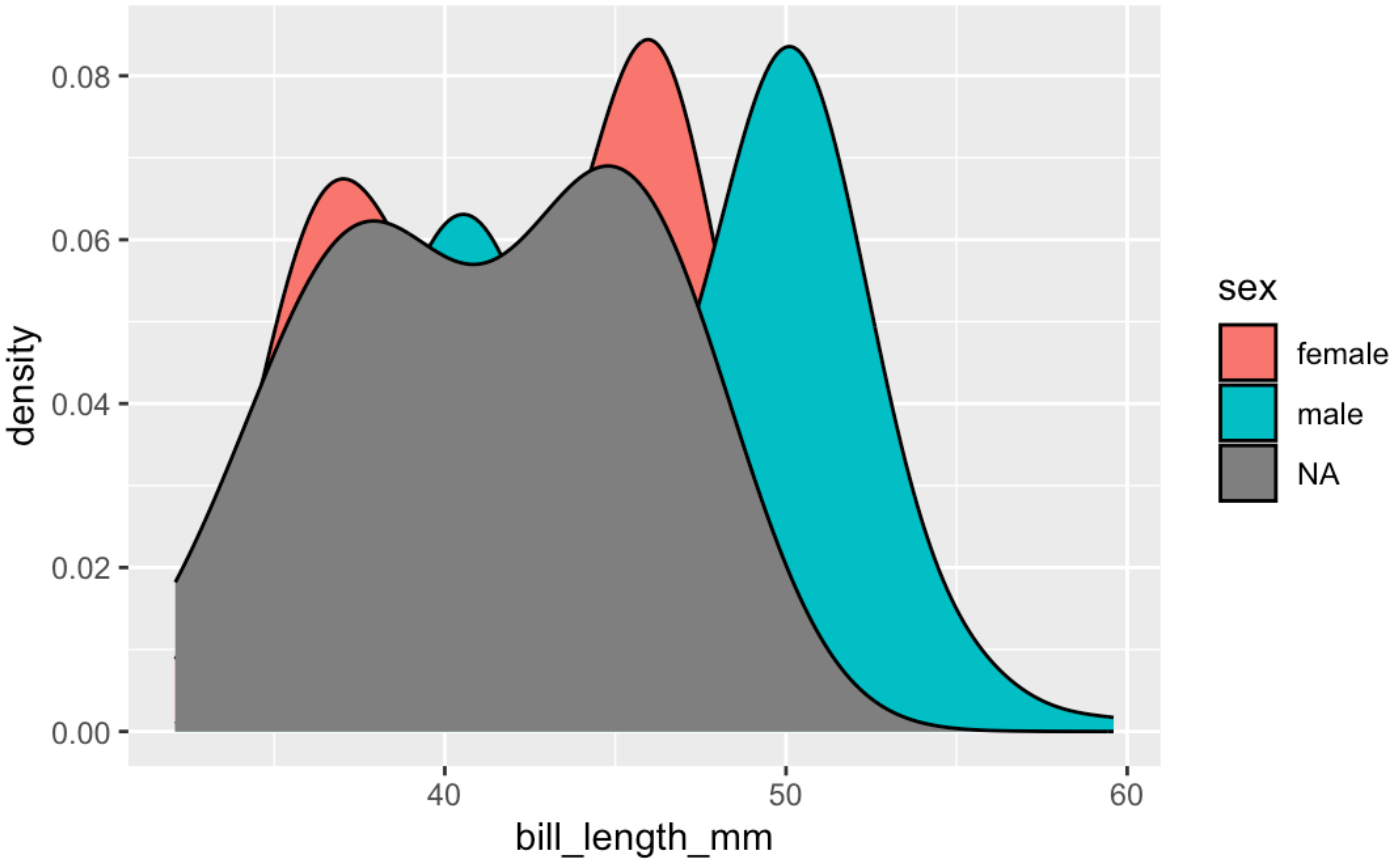
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

More aes mapping

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density(aes(fill=sex))
```



layers

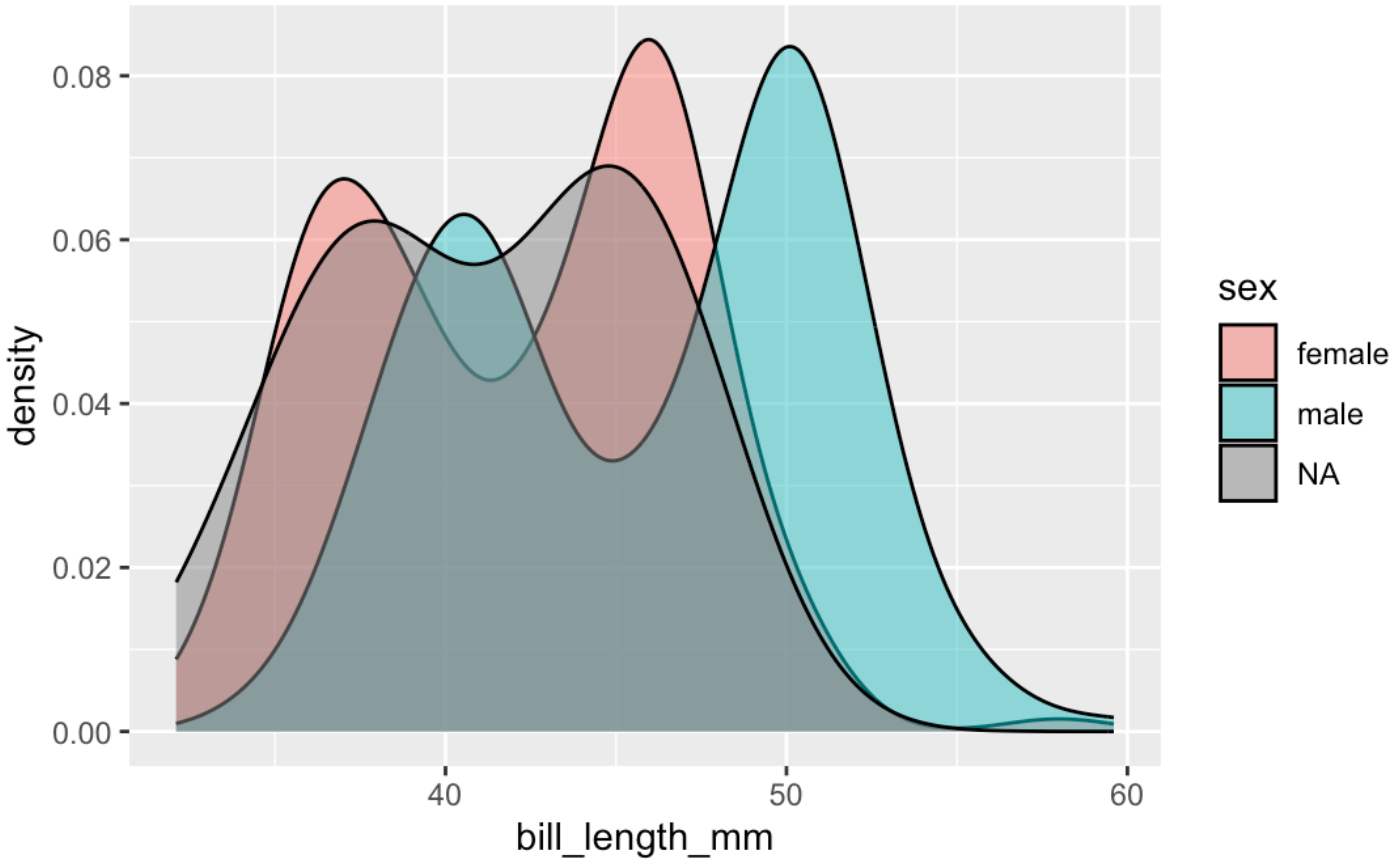
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Transparency for clarity

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density(aes(fill=sex), alpha=0.5)
```



data

ggplot(aes(mapping))

layers

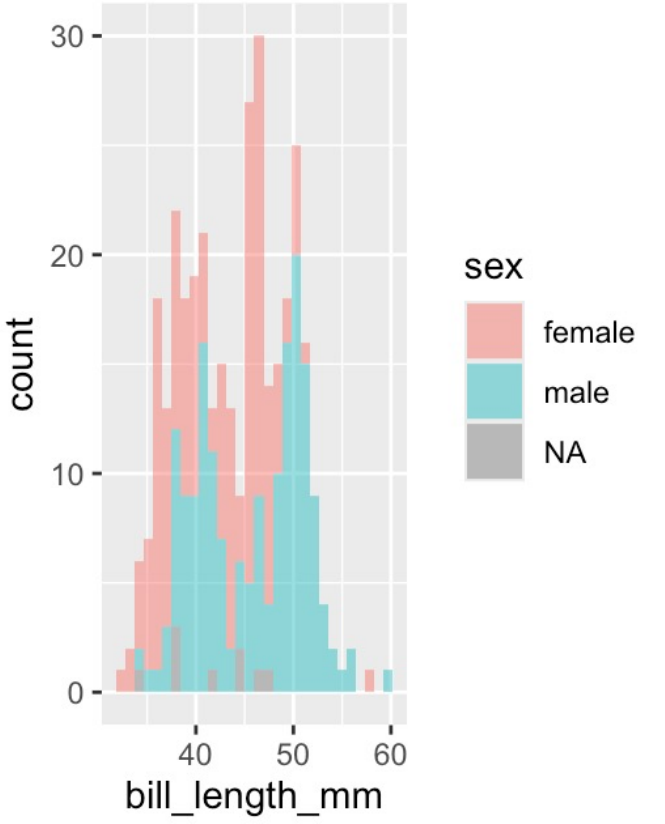
geom_function()

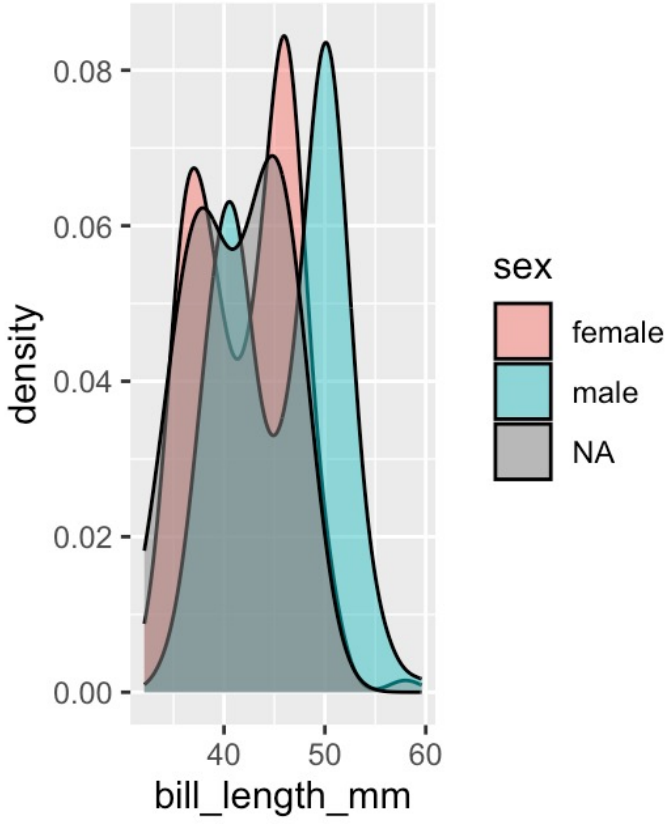
additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

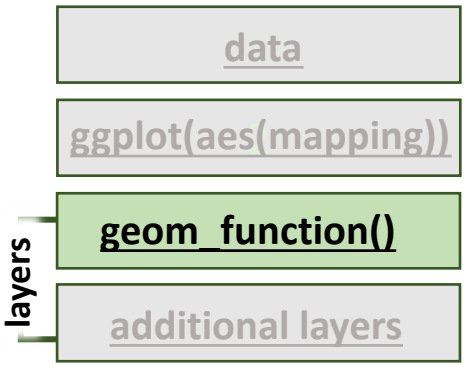
Other - Heatmaps: geom_tile

Histogram vs Density





What is the difference that you see? When would you use one vs another?



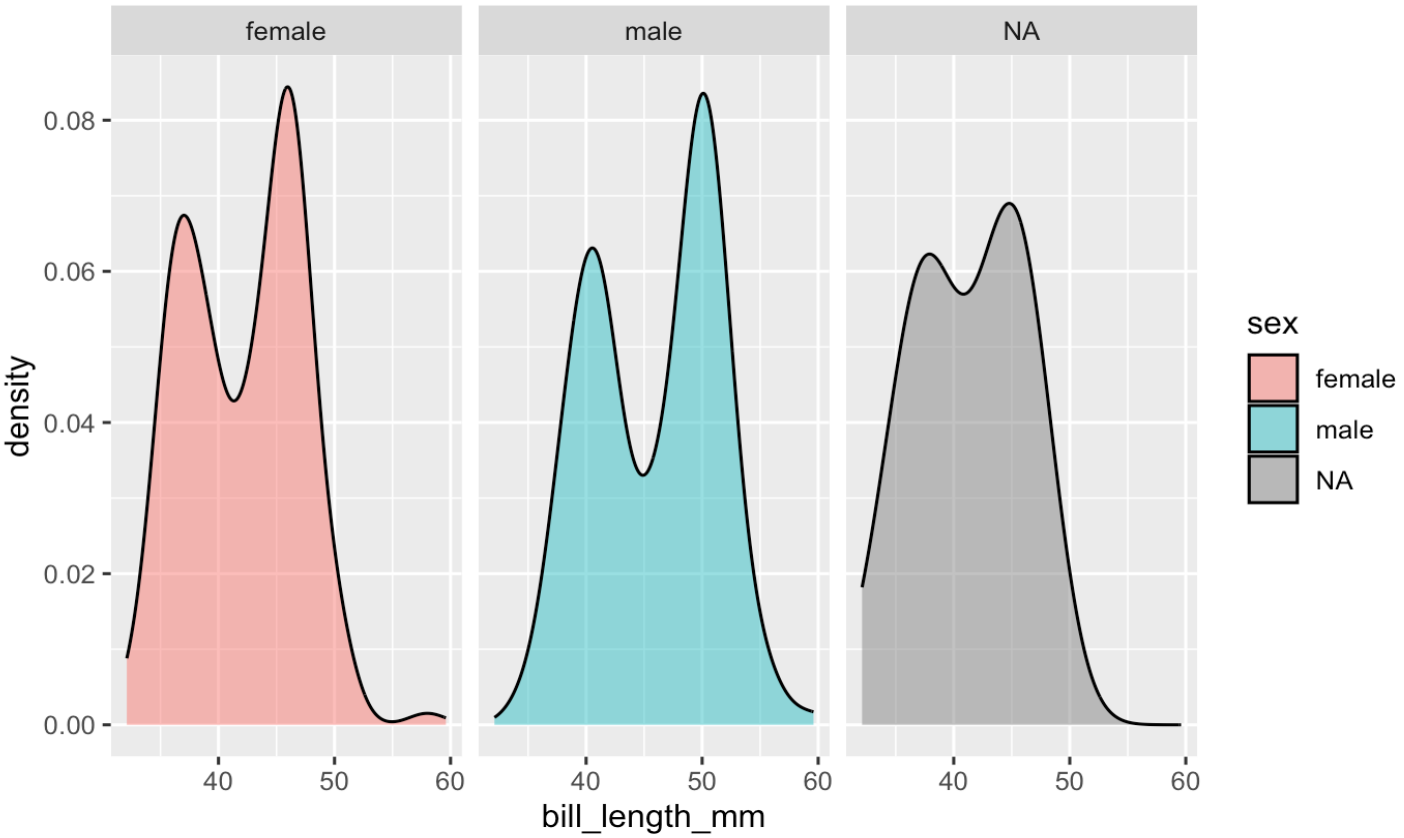
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

facet_wrap()

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density() +
  facet_wrap(~sex)
```

Remember to use ~



UNIVERSITY OF OREGON

data

ggplot(aes(mapping))

layers

geom_function()

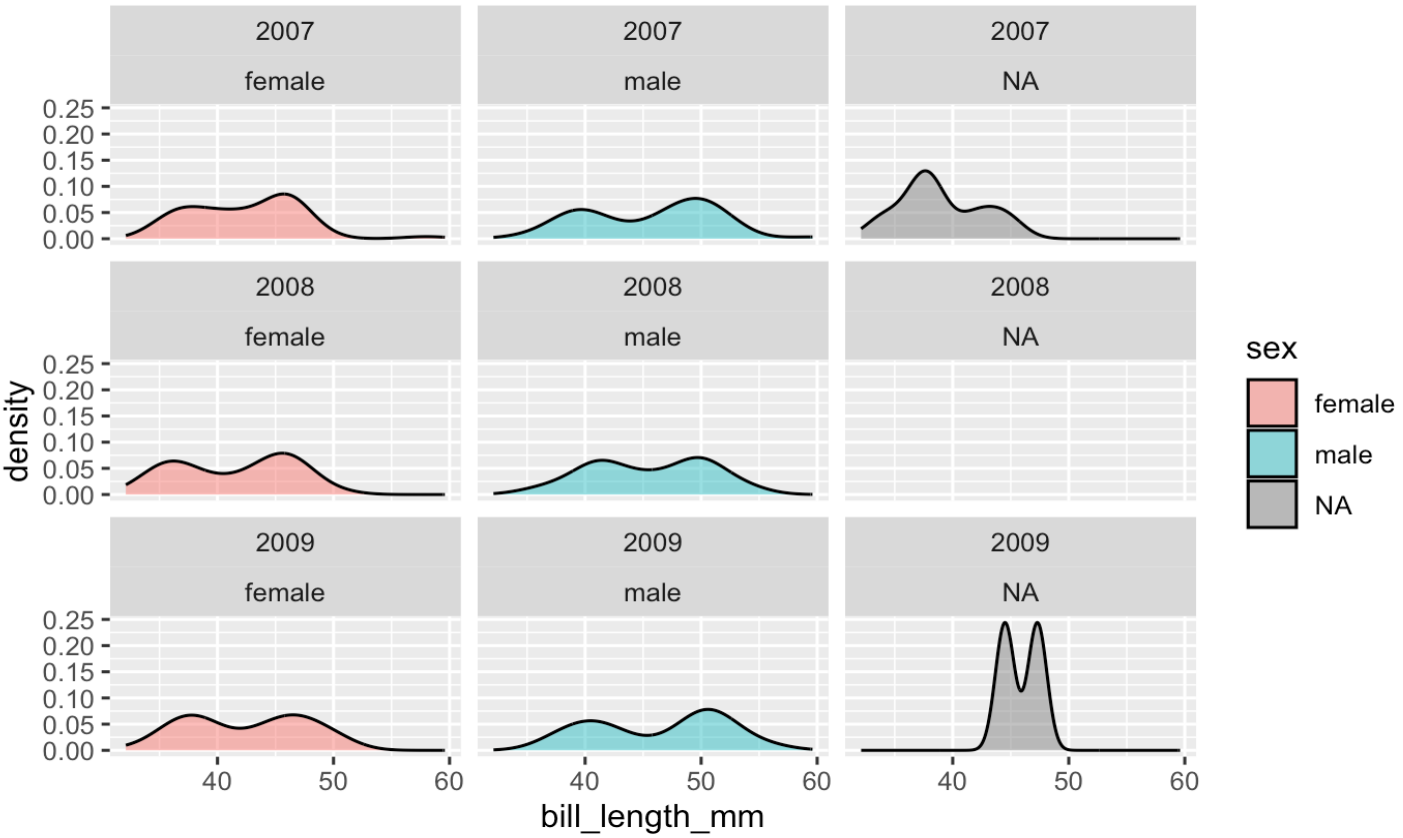
additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

⊞ facet_wrap()

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density() +
  facet_wrap(year~sex)
```



data

ggplot(aes(mapping))

layers

geom_function()

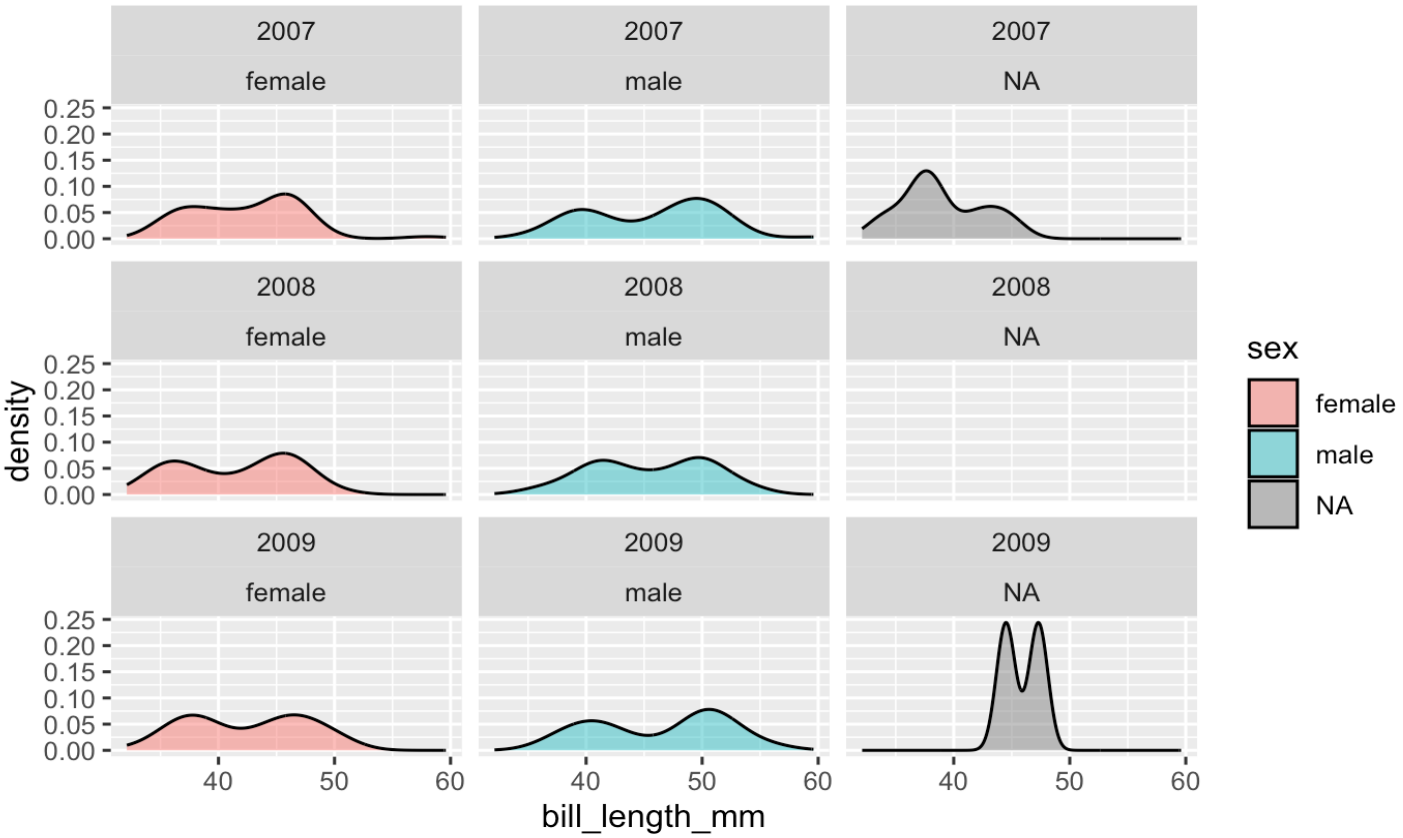
additional layers

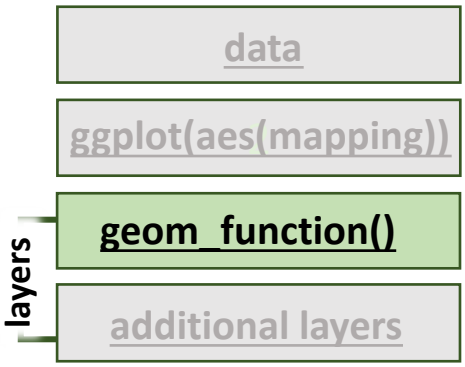
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

⊞ facet_wrap() – vars()

```
penguins |>
  ggplot(aes(x=bill_length_mm)) +
  geom_density() +
  facet_wrap(vars(year, sex))
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1: Continuous		geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1: Discrete		geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_density_ridges()

```
library(ggridges)

penguins |>
  ggplot(aes(bill_length_mm, sex)) +
  geom_density_ridges()
```

layers

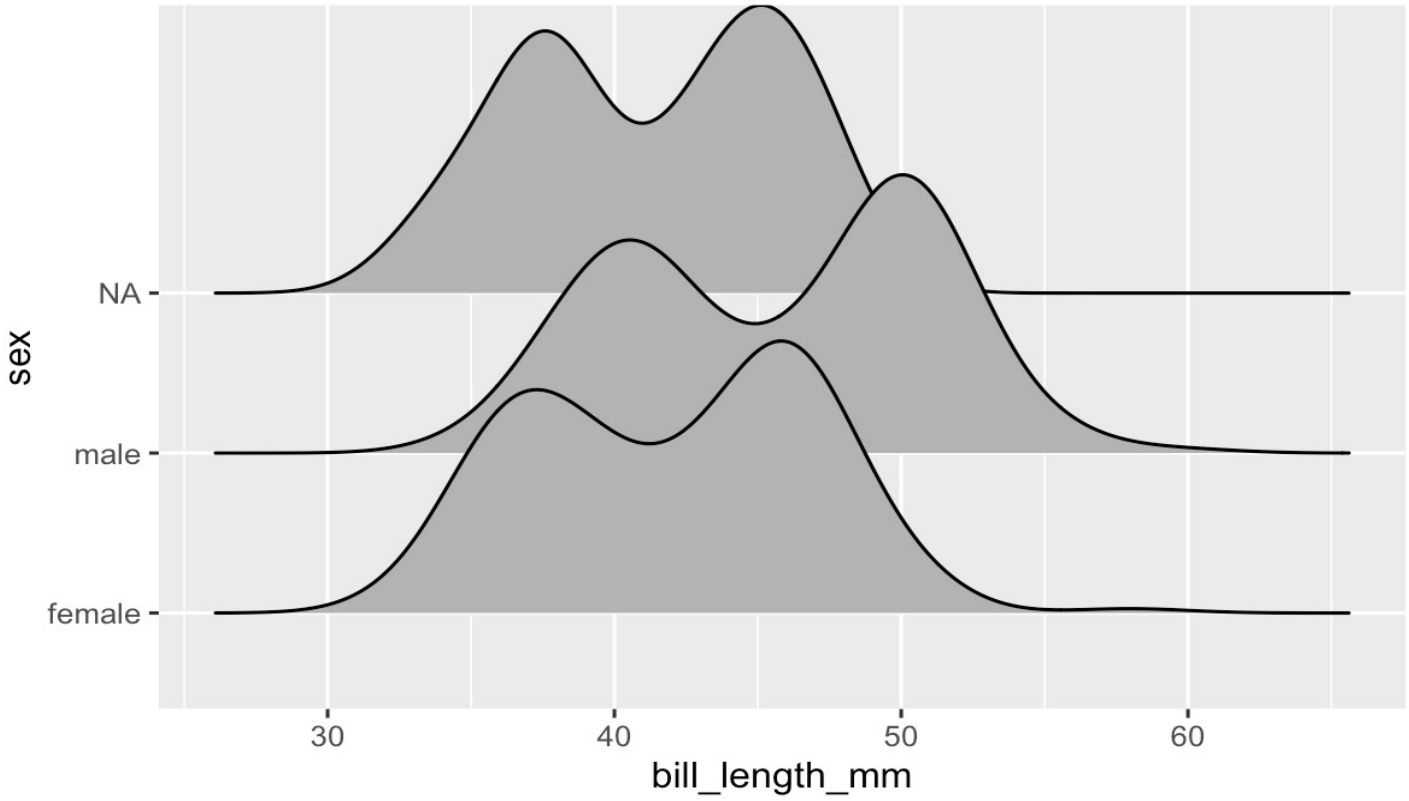
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

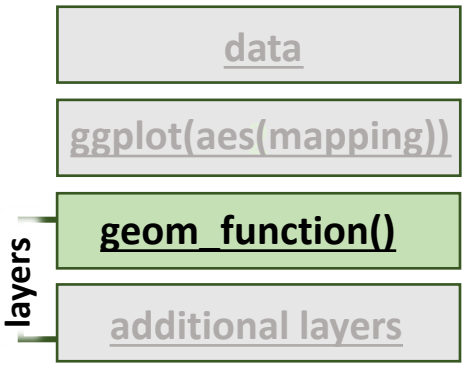
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_density_ridges()

```
library(ggridges)

penguins |>
  ggplot(aes(bill_length_mm, sex)) +
  geom_density_ridges()
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_point()

General research question:

- How are two numeric variables related? (raw observations)

Our data specific question:

- What is the relationship between penguin's bill length and body mass?

layers

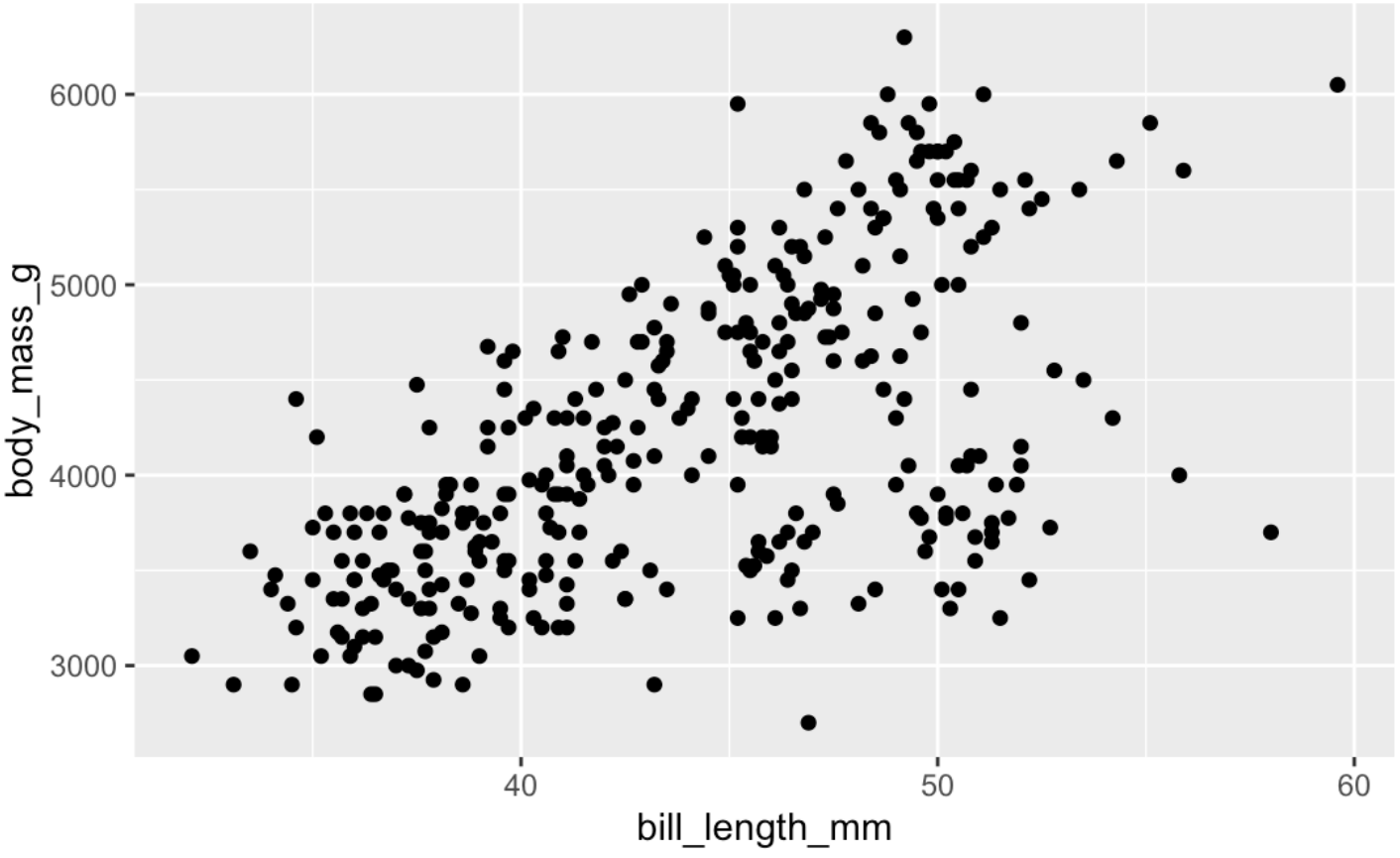
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

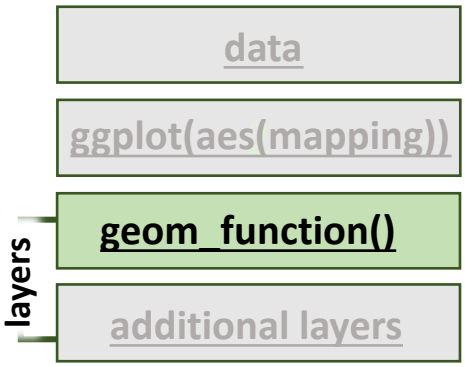
Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_point()

```
penguins |>
  ggplot(aes(x=bill_length_mm,
             y=body_mass_g)) +
  geom_point()
```





Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Color

```
penguins |>
  ggplot(aes(x=bill_length_mm,
             y=body_mass_g)) +
  geom_point(color="magenta")
```

Would this apply to all the data or certain data points?

layers

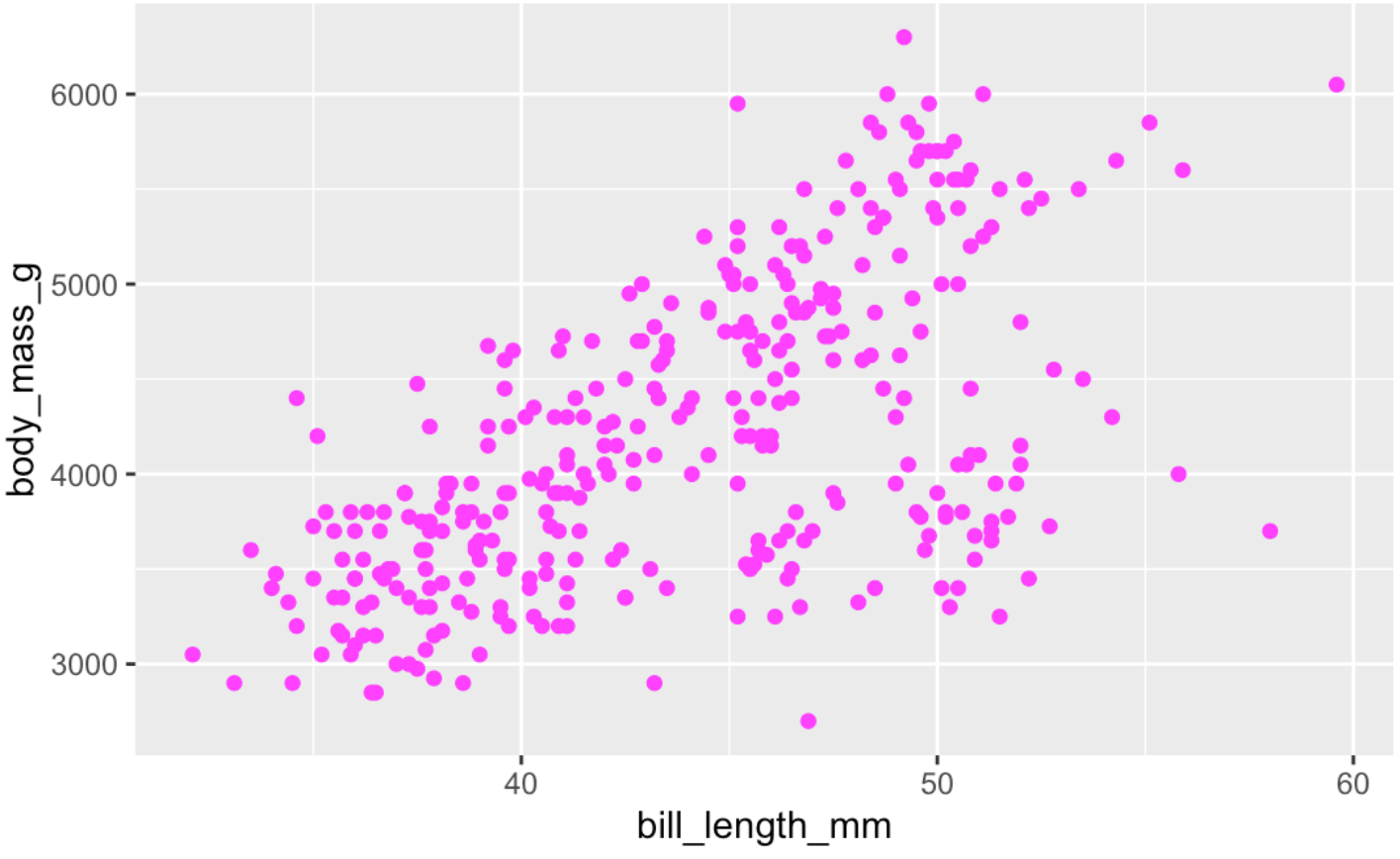
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

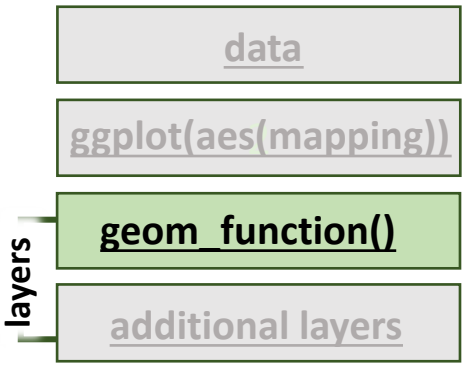
Other - Heatmaps: geom_tile

Color

```
penguins |>
  ggplot(aes(x=bill_length_mm,
             y=body_mass_g)) +
  geom_point(color="magenta")
```



UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Emphasis/Additional Layer

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)) +
  geom_point(color = "magenta") +
  geom_point(data=filter(penguins,
             island == "Torgersen"),
             color = "blue")
```

layers

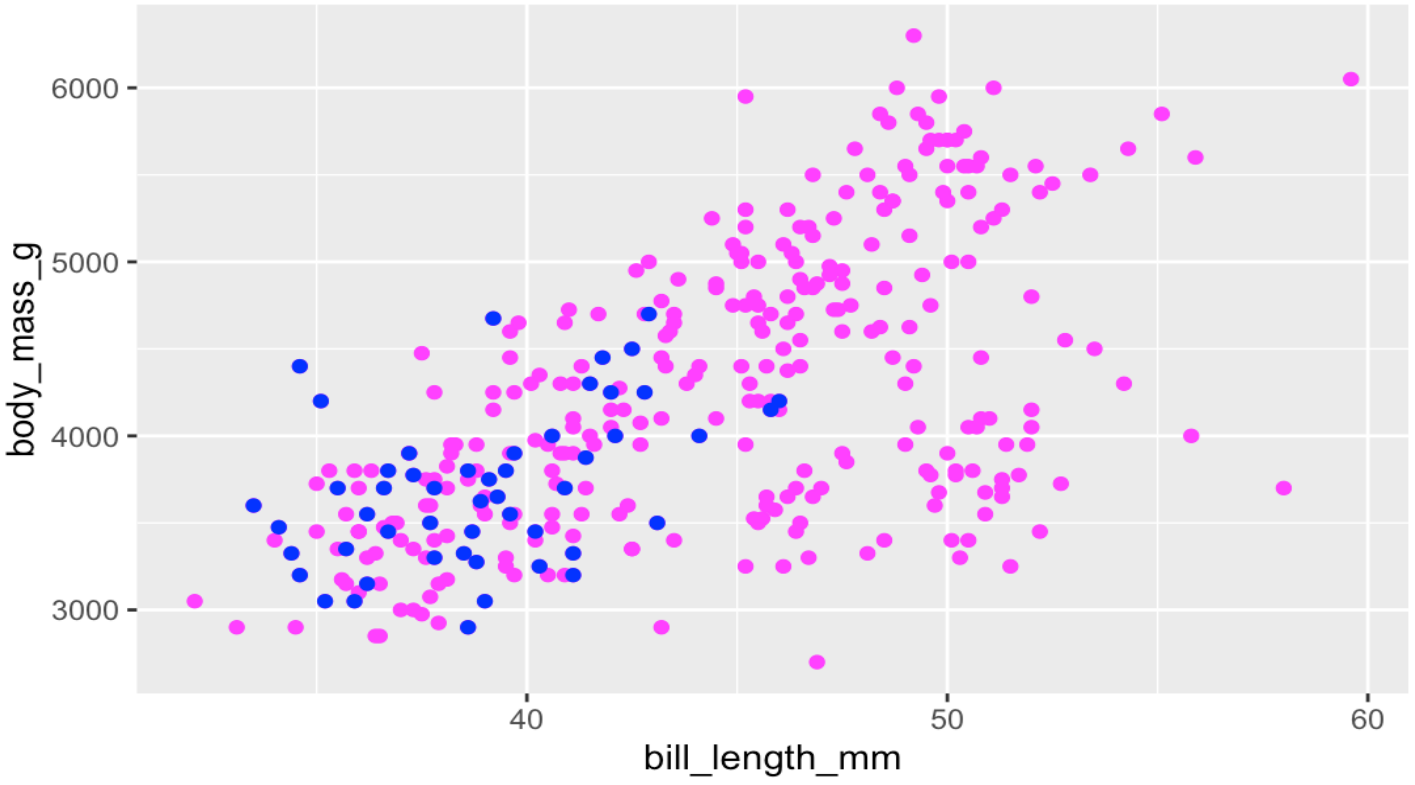
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

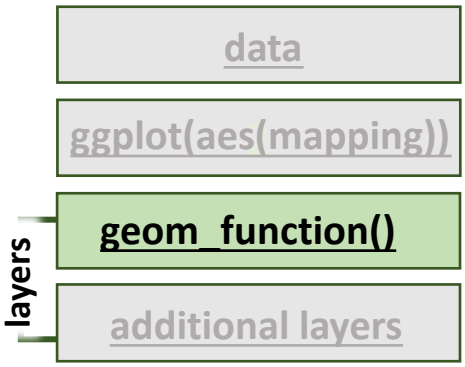
Other - Heatmaps: geom_tile

Emphasis/Additional Layer

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)) +
  geom_point(color = "magenta") +
  geom_point(data=filter(penguins,
             island == "Torgersen"),
             color = "blue")
```



UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Emphasis/Additional Layer

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)) +
  geom_point(data=filter(penguins,
             island == "Torgersen"),
             color = "blue") +
  geom_point(color = "magenta"
```

What do you think will happen if we switched the order of the two geom_points?

layers

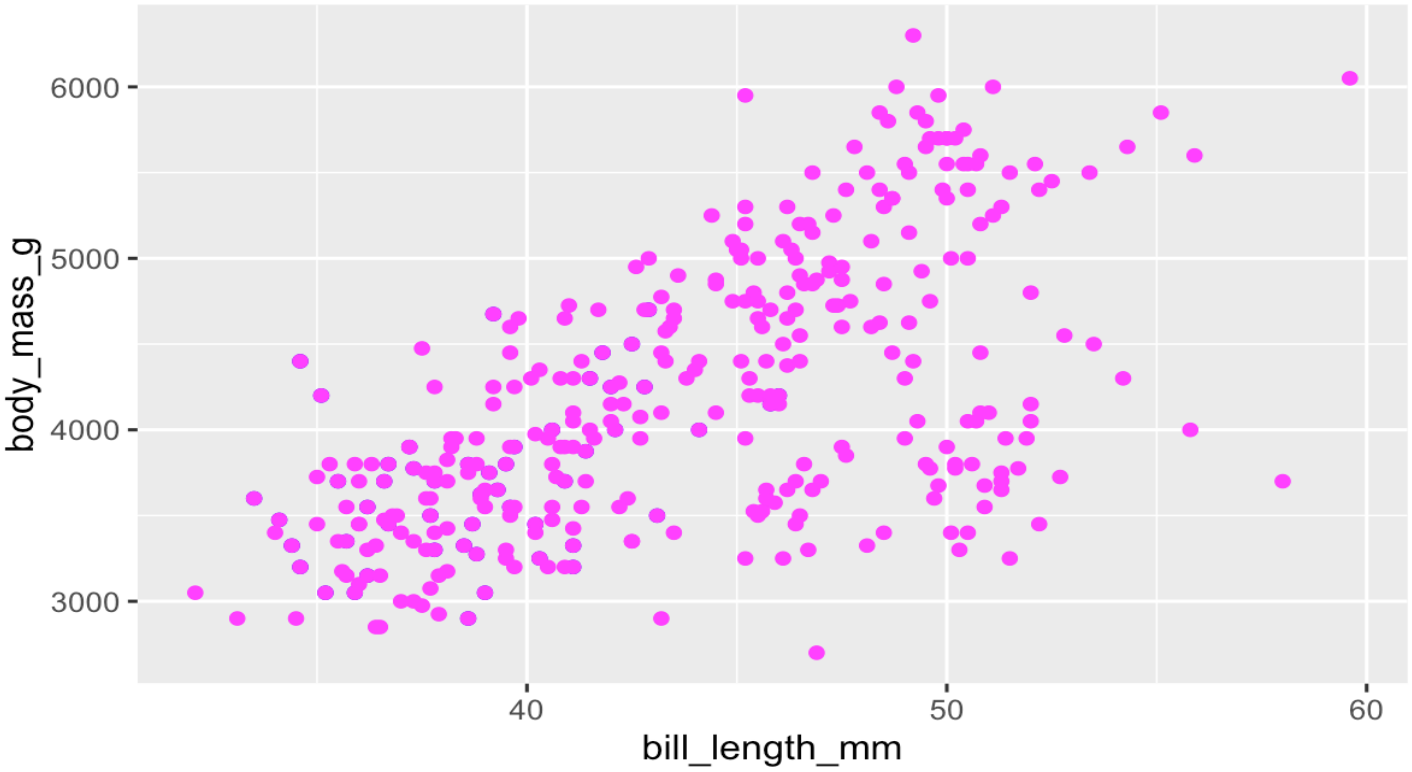
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

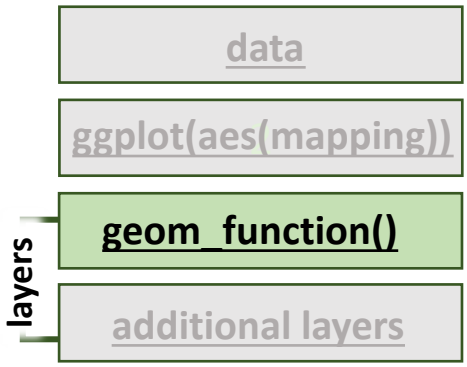
Other - Heatmaps: geom_tile

Emphasis/Additional Layer

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)) +
  geom_point(data=filter(penguins,
             island == "Torgersen"),
             color = "blue") +
  geom_point(color = "magenta")
```



UNIVERSITY OF OREGON

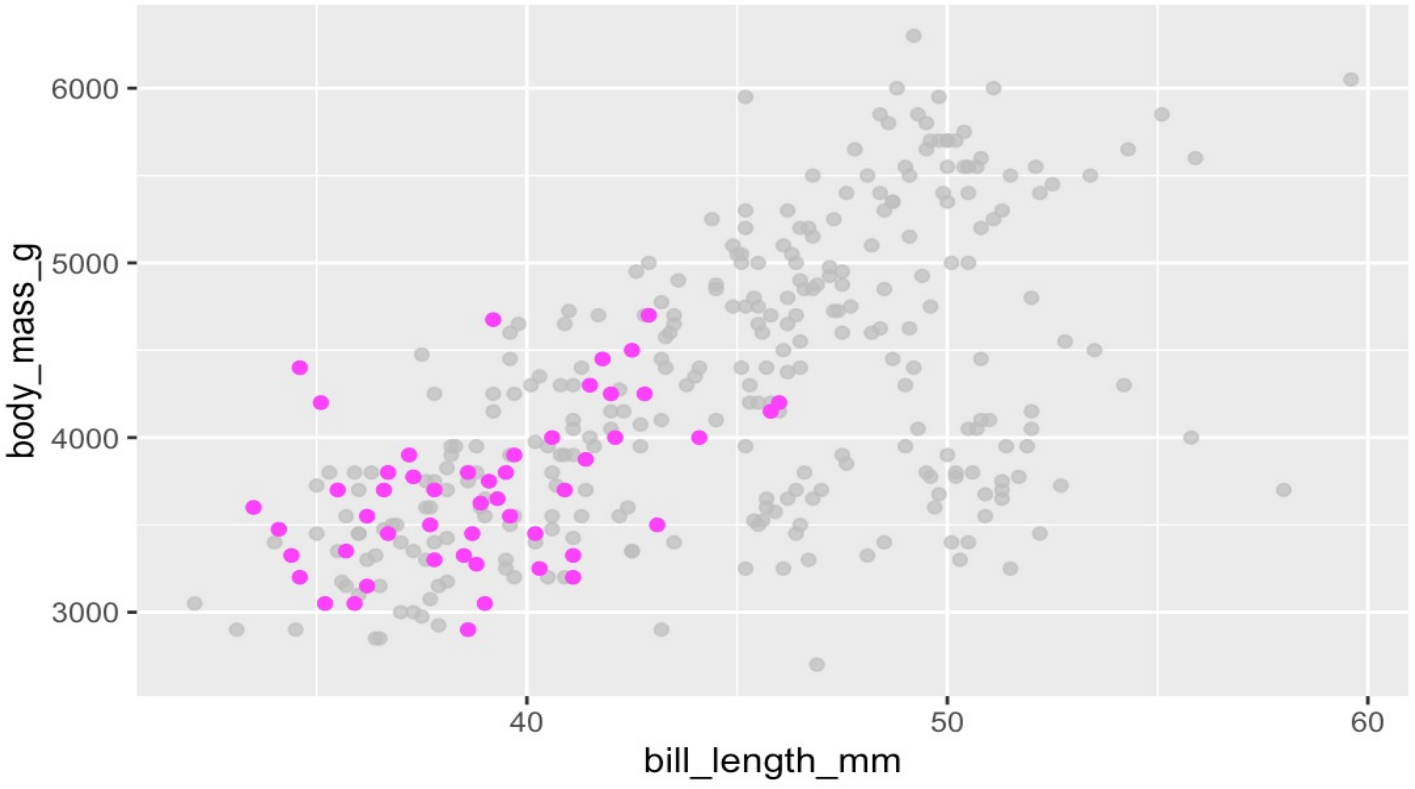


Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	**geom_point** geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

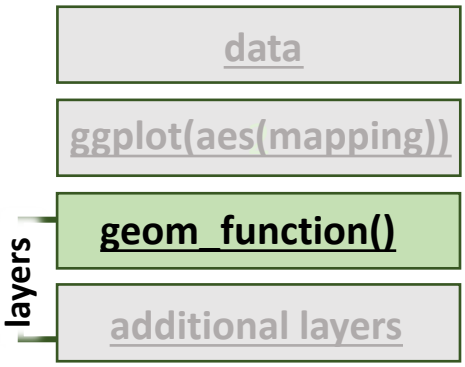
Other - Heatmaps: geom_tile

Emphasis (another way)

```
penguins |>
  ggplot(aes(x = bill_length_mm,
             y = body_mass_g)) +
  geom_point(color = "magenta") +
  gghighlight::gghighlight(island ==
                           "Torgersen")
```



UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1: Continuous		geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1: Discrete		geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_smooth()

General research question:

- What is the pattern of relationship of two continuous variables? (trend)

Our data specific question:

- What is the trend or pattern of relationship between penguin's bill length and body mass?

layers

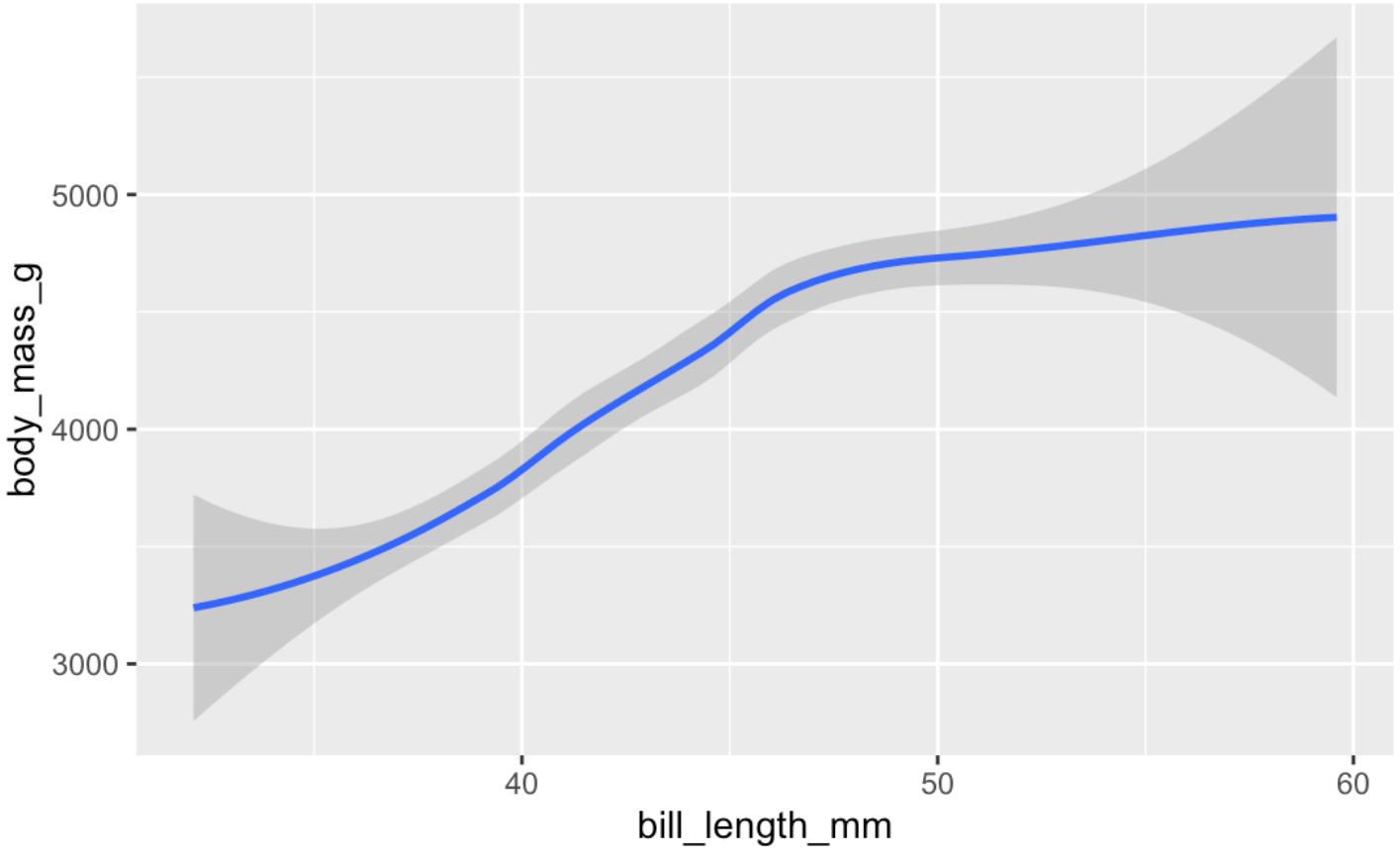
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_smooth()

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth()
```



data

ggplot(aes(mapping))

layers

geom_function()

additional layers

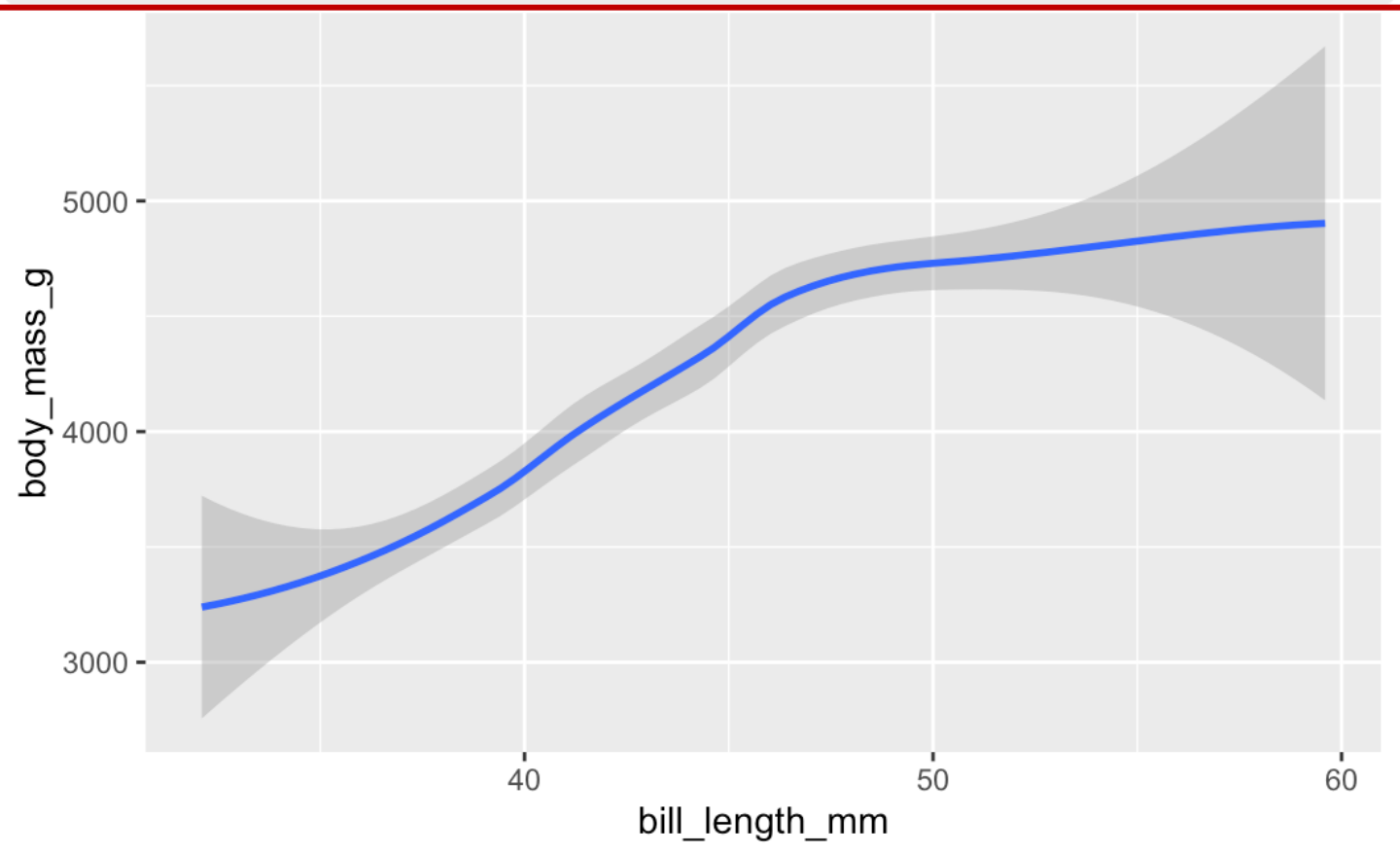
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_smooth()

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth()
```

```
[38;5;232m`geom_smooth()` using method = 'loess' and formula = 'y ~ x' [39m
```



layers

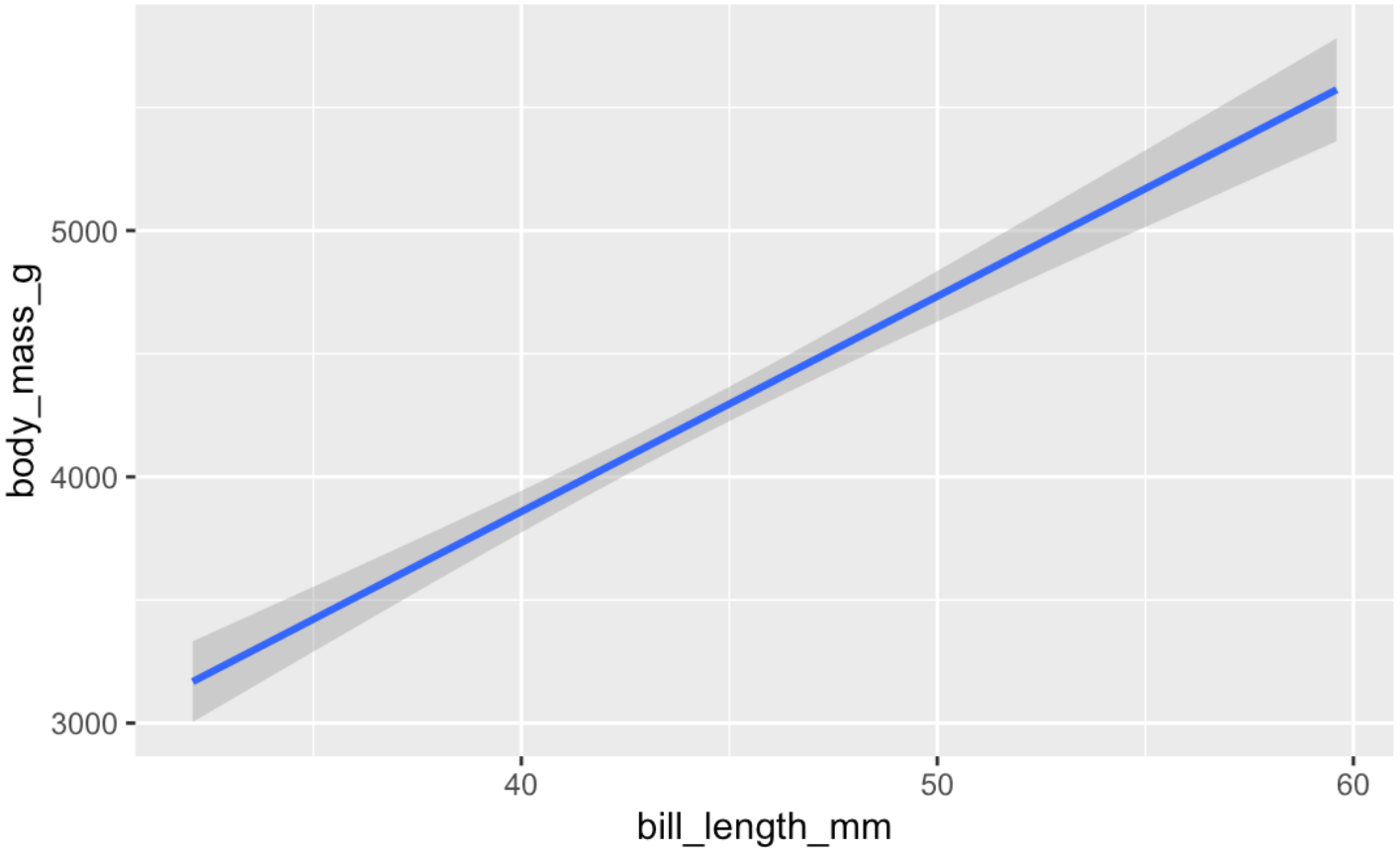
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Method

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth(method = "lm")
```



data

ggplot(aes(mapping))

layers

geom_function()

additional layers

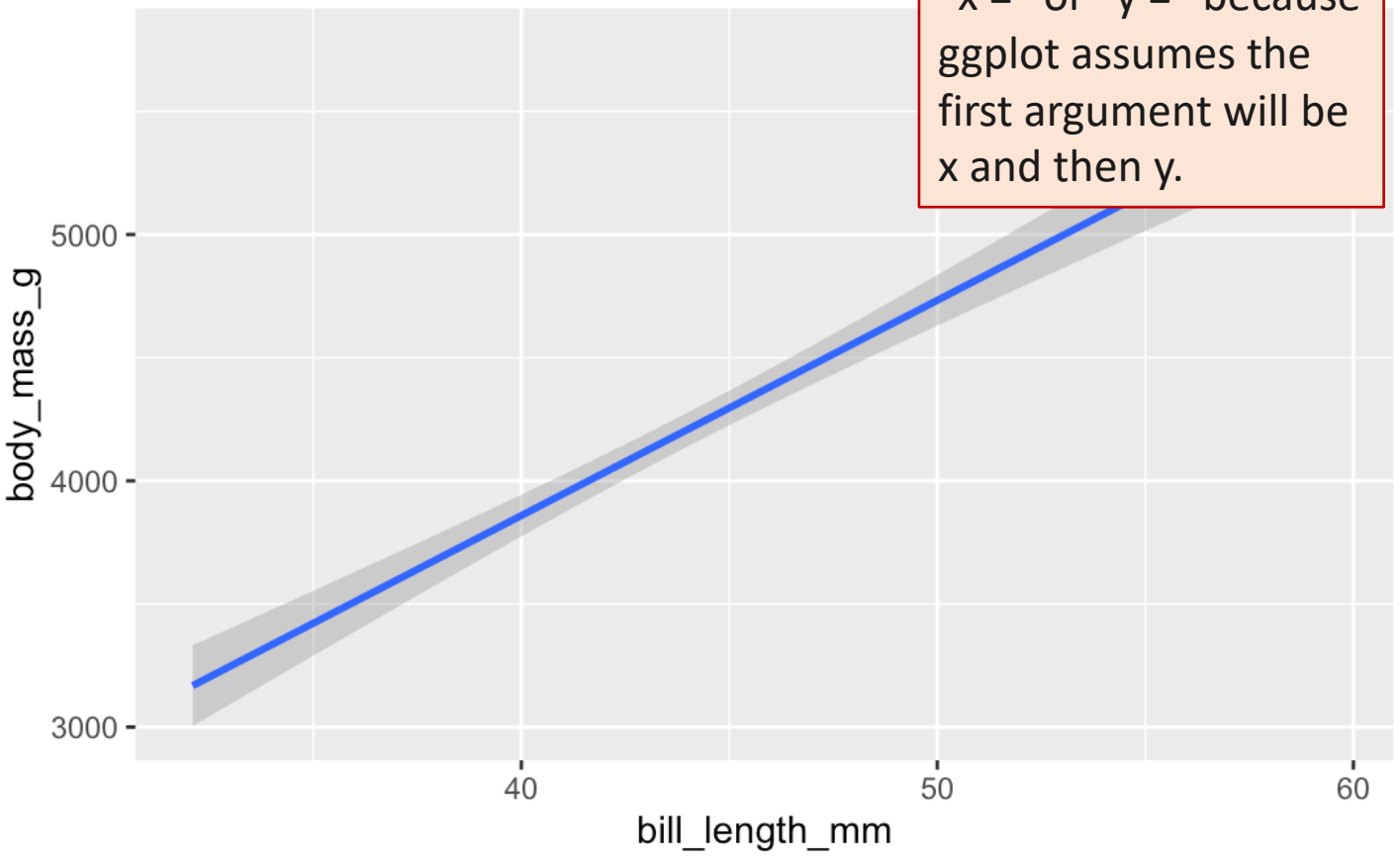
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1 Continuous		geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1 Discrete		geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Method

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth(method = "lm")
```

No need to include "x =" or "y =" because ggplot assumes the first argument will be x and then y.



UNIVERSITY OF OREGON

layers

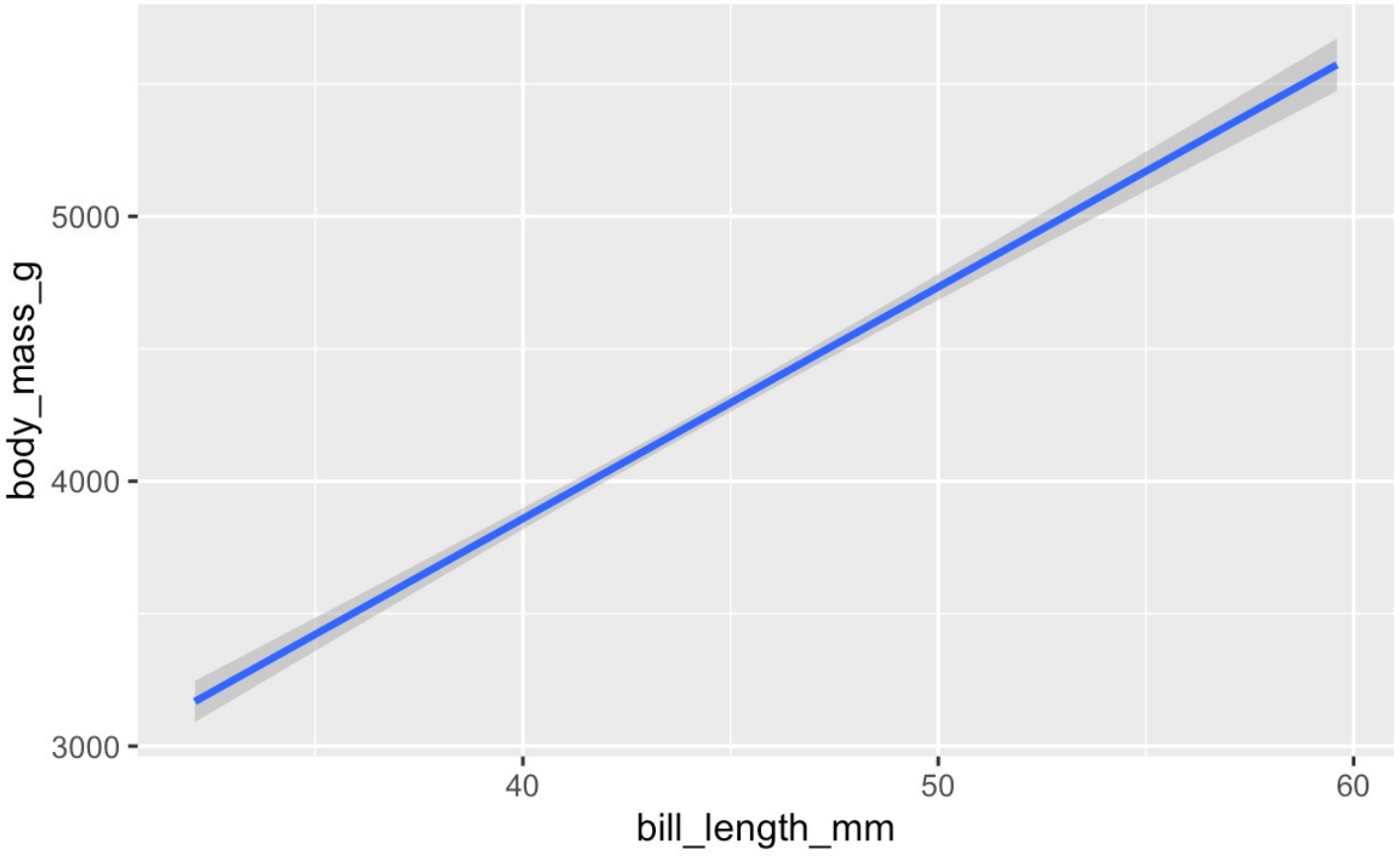
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

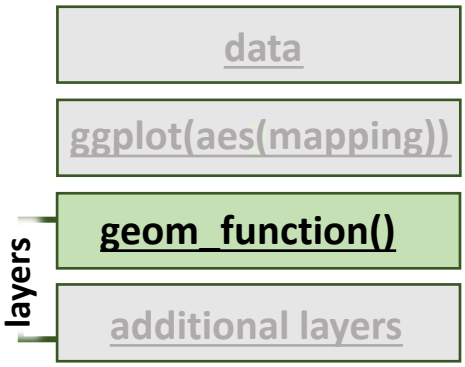
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Method

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth(method = "lm", level = .65)
```



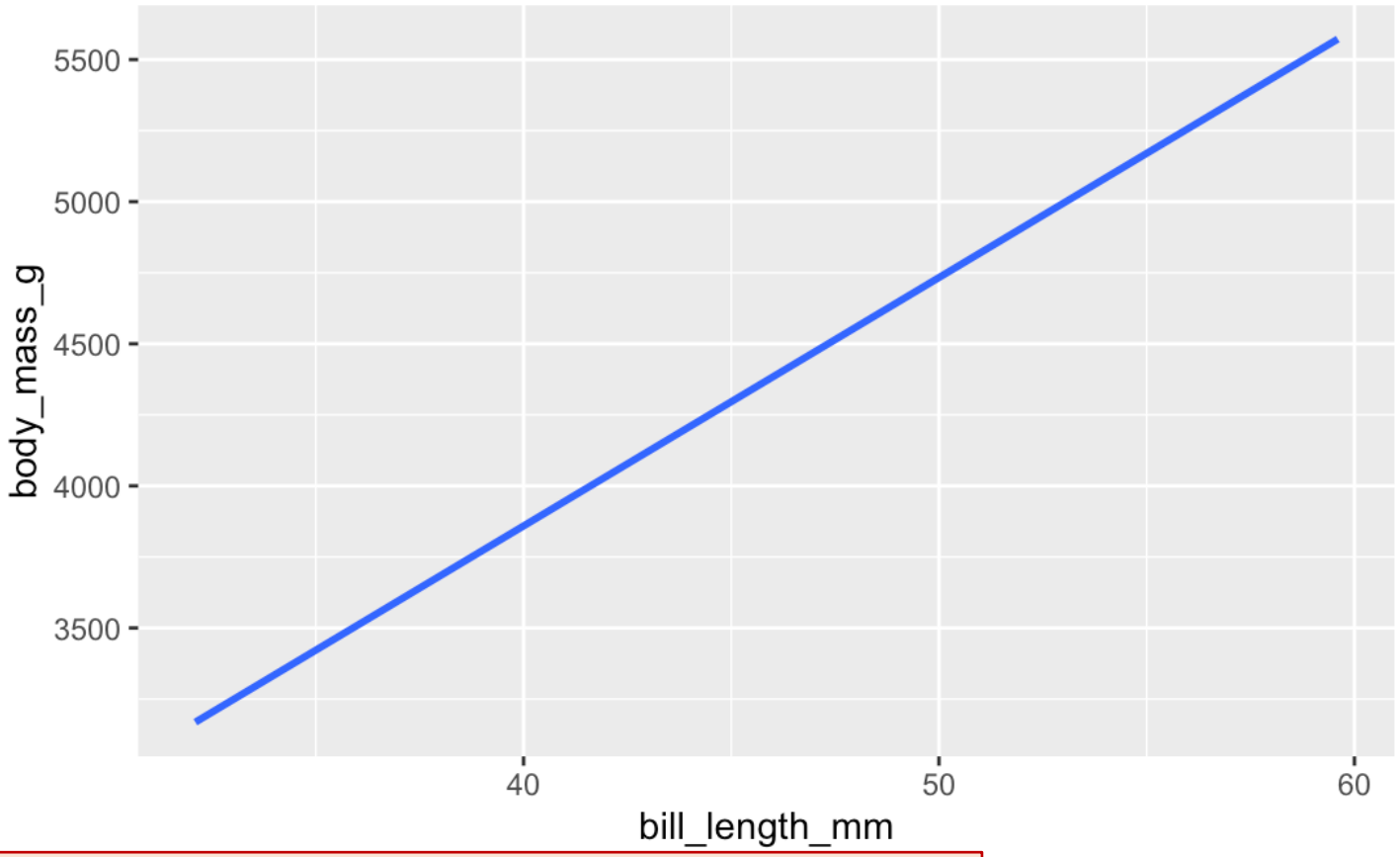


		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

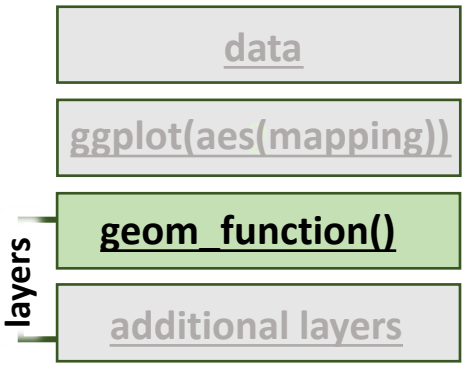
Method

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_smooth(method = "lm", se = FALSE)
```



Note: This is not the same as geom_line(). We are fitting a line of best fit with geom_smooth()

UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Adding layers

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_point() +
  geom_smooth(method = "lm")
```

layers

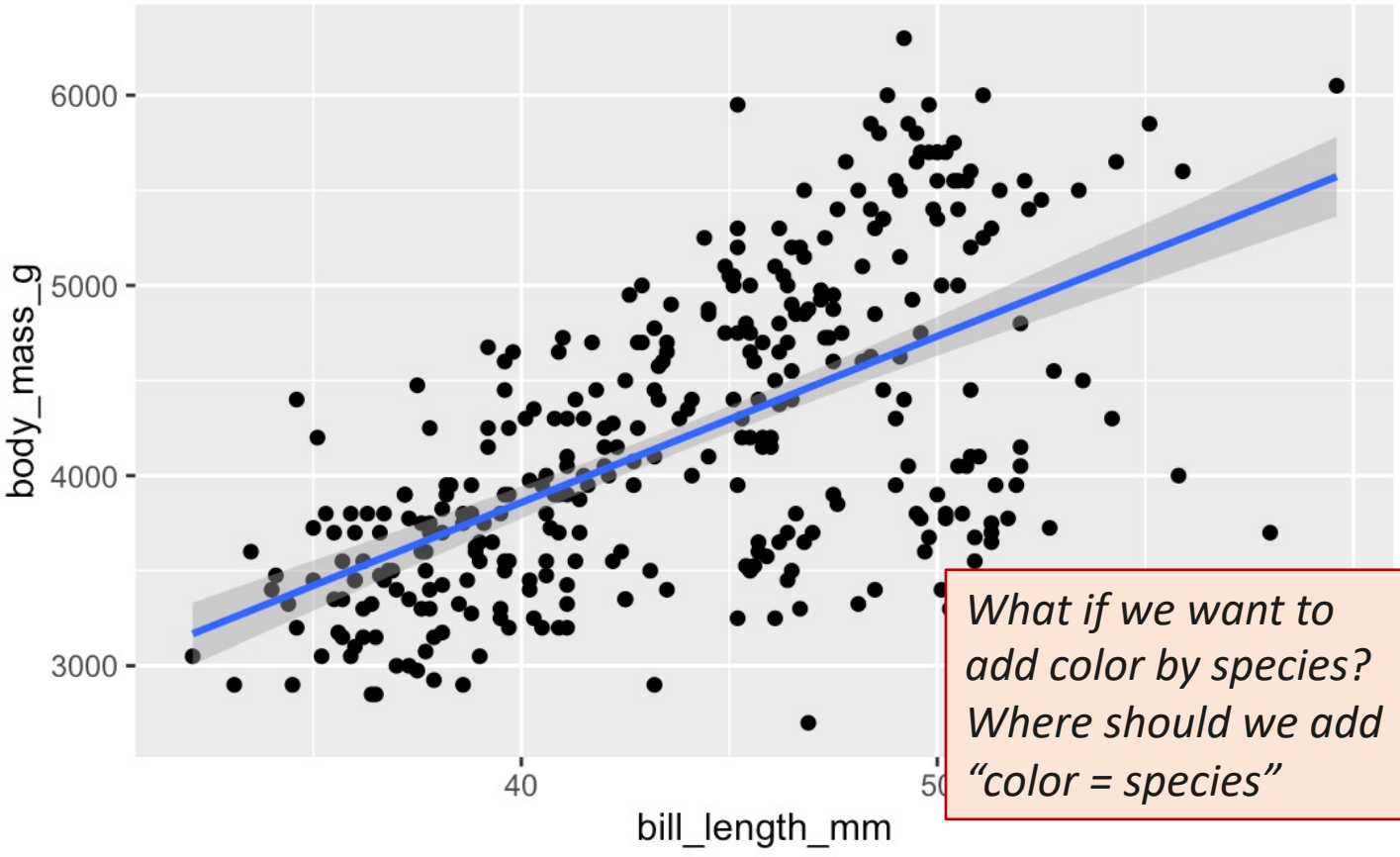
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

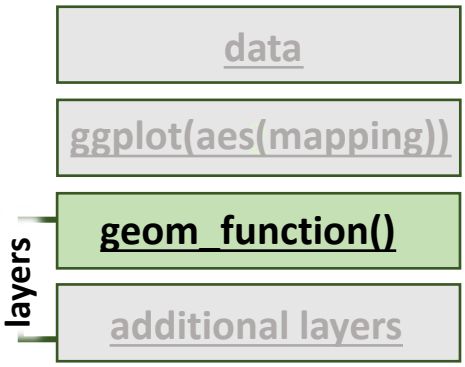
Other - Heatmaps: geom_tile

Adding layers

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_point() +
  geom_smooth(method = "lm")
```



What if we want to add color by species? Where should we add "color = species"



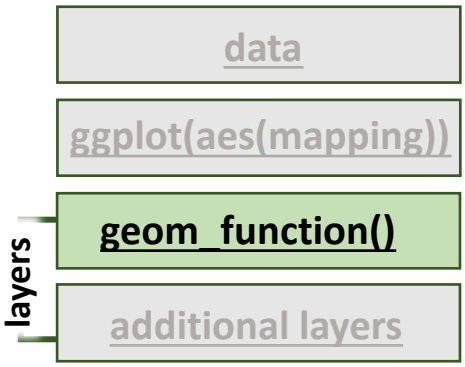
Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Global

If we use something like color = “x” in the first aesthetic, it will carry on through all additional layers.

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g,
             color = species)) +
  geom_point() +
  geom_smooth(method = “lm”)
```



Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Global

If we use something like color = "x" in the first aesthetic, it will carry on through all additional layers.

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g,
             color = species)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g))+
  geom_point(aes(color = species)) +
  geom_smooth(aes(color = species),
              method = "lm")
```

layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

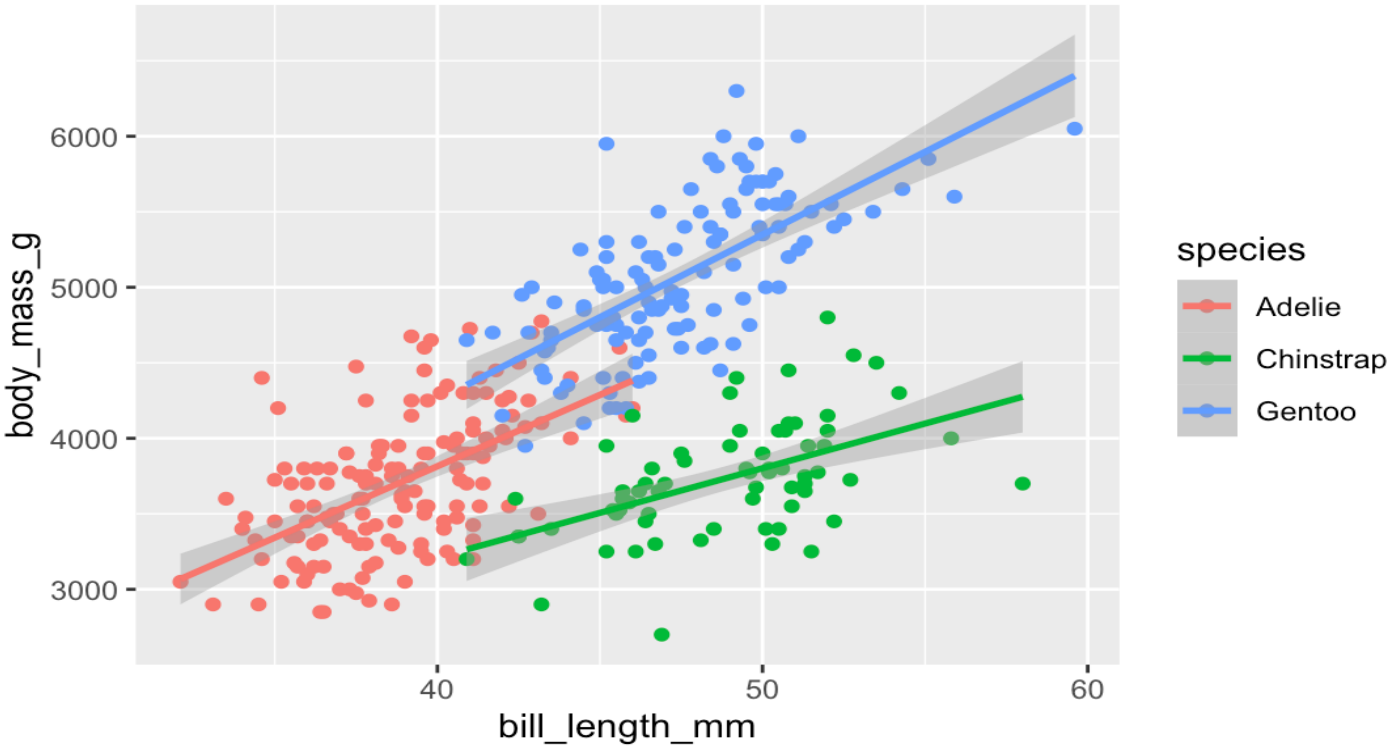
Other - Heatmaps: geom_tile

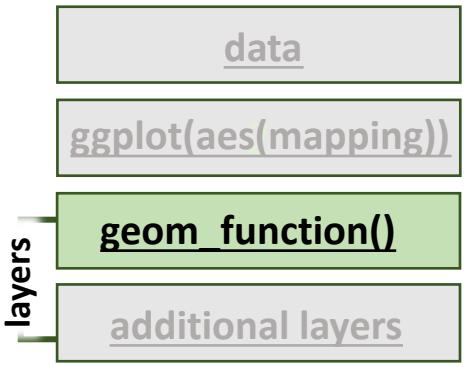
Global

If we use something like color = "x" in the first aesthetic, it will carry on through all additional layers.

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g,
             color = species)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g))+
  geom_point(aes(color = species)) +
  geom_smooth(aes(color = species),
              method = "lm")
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Local

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g))+
  geom_point(aes(color = species)) +
  geom_smooth(method = "lm")
```

How would this plot look like?

layers

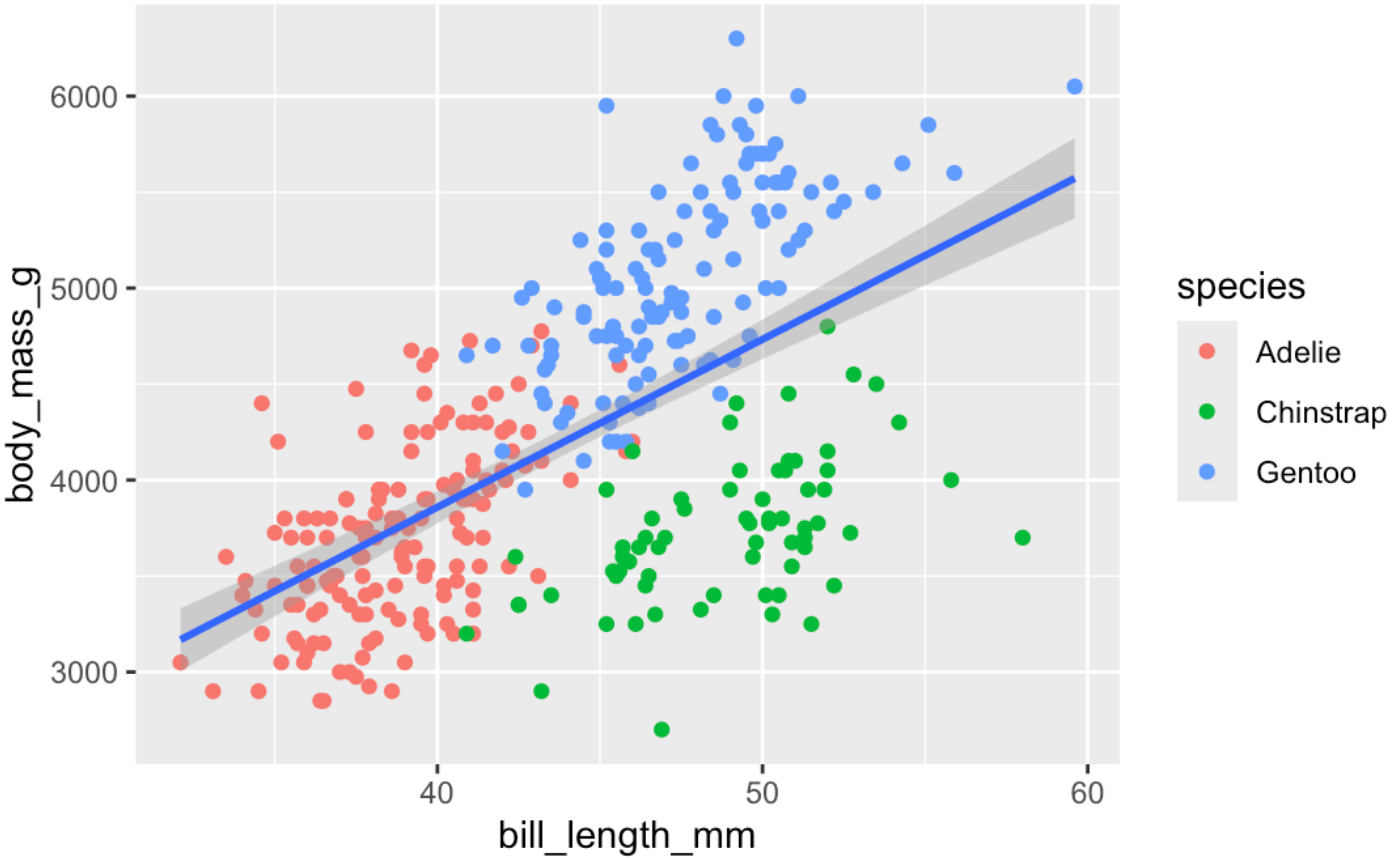
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

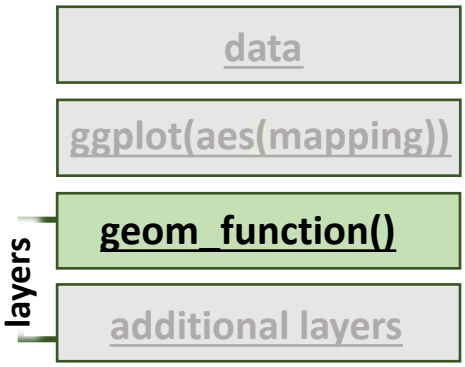
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Local

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g))+
  geom_point(aes(color = species)) +
  geom_smooth(method = "lm")
```





Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

geom_point: raw data, not linked

geom_smooth: trend/pattern

geom_line: raw data linked

layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

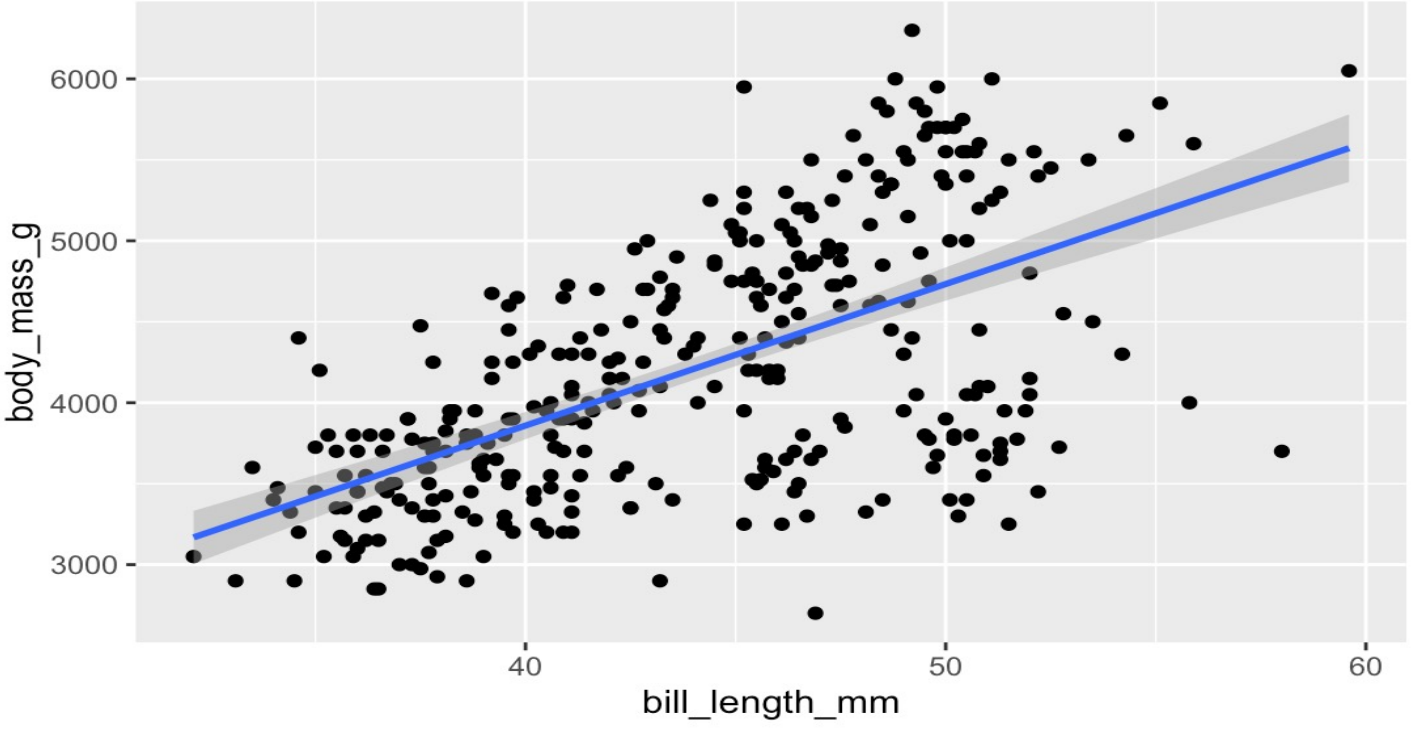
Other - Heatmaps: geom_tile

geom_line()

geom_point: raw data, not linked

geom_smooth: trend/pattern

geom_line: raw data linked



layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1: Continuous		geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1: Discrete		geom_bar	X	geom_count

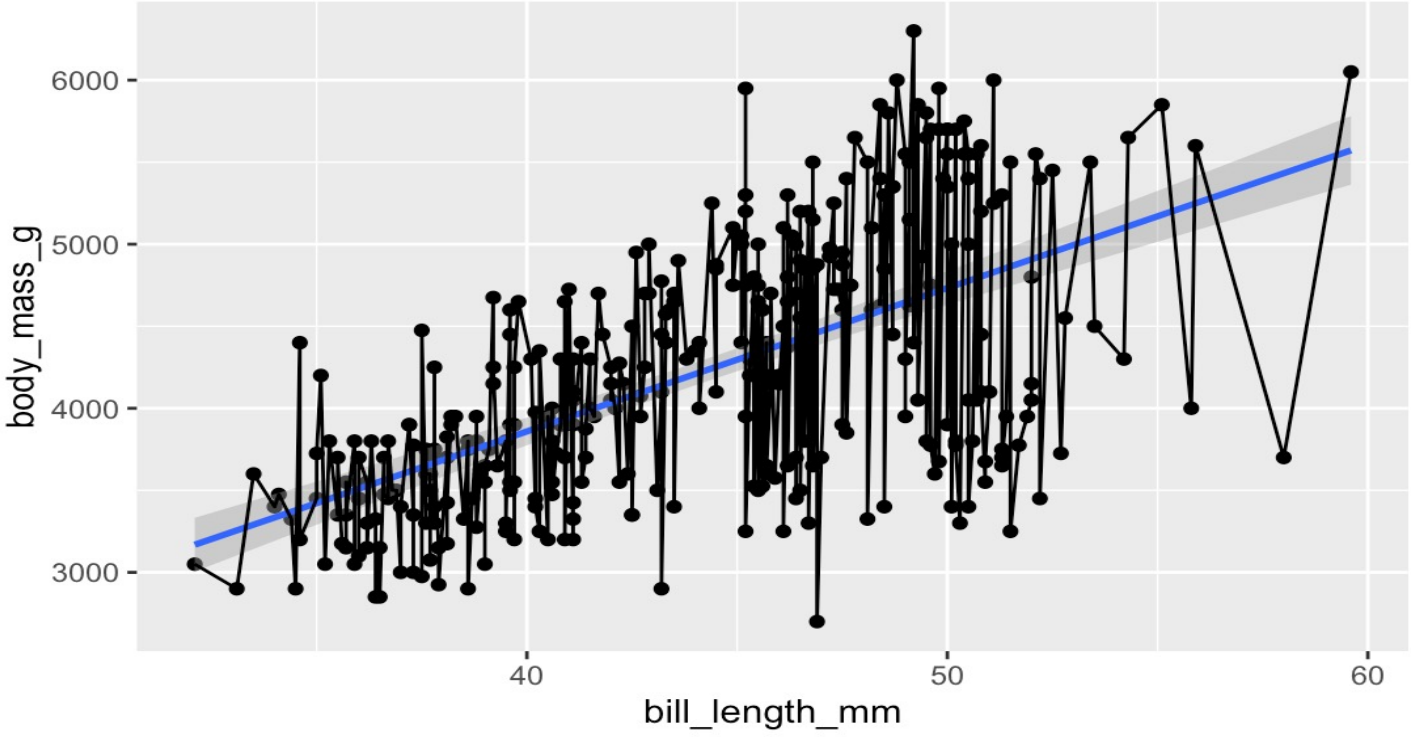
Other - Heatmaps: geom_tile

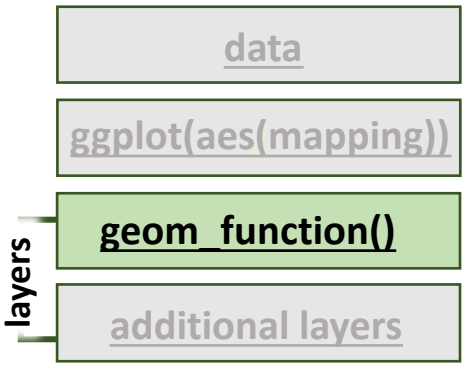
geom_line()

geom_point: raw data, not linked

geom_smooth: trend/pattern

geom_line: raw data linked



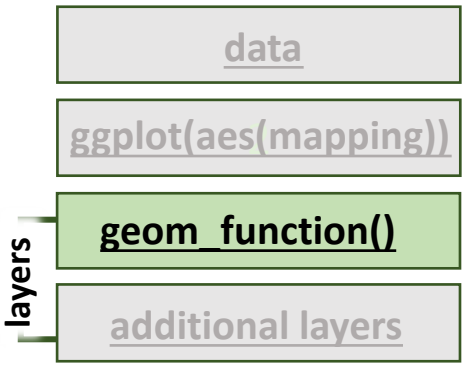


Variable 1	Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

- When should you use line plots?
  - Usually when time is involved
  - One time point per line or per group
  - Shows linkage



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

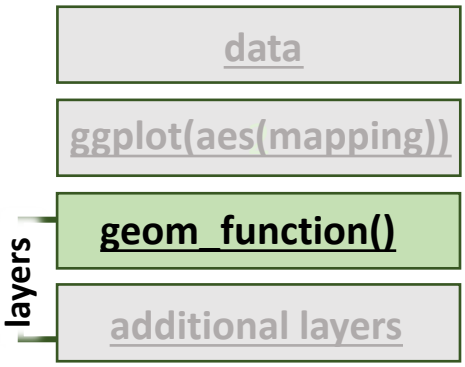
```
head(penguins)
```

A tibble: 6 x 8

species <fctr>	island <fctr>	bill_length_mm <dbl>	bill_depth_mm <dbl>	flipper_length_mm <int>	body_mass_g <int>	sex <fctr>	year <int>
Adelie	Torgersen	39.1	18.7	181	3750	male	2007
Adelie	Torgersen	39.5	17.4	186	3800	female	2007
Adelie	Torgersen	40.3	18.0	195	3250	female	2007
Adelie	Torgersen	*NA*	*NA*	*NA*	*NA*	*NA*	2007
Adelie	Torgersen	36.7	19.3	193	3450	female	2007
Adelie	Torgersen	39.3	20.6	190	3650	male	2007

6 rows

New data: One data point for each year



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

```
penguins_year <- penguins |>
  group_by(year) |>
  summarize(avg_bill = mean(bill_length_mm,
  na.rm=TRUE))
head(penguins_year)
```

A tibble: 3 × 2

year <int>	avg_bill <dbl>
2007	43.74037
2008	43.54123
2009	44.45294

3 rows

layers

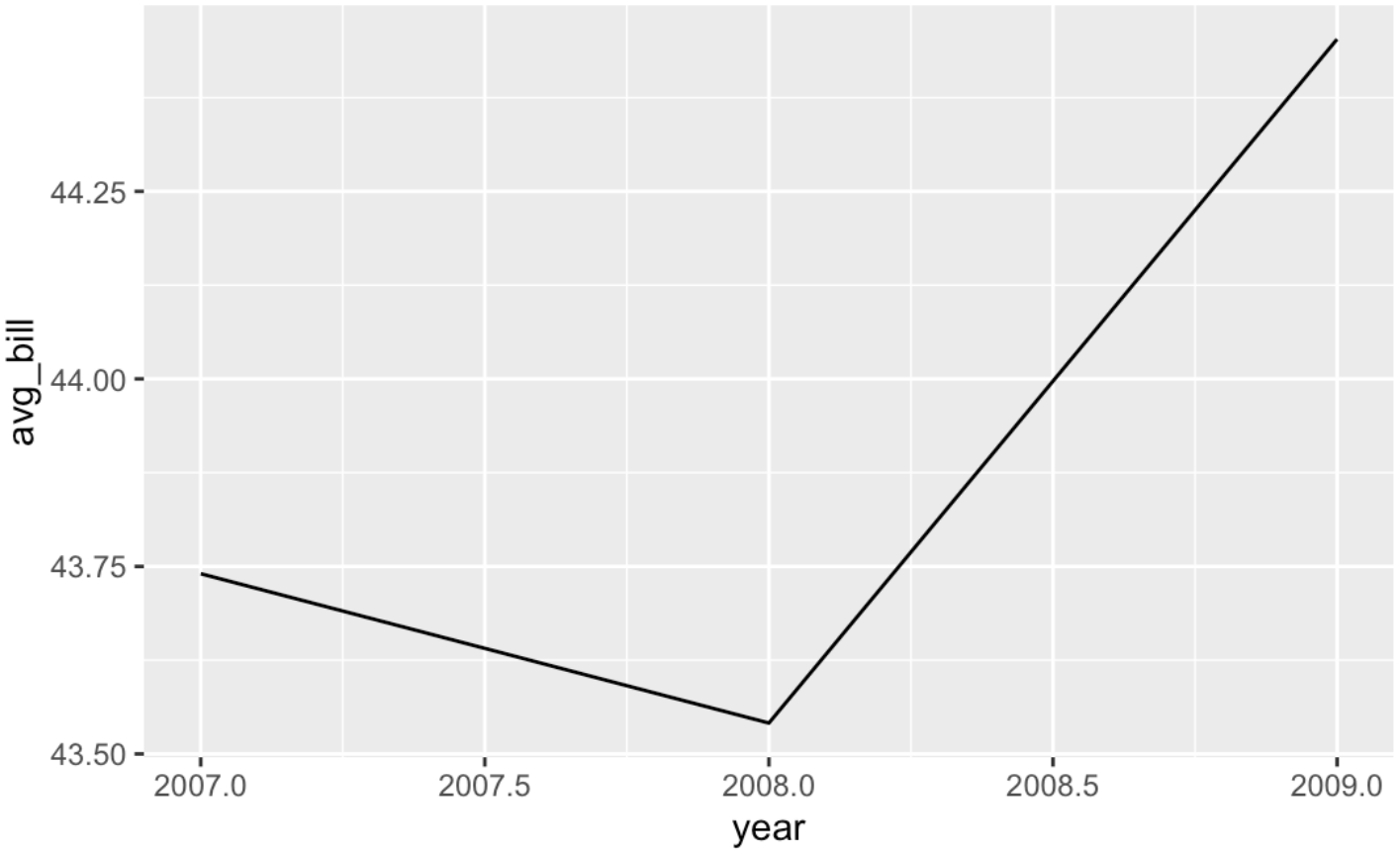
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

```
penguins_year |>
  ggplot(aes(year, avg_bill)) +
  geom_line()
```



layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

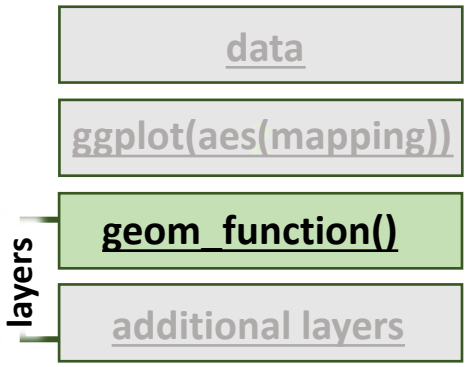
```
penguins_year_species <- penguins |>
  group_by(year, island) |>
  summarize(avg_bill = mean(bill_length_mm,
  na.rm=TRUE))

head(penguins_year_species)
```

A tibble: 6 × 3 | Groups: year [2]

year <int>	island <fctr>	avg_bill <dbl>
2007	Biscoe	45.03864
2007	Dream	44.53913
2007	Torgersen	38.80000
2008	Biscoe	44.62031
2008	Dream	43.75588
2008	Torgersen	38.76875

6 rows

New data: One data point for each year for each species

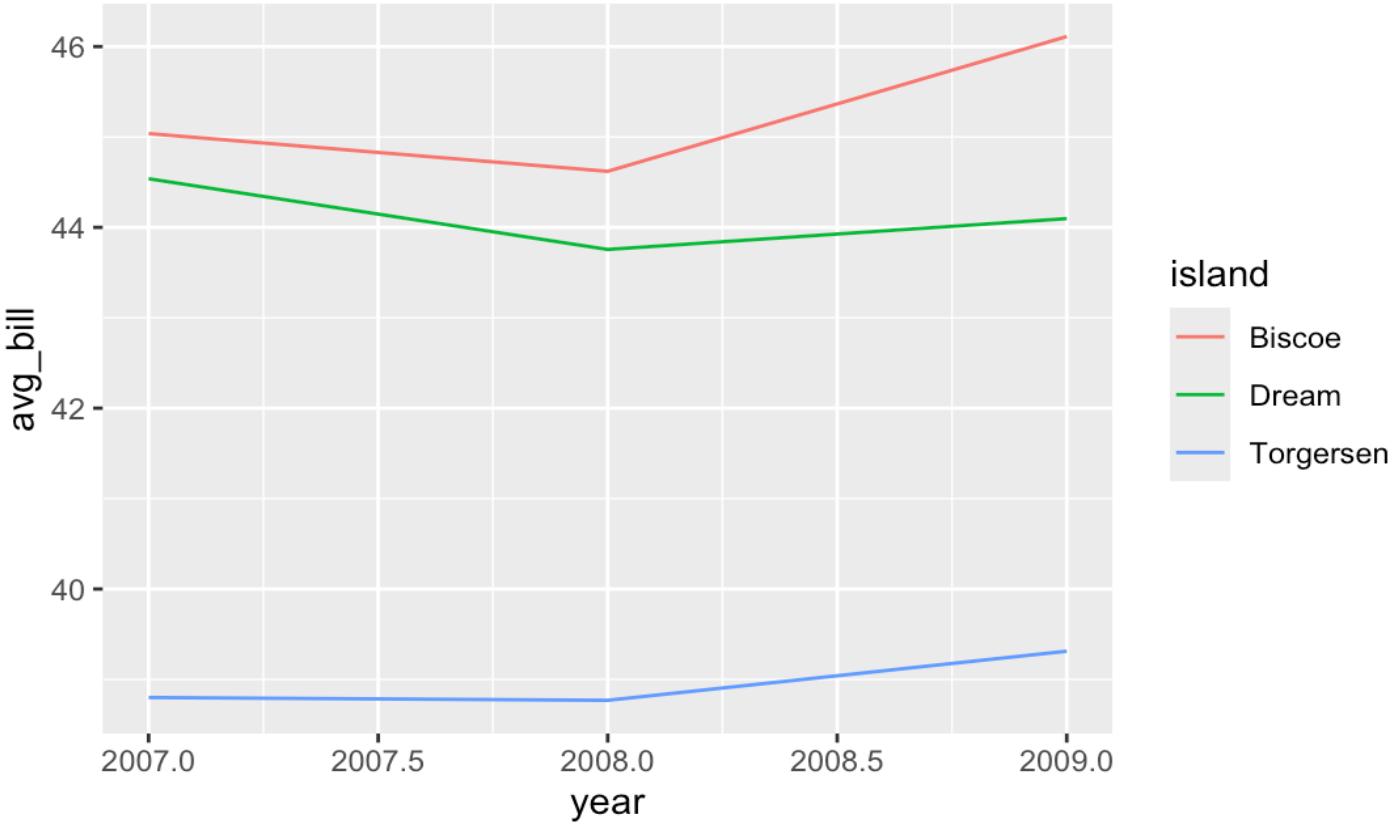


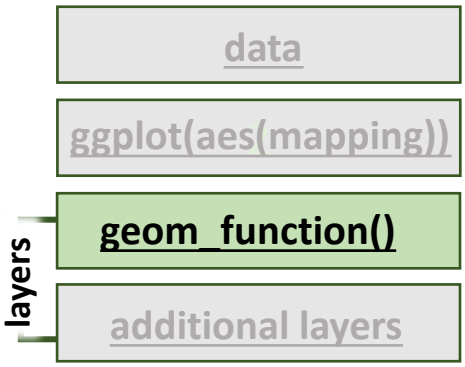
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_line()

```
penguins_year_species |>
  ggplot(aes(year, avg_bill, group =
  island, color = island)) +
  geom_line()
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_boxplot()

- General research question
  - How is a continuous variable distributed across groups, and how do the medians, quartiles, and potential outliers compare?

layers

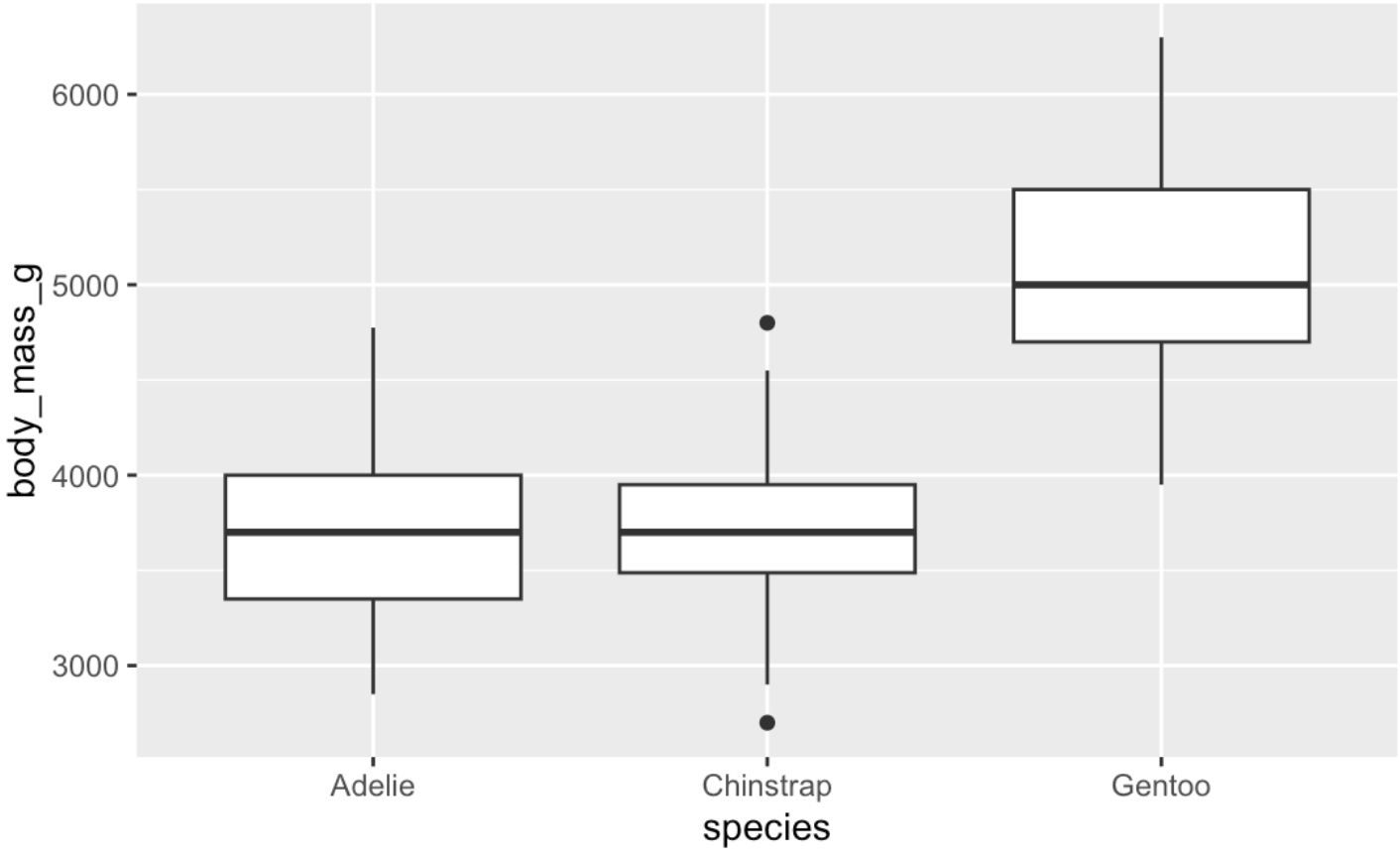
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

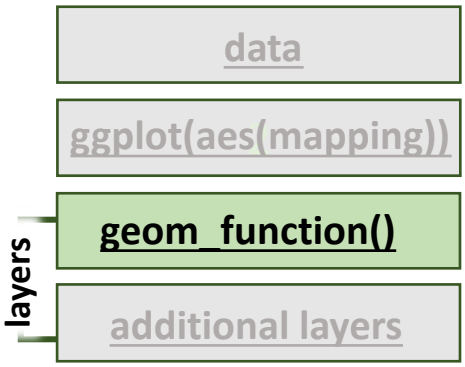
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_boxplot()

```
penguins |>
  ggplot(aes(species, body_mass_g)) +
  geom_boxplot()
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_violin()

- General research question
  - How is the full distribution of a continuous variable shaped across groups?

layers

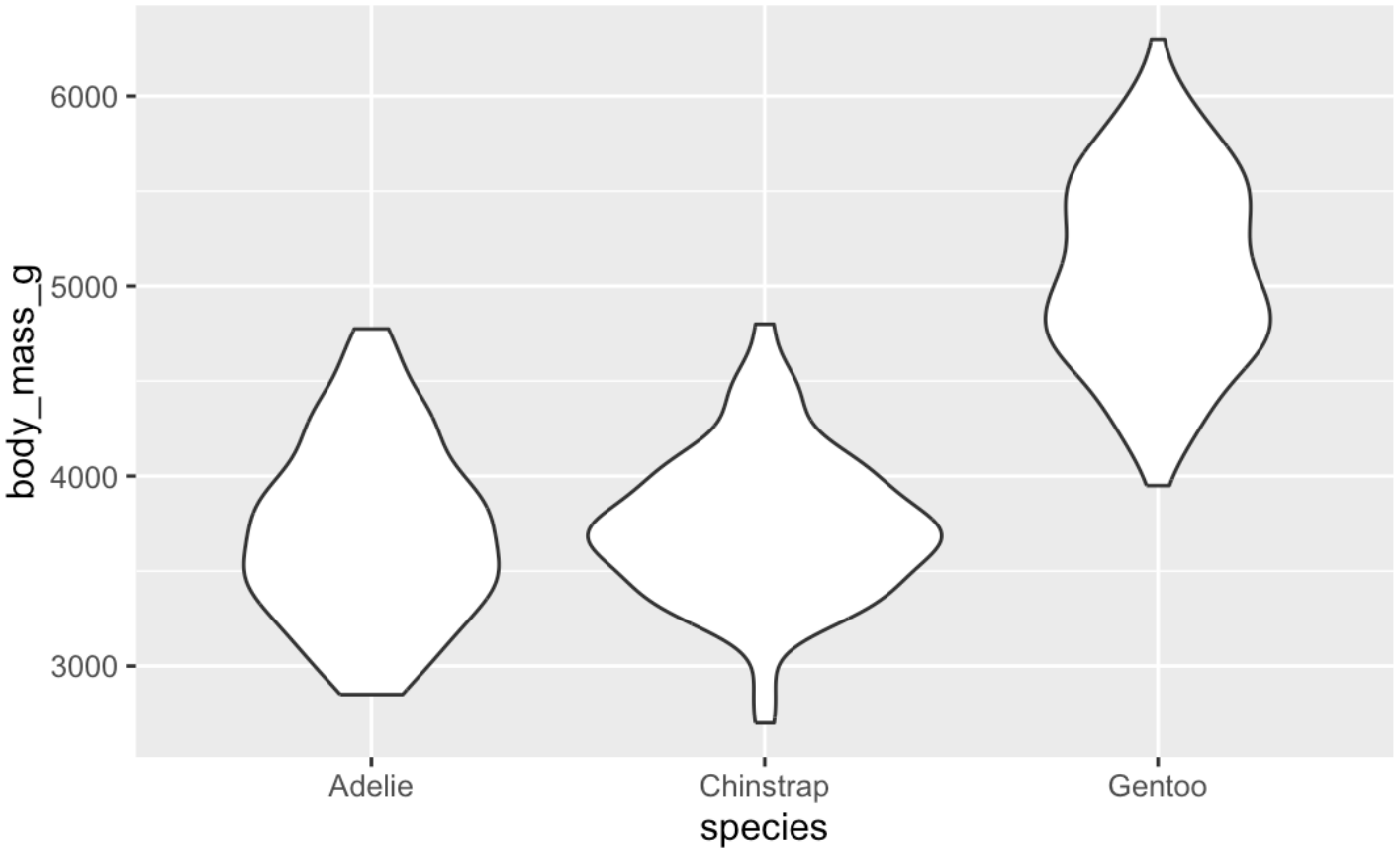
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

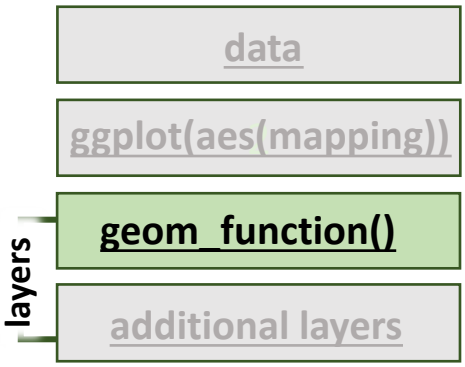
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_violin()

```
penguins |>
  ggplot(aes(species, body_mass_g)) +
  geom_violin()
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_bar() vs geom_col()

- **1 discrete variable**
- counts rows
- height of the bar is proportional to the number of cases in each group

- **2 variables**:
  - 1 continuous
  - 1 discrete (unique)
- need to have a variable with numbers in your data (average, proportion, count)

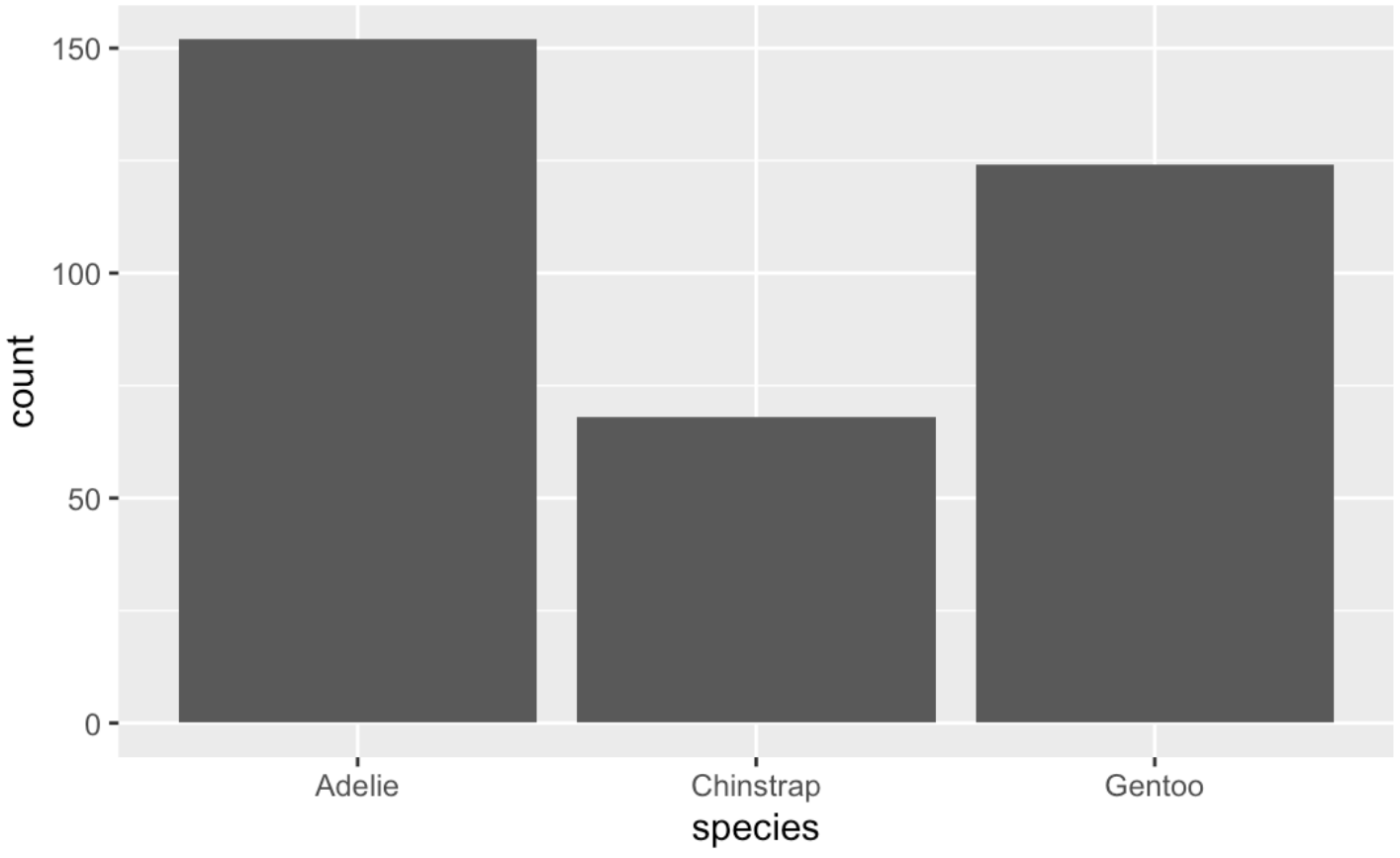
layers

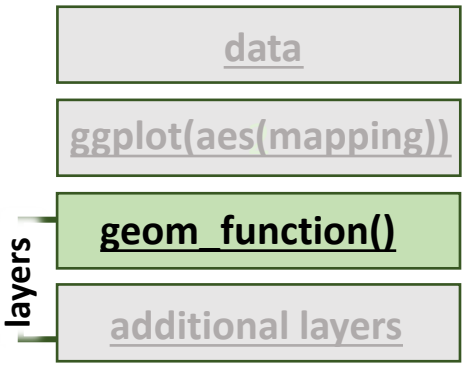
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

geom_bar()

```
penguins |>
  ggplot(aes(species)) +
  geom_bar()
```





		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_col(): data

```
summarized_penguins <- penguins |>
  group_by(species) |>
  summarize(N = n())

head(summarized_penguins)
```

A tibble: 3 × 2

species <fctr>	N <int>
Adelie	152
Chinstrap	68
Gentoo	124

3 rows

layers

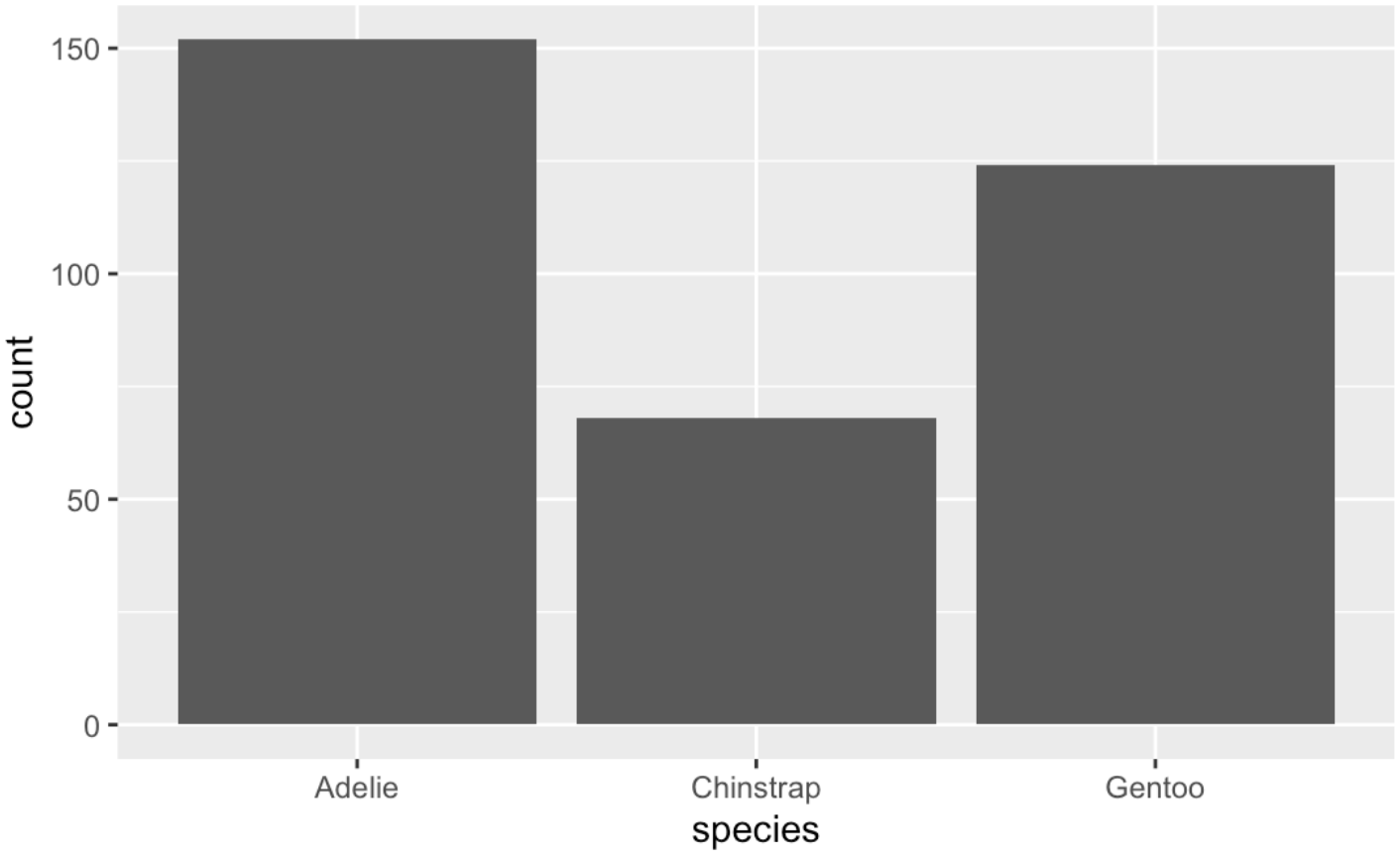
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

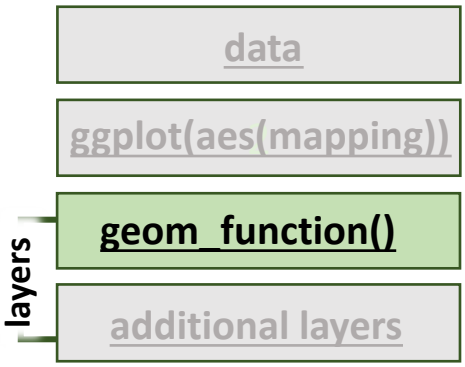
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_col()

```
summarized_penguins |>
  ggplot(aes(species,N)) +
  geom_col()
```





		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

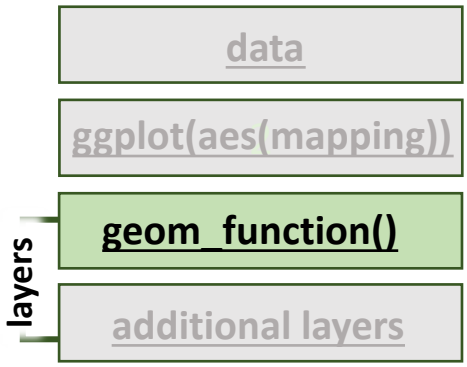
More aes mapping

```
summarized_penguins2 <- penguins |>
  group_by(species, sex) |>
  na.omit()
  summarize(bill_length_avg =
                      mean(bill_length_mm))
head(summarized_penguins2)
```

A tibble: 6 × 3 | Groups: species [3]

species <fctr>	sex <fctr>	bill_length_avg <dbl>
Adelie	female	37.25753
Adelie	male	40.39041
Chinstrap	female	46.57353
Chinstrap	male	51.09412
Gentoo	female	45.56379
Gentoo	male	49.47377

6 rows

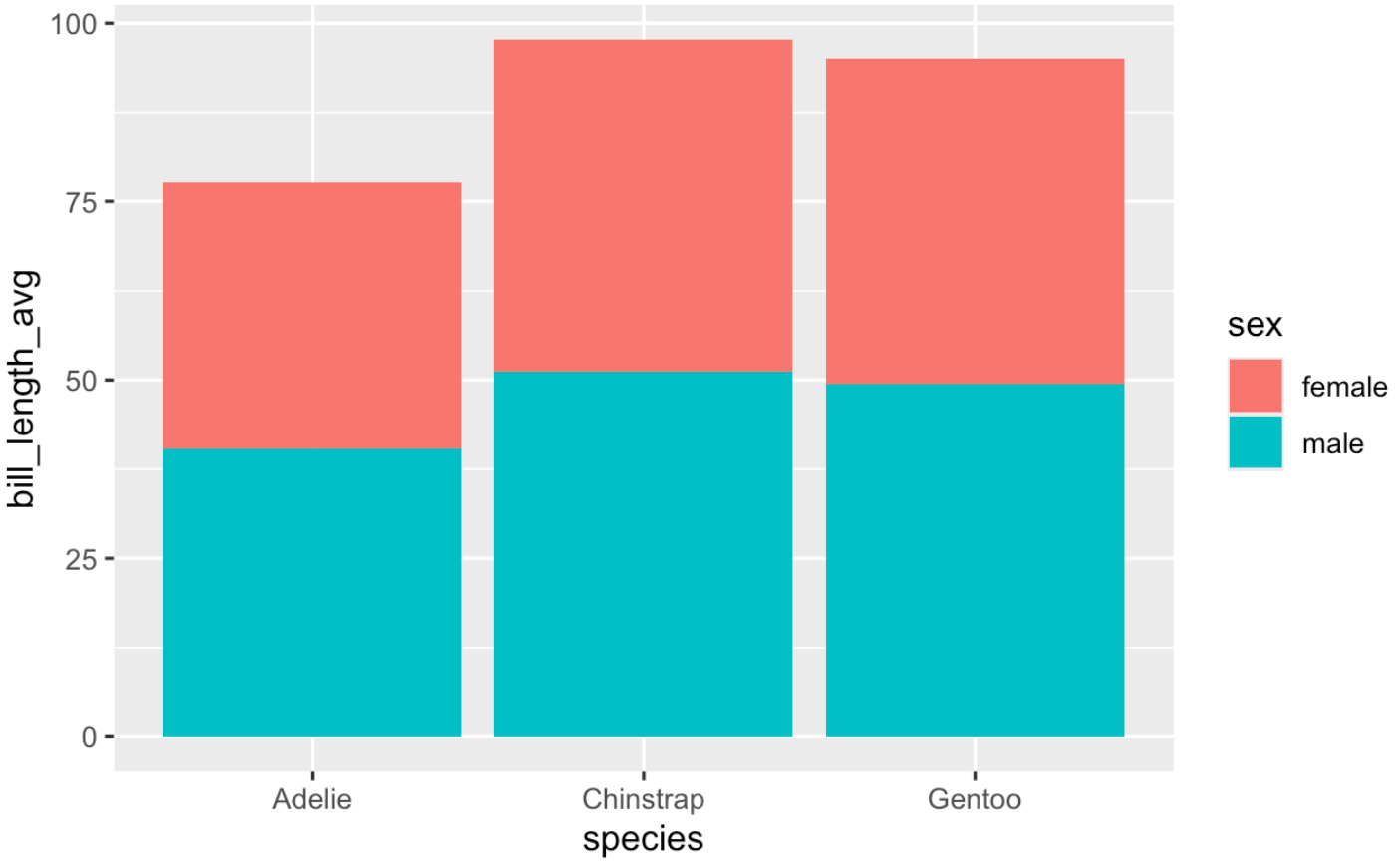


		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

More aes mapping

```
ggplot(summarized_penguins2, aes(species,
  bill_length_avg)) +
   geom_col(aes(fill = sex))
```



UNIVERSITY OF OREGON

layers

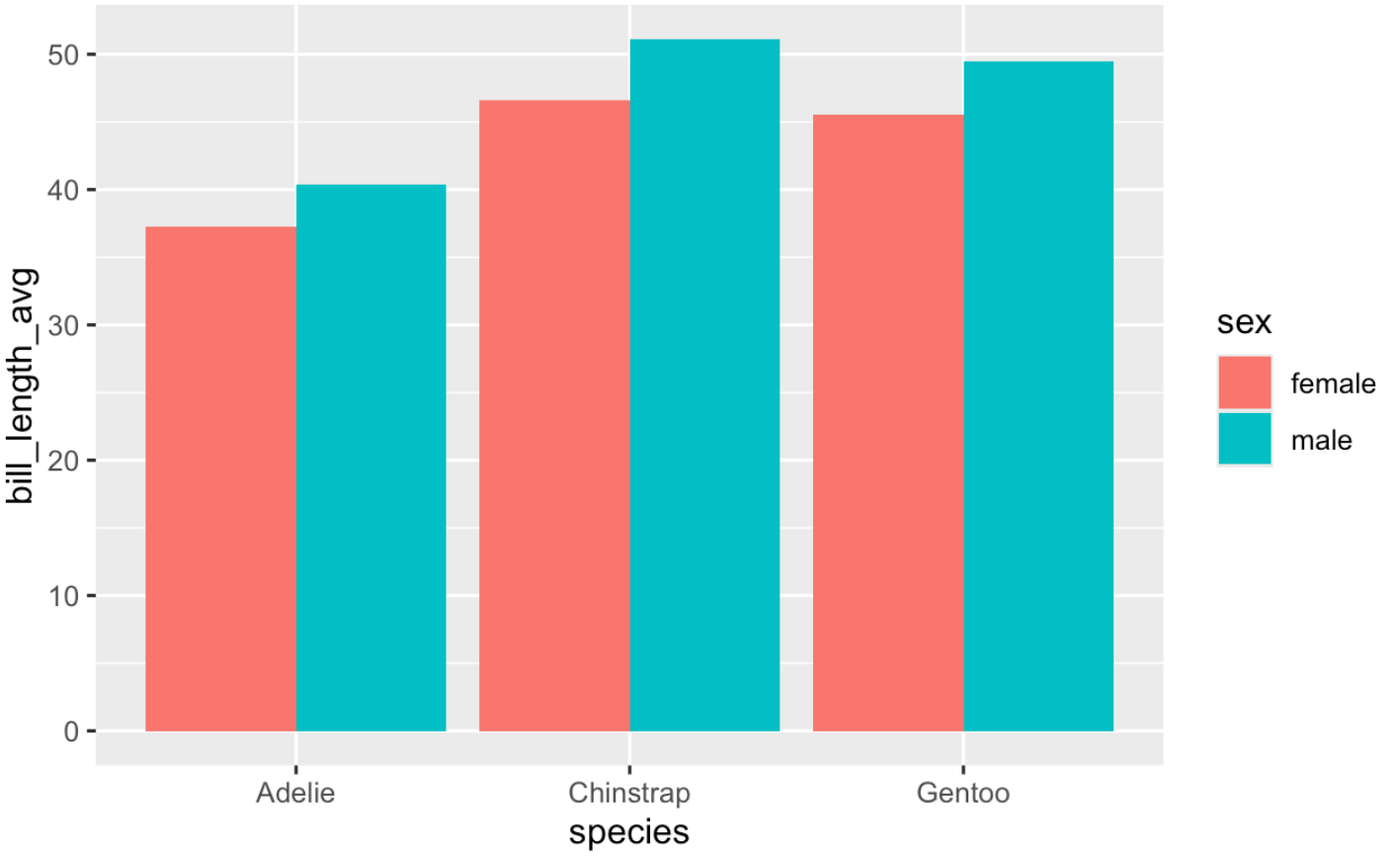
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Position

```
ggplot(summarized_penguins2, aes(species,
  bill_length_avg)) +
   geom_col(aes(fill = sex),
            position = "dodge")
```



UNIVERSITY OF OREGON

data

ggplot(aes(mapping))

layers

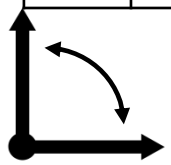
geom_function()

additional layers

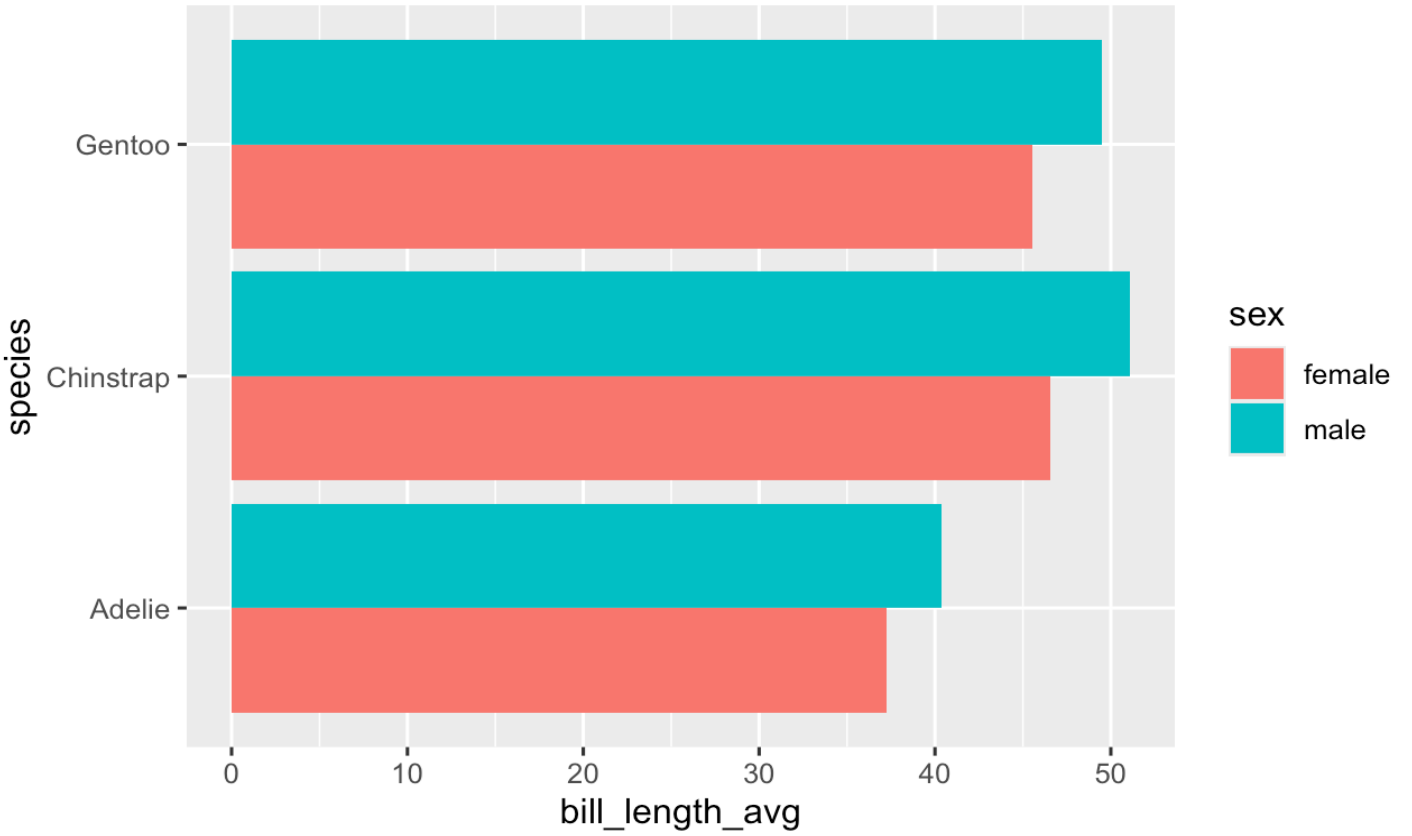
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

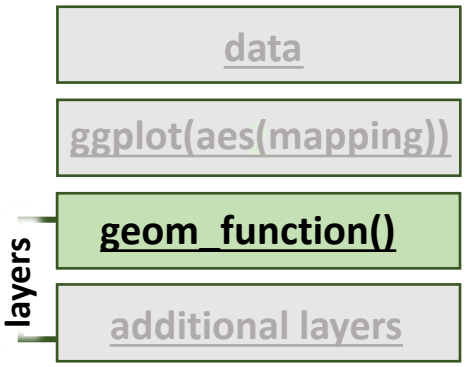
coord_flip



```
ggplot(summarized_penguins2, aes(species,
  bill_length_avg)) +
   geom_col(aes(fill = sex), position =
   "dodge")+
coord_flip()
```



UNIVERSITY OF OREGON

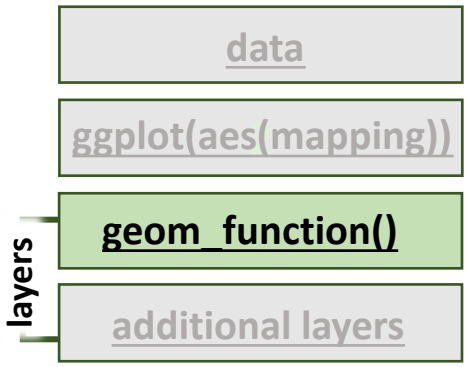


		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges (from {ggridges}) geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_count()

- General research question
  - How many observations fall in each category pair?
- Our data specific question
  - How many of each species live in each island?

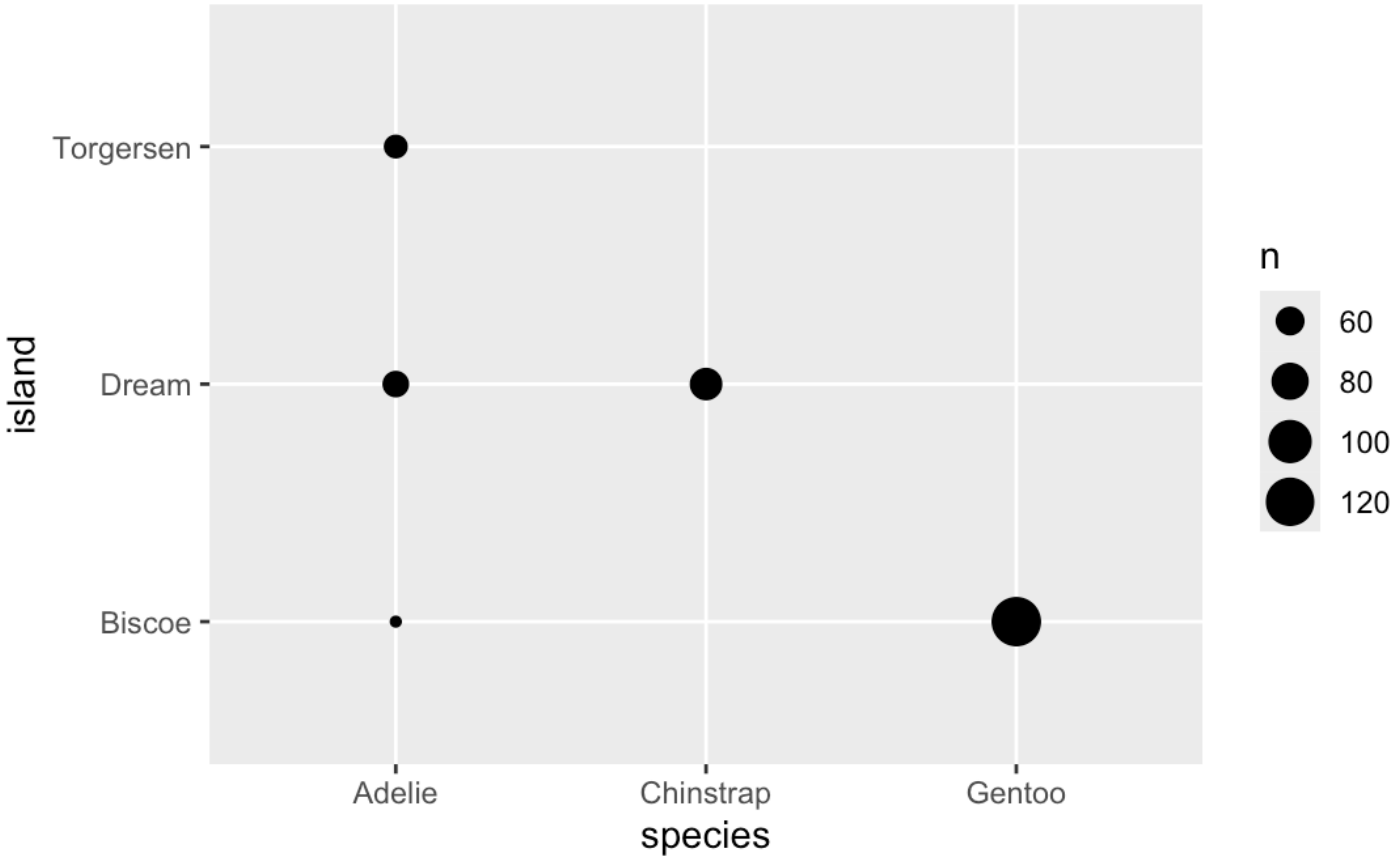


		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_count()

```
penguins |>
  ggplot(aes(species, island)) +
  geom_count()
```



How might we increase the contrast to enhance clarity?

UNIVERSITY OF OREGON

layers

- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

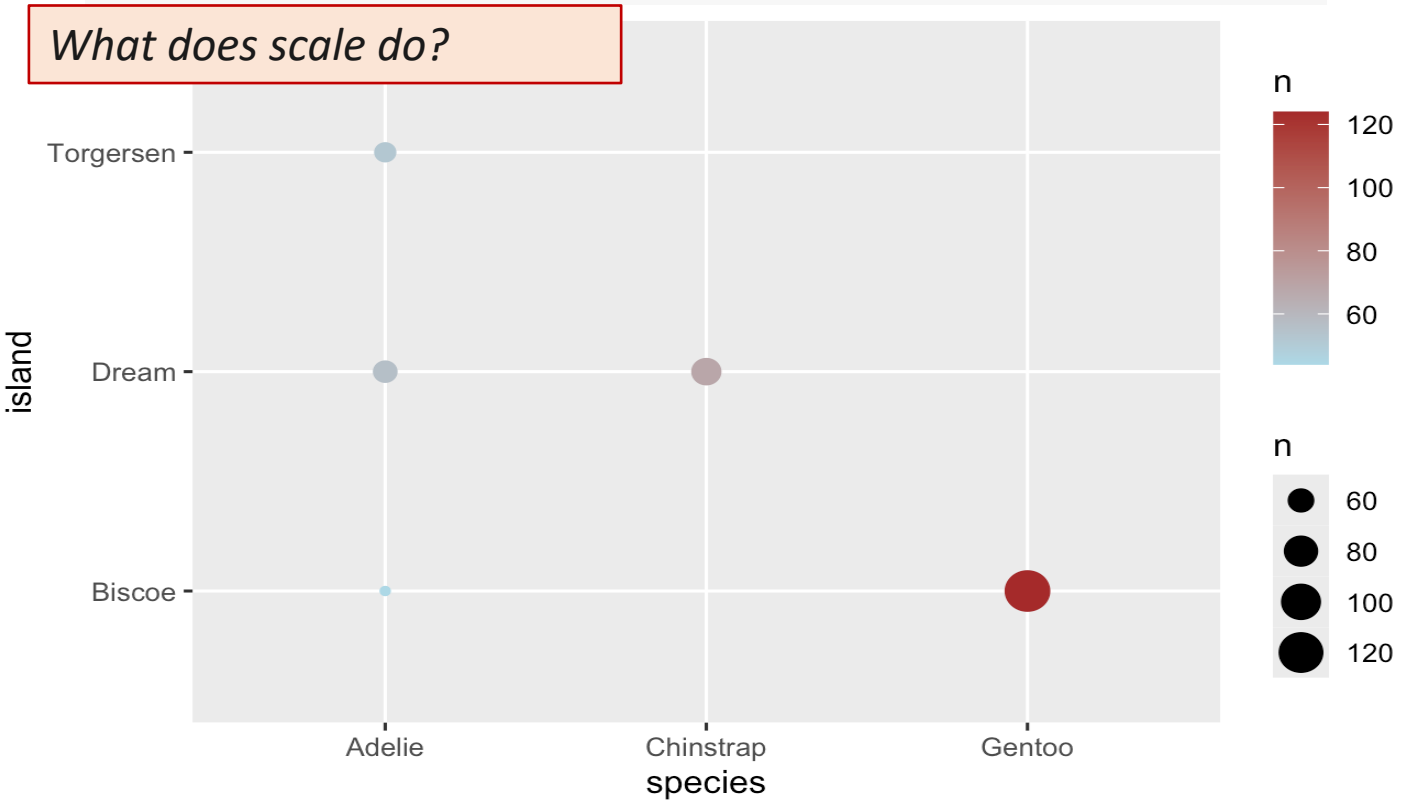
		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

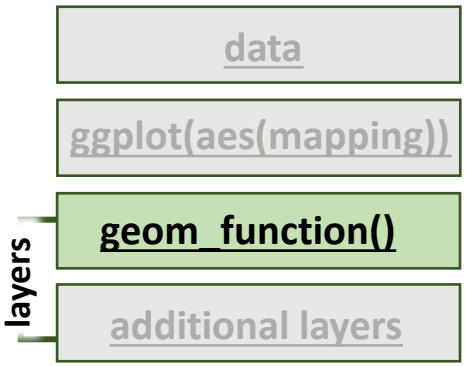
Scales

```
penguins |>
  ggplot(aes(x = species, y = island)) +
  geom_count(aes(color = after_stat(n)))+
  scale_color_gradient(low = "lightblue",
                       high = "brown")
```

What does scale do?



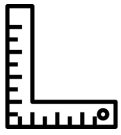
UNIVERSITY OF OREGON



		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Scales



- What do scales do?
  - Scales control how the mappings you added to aes are displayed (e.g., color range, size range, breaks and labels, range or limits)
  - Template: scale_*
    - Most aes mappings: x, y, size, color, fill, line, alpha, etc.

layers

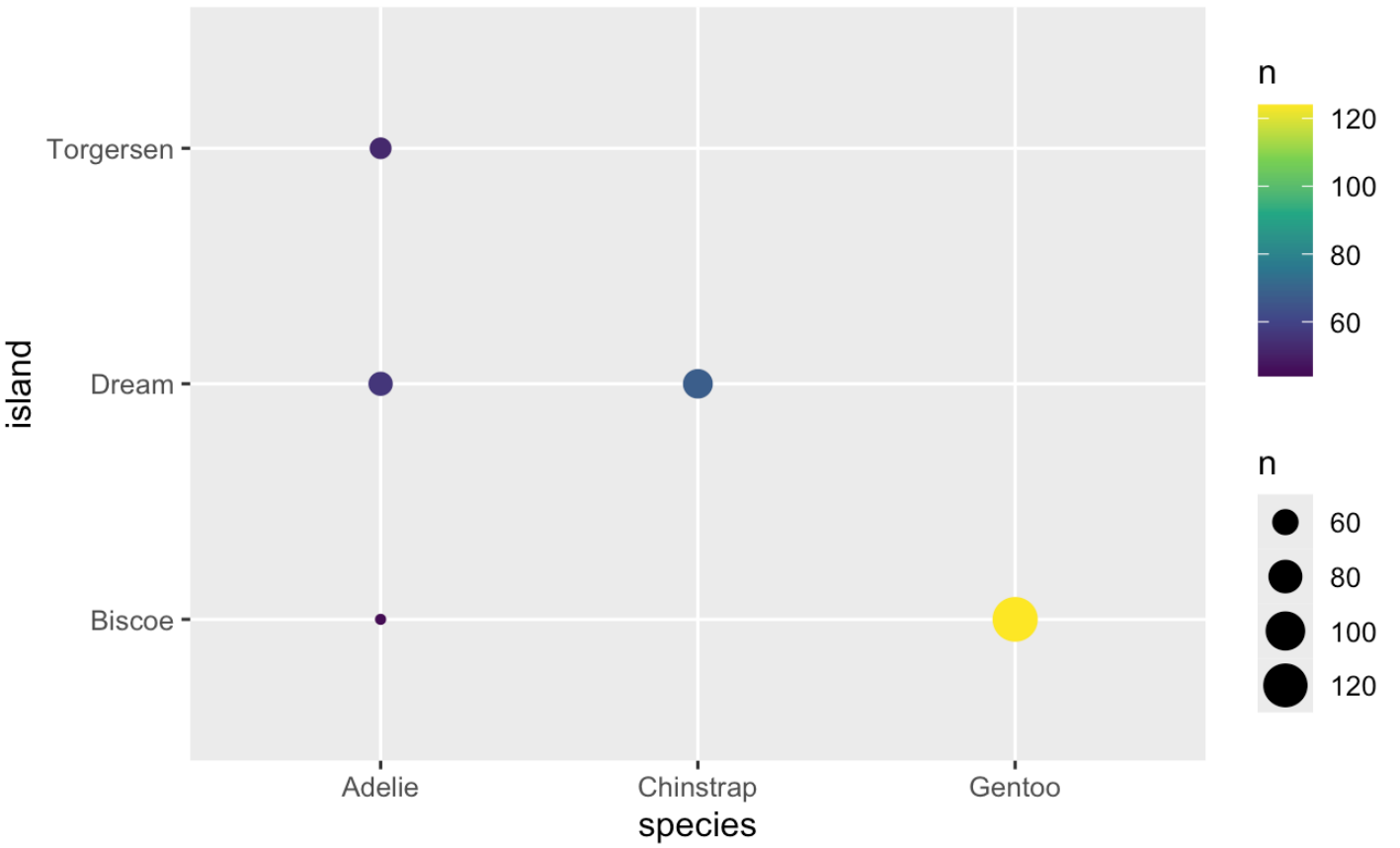
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2: Continuous	Variable 2: Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Scales

```
penguins |>
  ggplot(aes(x = species, y = island)) +
  geom_count(aes(color = after_stat(n)))+
  scale_color_viridis_c()
```



data

ggplot(aes(mapping))

geom_function()

additional layers

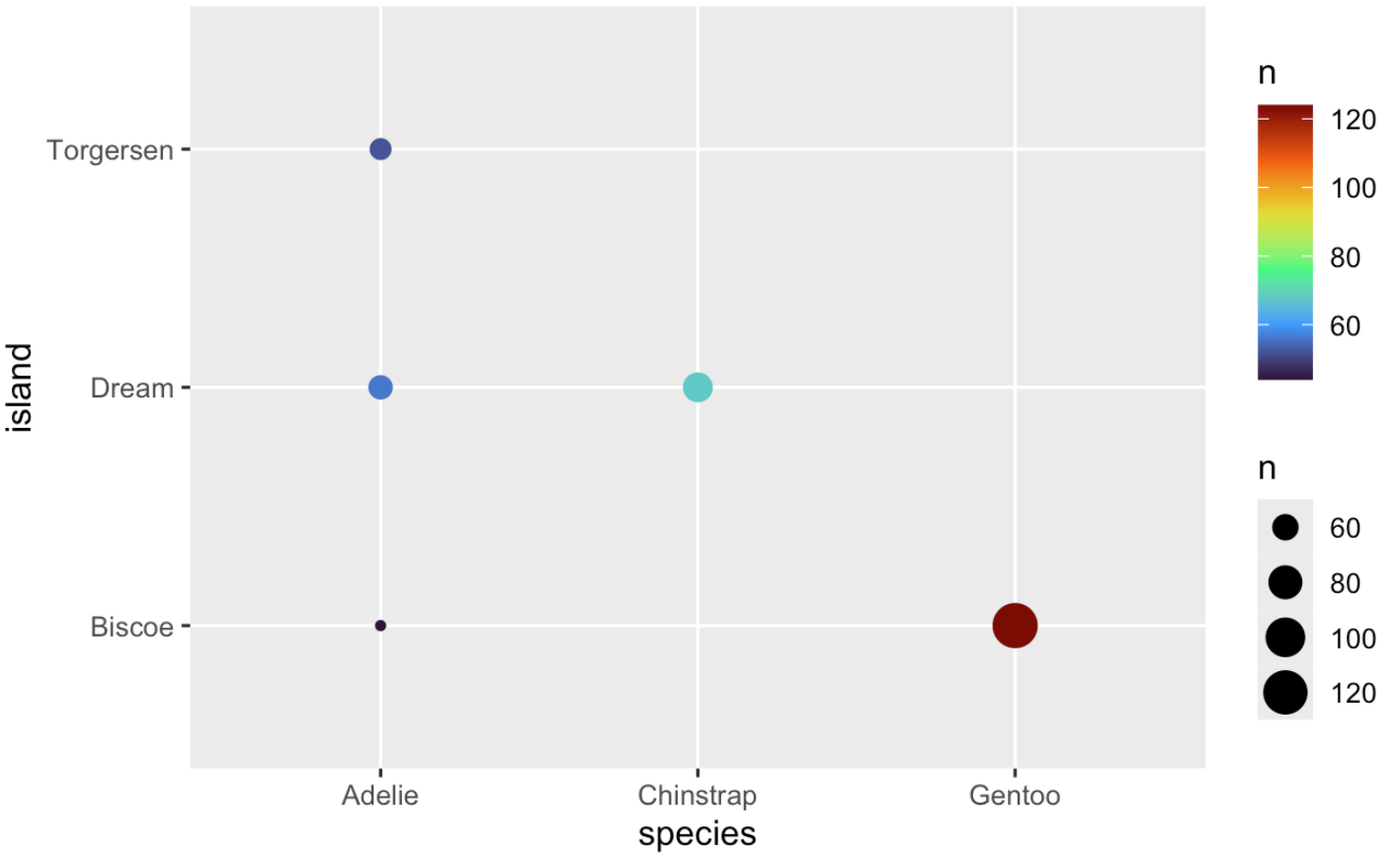
layers

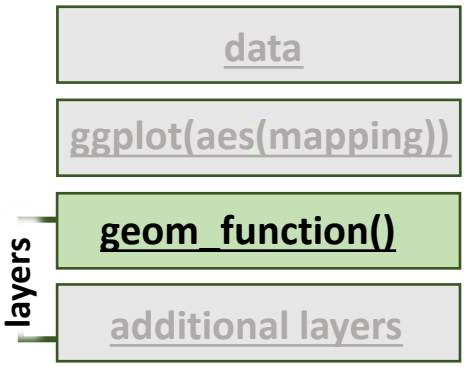
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

Scales

```
penguins |>
  ggplot(aes(x = species, y = island)) +
  geom_count(aes(color = after_stat(n)))+
  scale_color_viridis_c(option = "turbo")
```





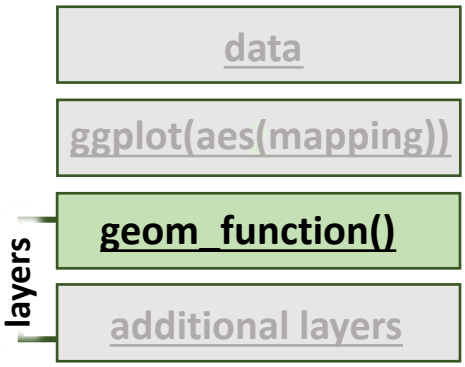
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_tile()

General research question:

- What's the value of a numerical measure (Z) for each (X, Y) pair?
  - What is the correlation (Z) for each X,Y pair?



		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_tile()

```
corr <- penguins |>
        select(bill_length_mm,
               bill_depth_mm,
               flipper_length_mm,
               body_mass_g) |>
  drop_na() |> cor()
pc <- corr |>
      as.data.frame() |>
      mutate(row = rownames(.)) |>
      pivot_longer(cols = -row,
                   names_to = "col",
                   values_to = "cor")
head(pc)
```

A tibble: 6 × 3

row <chr>	col <chr>	cor <dbl>
bill_length_mm	bill_length_mm	1.0000000
bill_length_mm	bill_depth_mm	-0.2350529
bill_length_mm	flipper_length_mm	0.6561813
bill_length_mm	body_mass_g	0.5951098
bill_depth_mm	bill_length_mm	-0.2350529
bill_depth_mm	bill_depth_mm	1.0000000

6 rows

layers

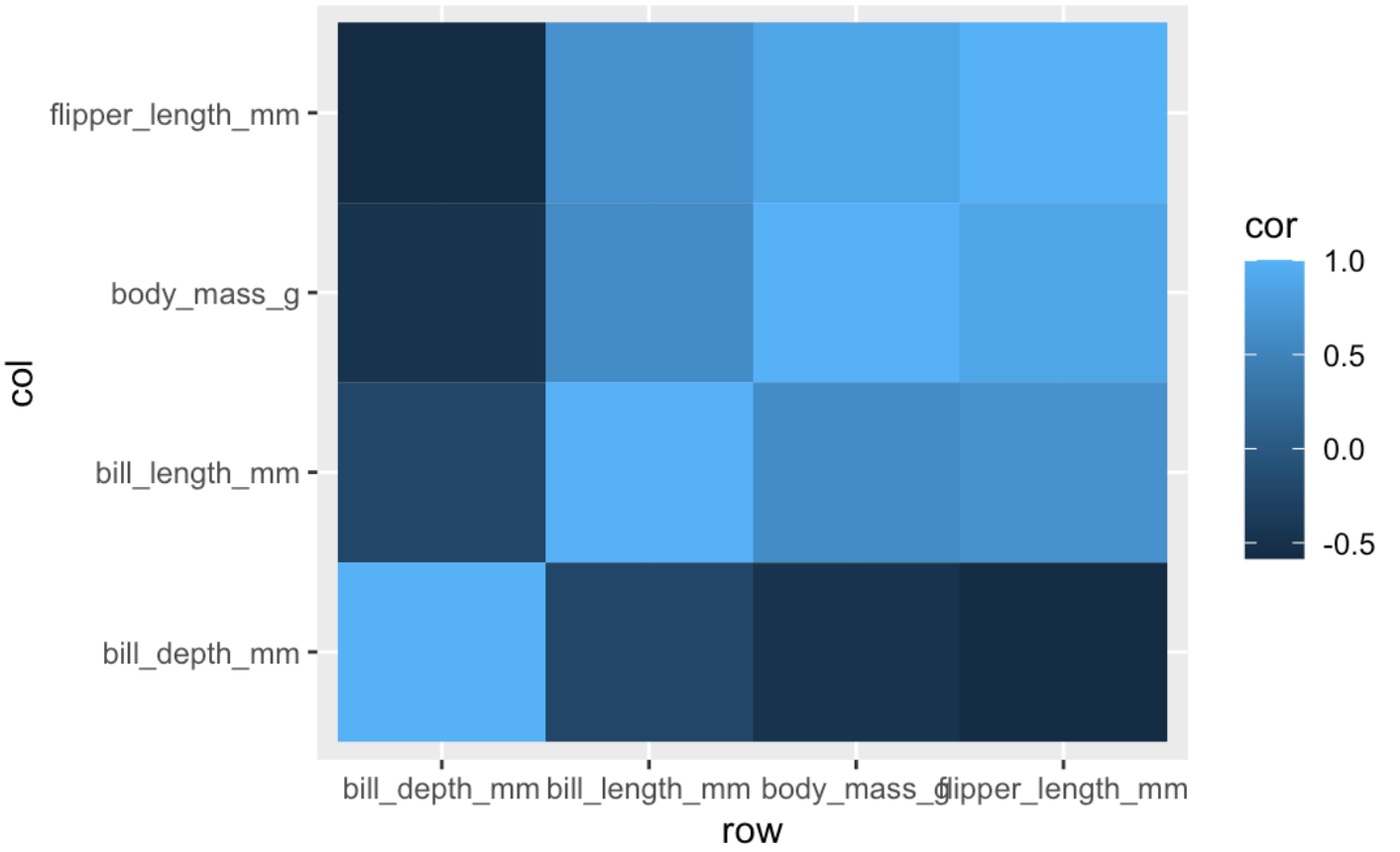
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_tile()

```
ggplot(pc, aes(row, col, fill = cor)) +
  geom_tile()
```



layers

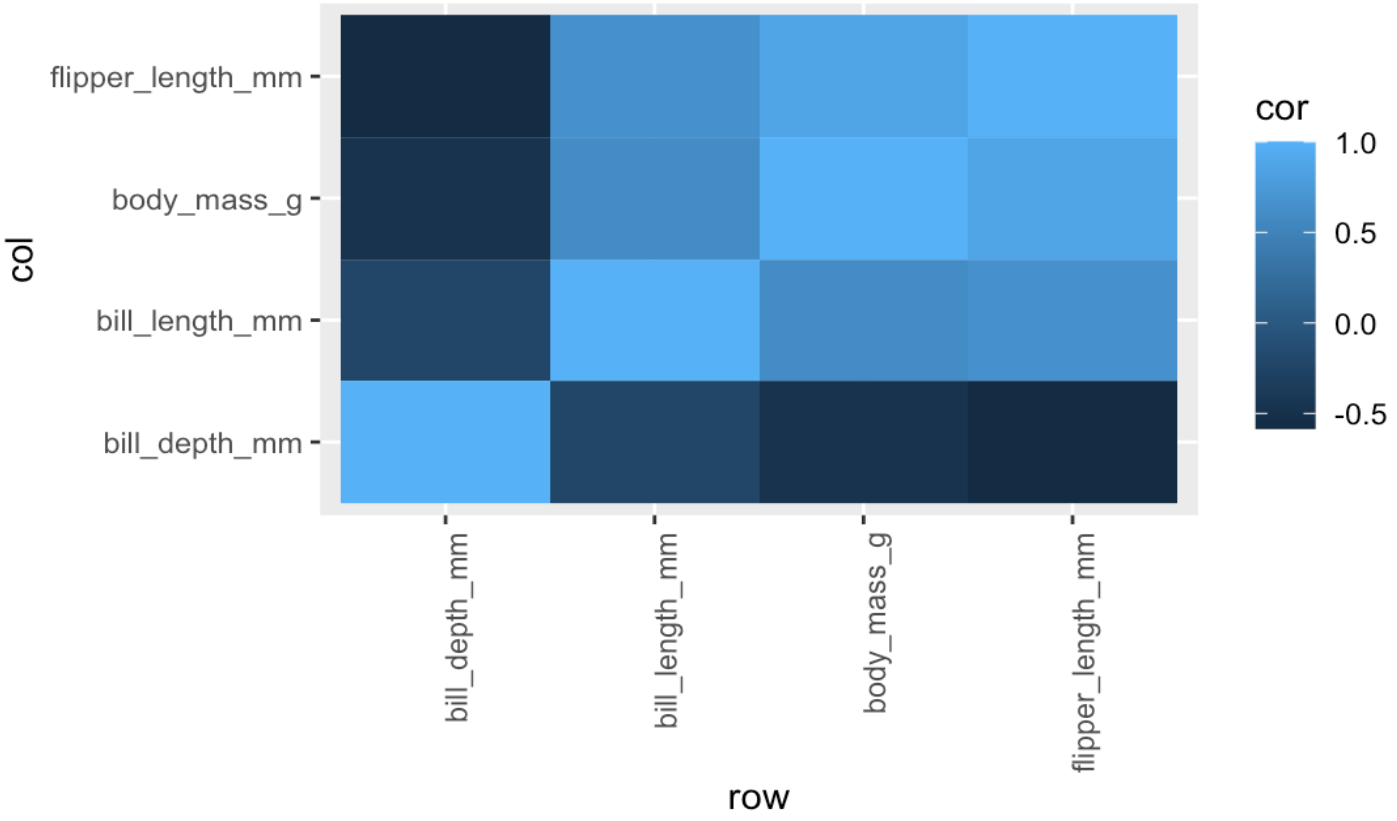
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_tile()

```
ggplot(pc, aes(row, col, fill = cor)) +
  geom_tile() +
  theme(axis.text.x = element_text(angle=90,
                                   hjust=1))
```



layers

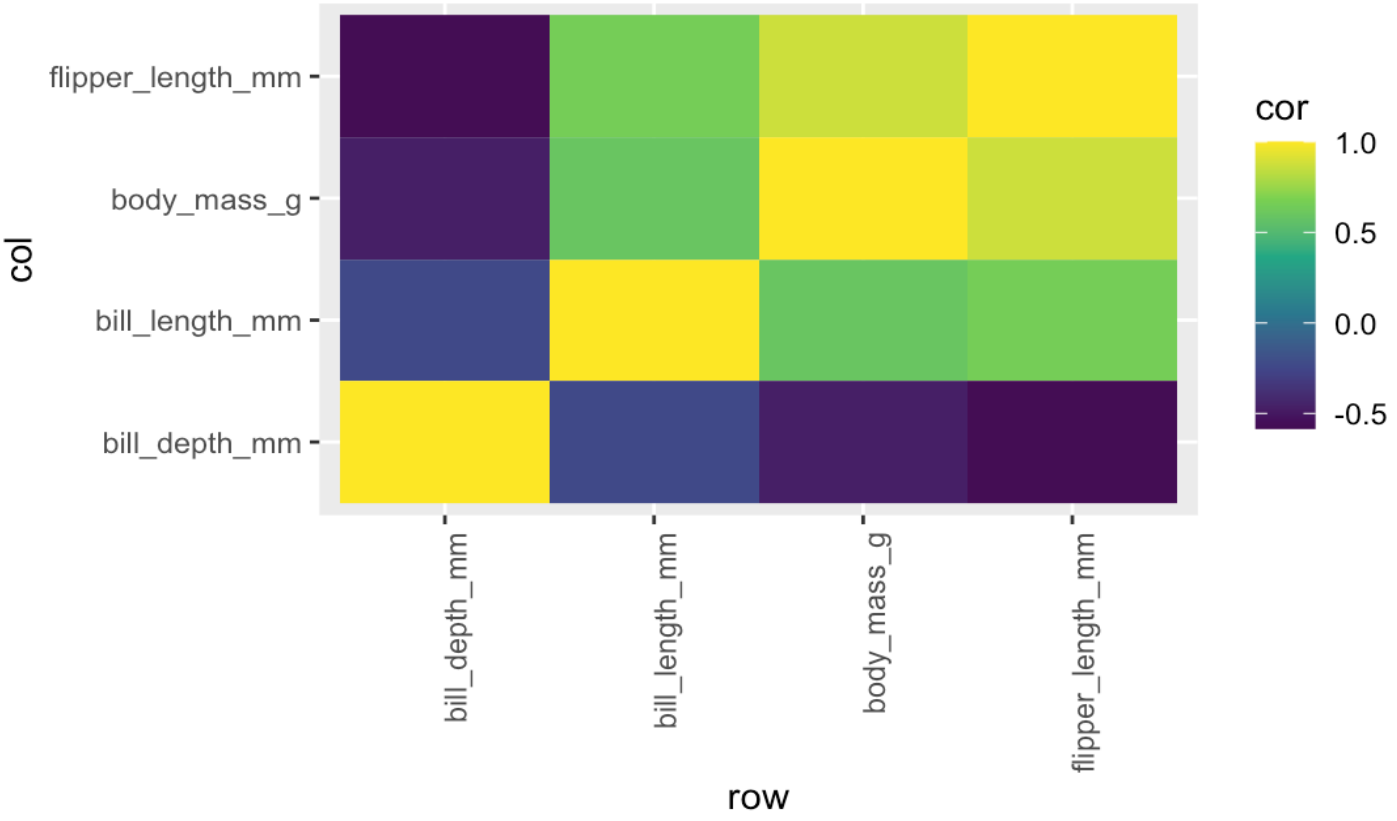
- data
- ggplot(aes(mapping))
- **geom_function()**
- additional layers

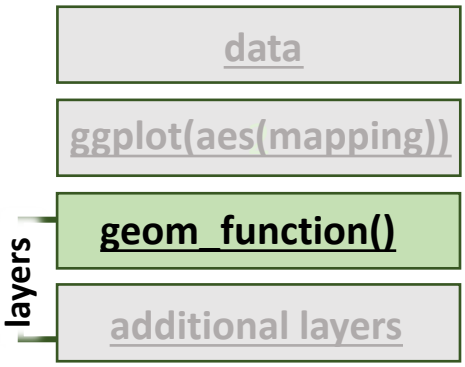
		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1	Discrete	geom_bar	X	geom_count

Other - Heatmaps: geom_tile

geom_tile()

```
ggplot(pc, aes(row, col, fill = cor)) +
  geom_tile() +
  scale_fill_viridis_c() +
  theme(axis.text.x = element_text(angle=90,
                                   hjust=1))
```





geom_function()

		Only 1 Variable	Variable 2	
			Continuous	Discrete
Variable 1	Continuous	geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
	Discrete	geom_bar	x	geom_count

Heatmaps: geom_tile

UNIVERSITY OF OREGON

data

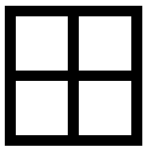
ggplot(aes(mapping))

layers

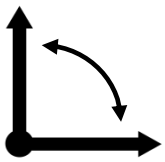
geom_function()

additional layers

additional layers



Facet wrap



Coord flip

Scales

Labels

Themes

data

ggplot(aes(mapping))

layers

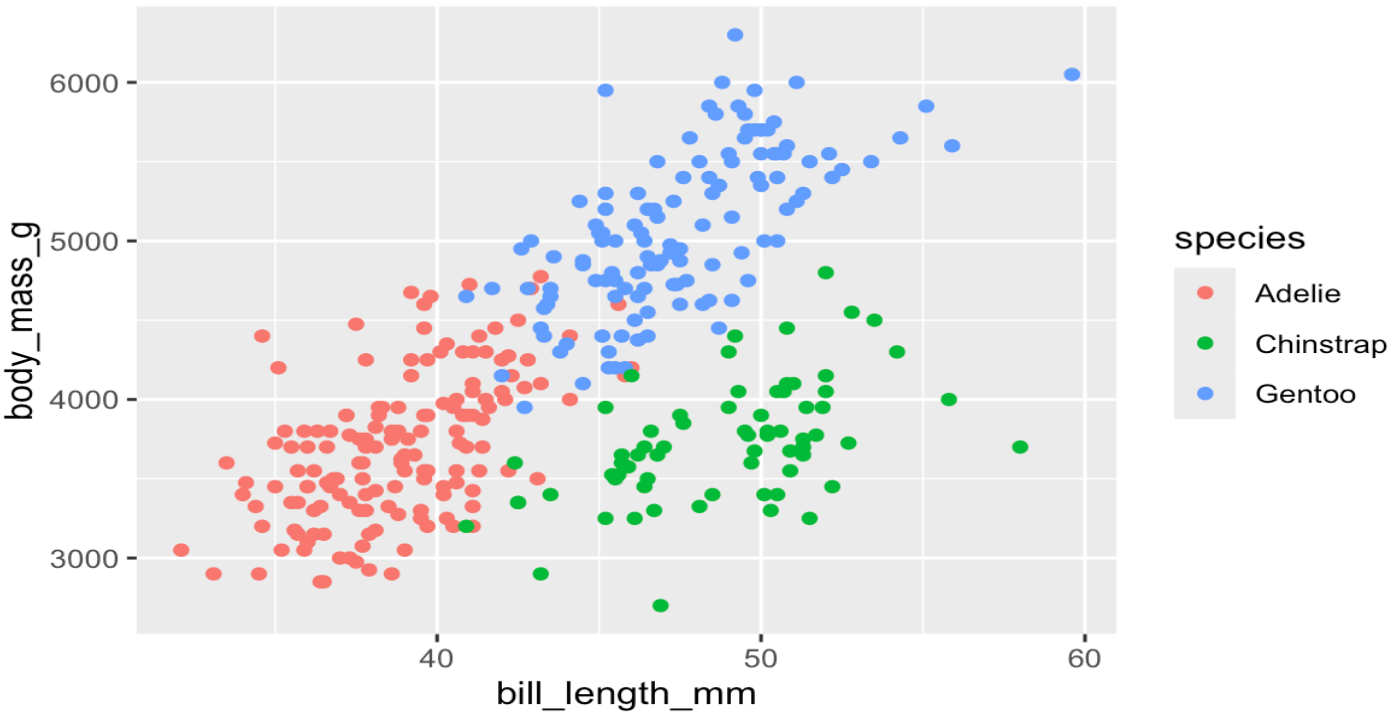
geom_function()

additional layers

Labels

Axis Labels

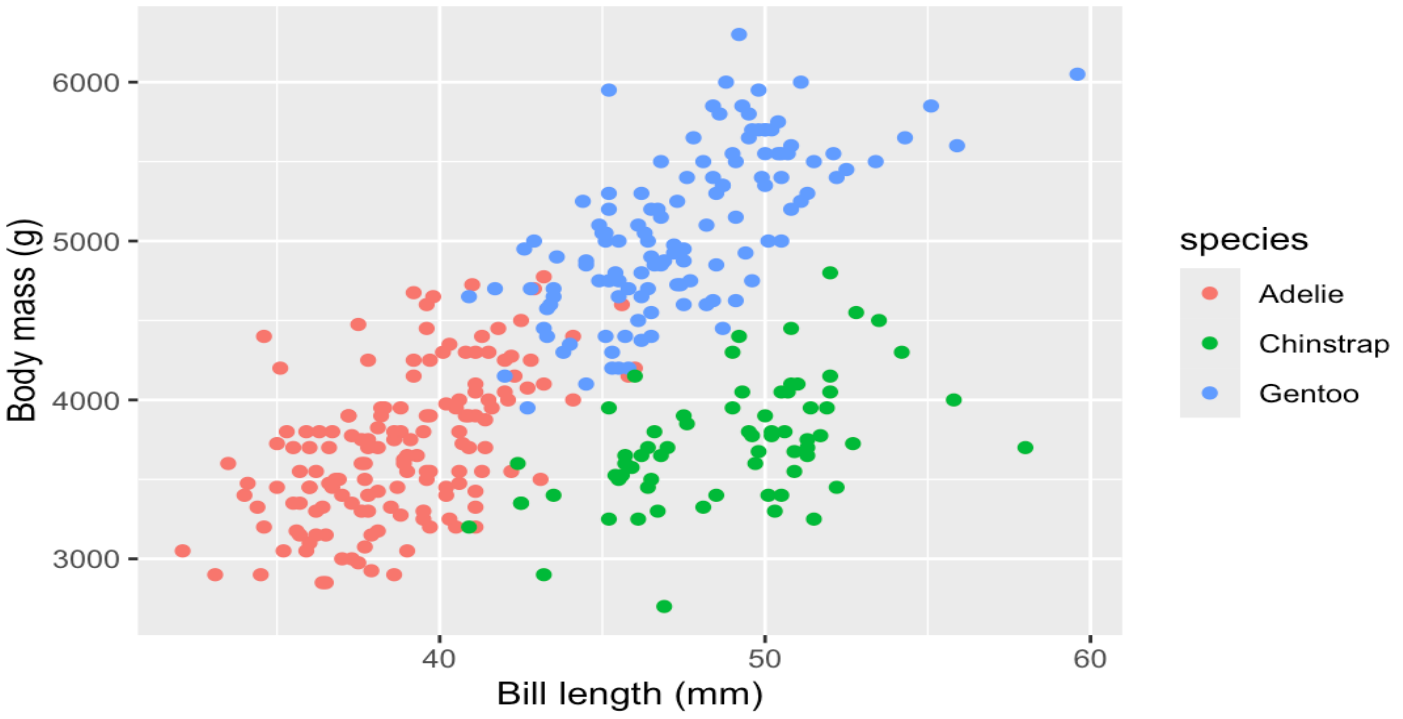
```
penguins |>
  ggplot(aes(bill_length_mm,
  body_mass_g)) +
  geom_point(aes(color = species))
```



Labels

Axis Labels

```
penguins |>
  ggplot(aes(bill_length_mm,
  body_mass_g)) +
  geom_point(aes(color = species)) +
  labs(x="Bill length (mm)",
       y="Body mass (g)")
```



data

ggplot(aes(mapping))

layers

geom_function()

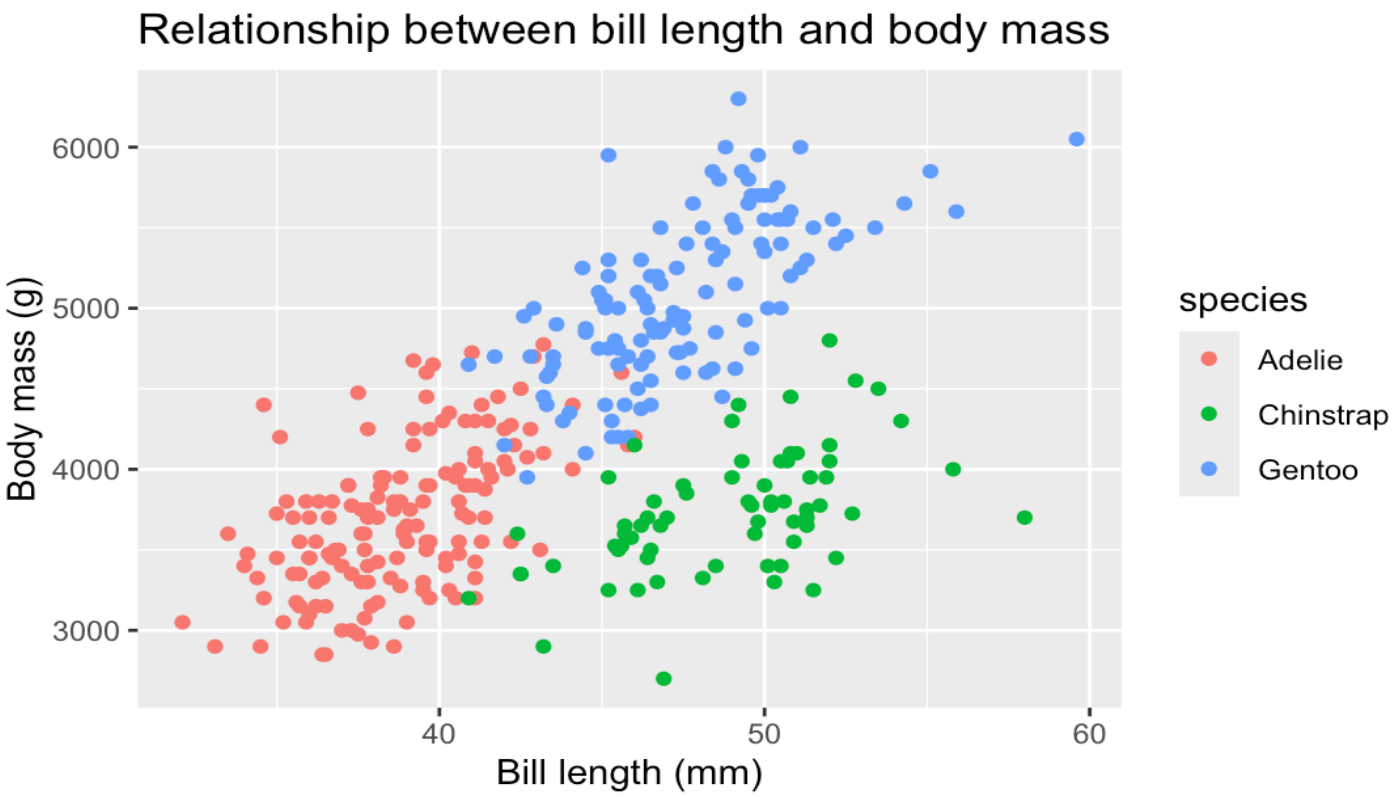
additional layers

Labels

Title

It is important to add “” around each labels

```
…
 labs(x="Bill length (mm)",
      y="Body mass (g)",
      title = "Relationship between bill
               length and body mass”)
```



data

ggplot(aes(mapping))

layers

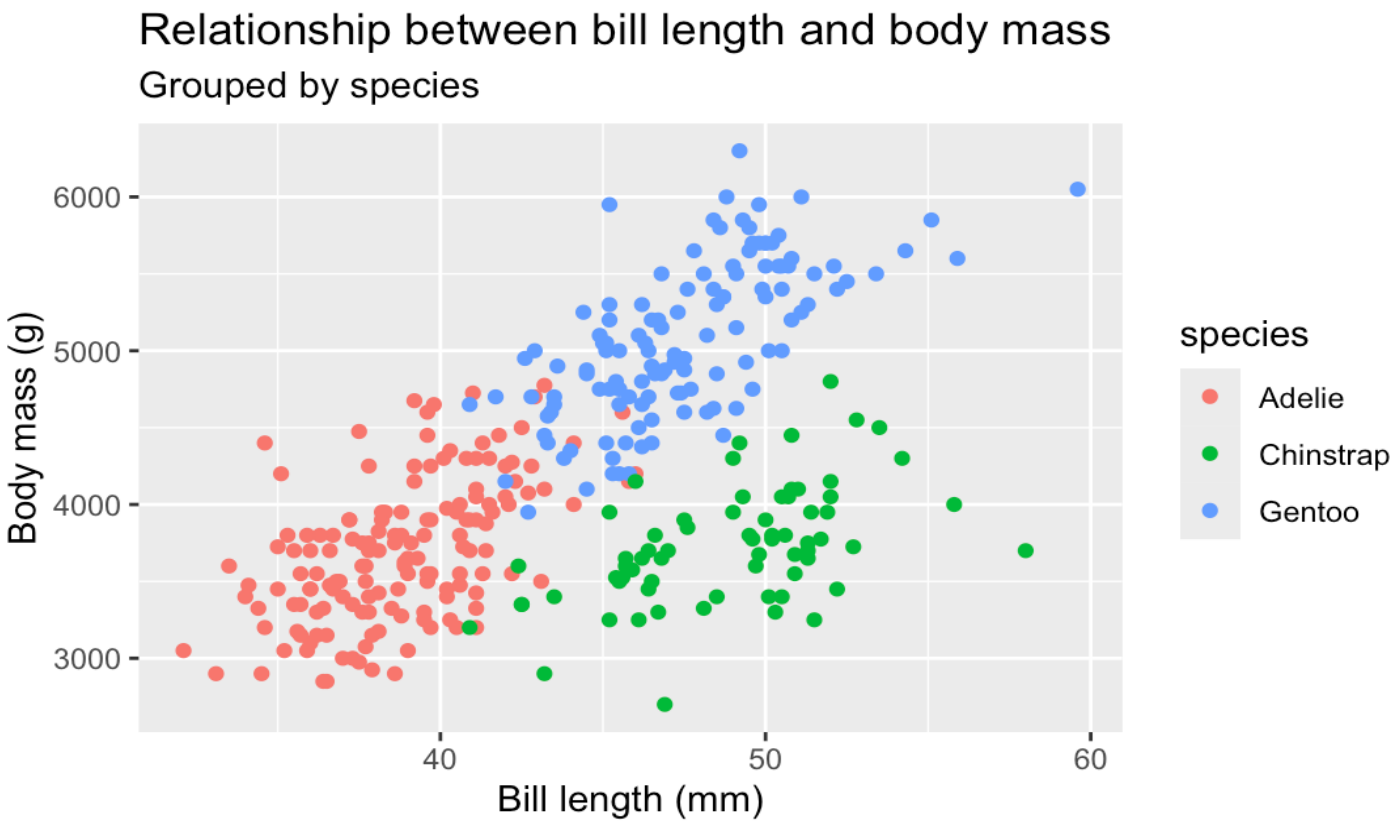
geom_function()

additional layers

Labels

Subtitle

```
…
 labs(x="Bill length (mm)",
       y="Body mass (g)",
       title = "Relationship between bill
                length and body mass”,
       subtitle=“Grouped by species”)
```



data

ggplot(aes(mapping))

layers

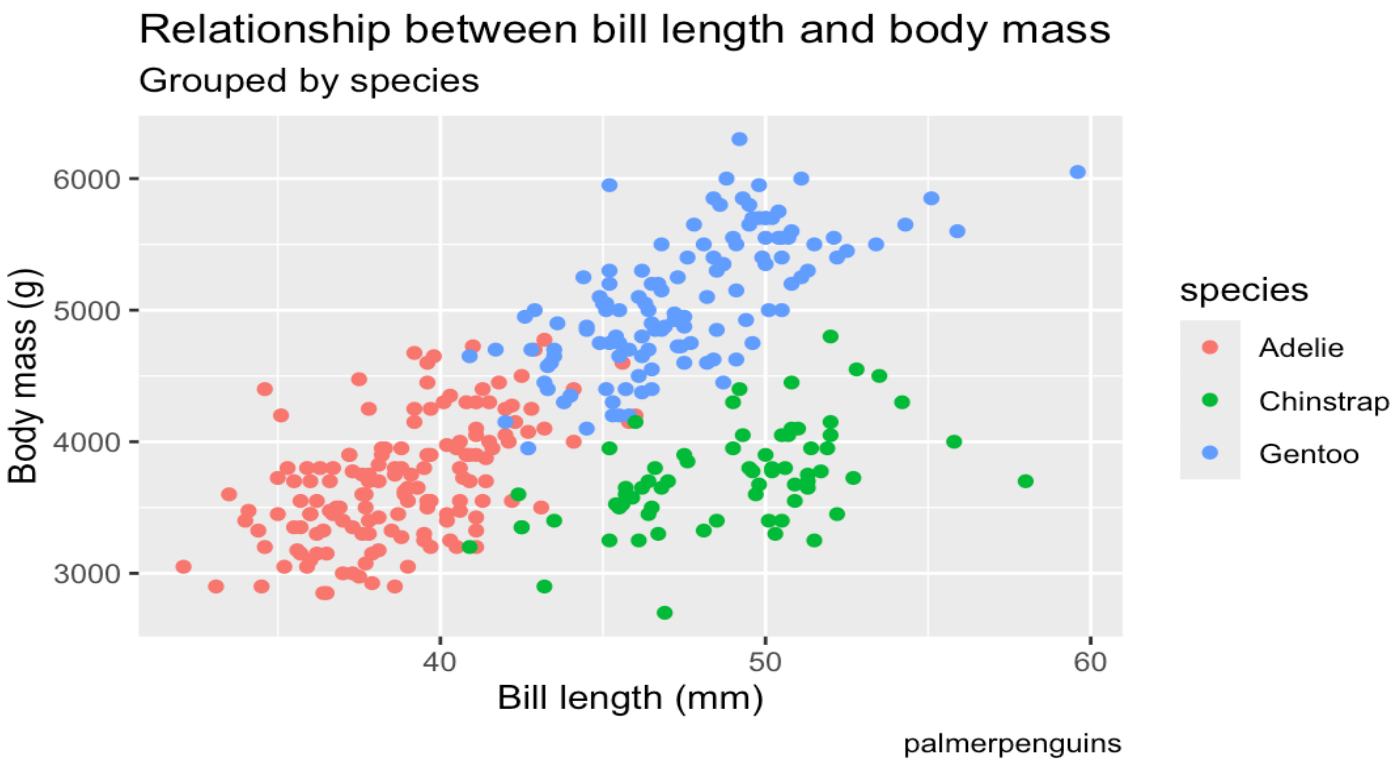
geom_function()

additional layers

Labels

Caption

```
…
 labs(x="Bill length (mm)",
       y="Body mass (g)",
       title = "Relationship between bill
                length and body mass",
       subtitle="Grouped by species",
       caption = "palmerpenguins")
```



data

ggplot(aes(mapping))

layers

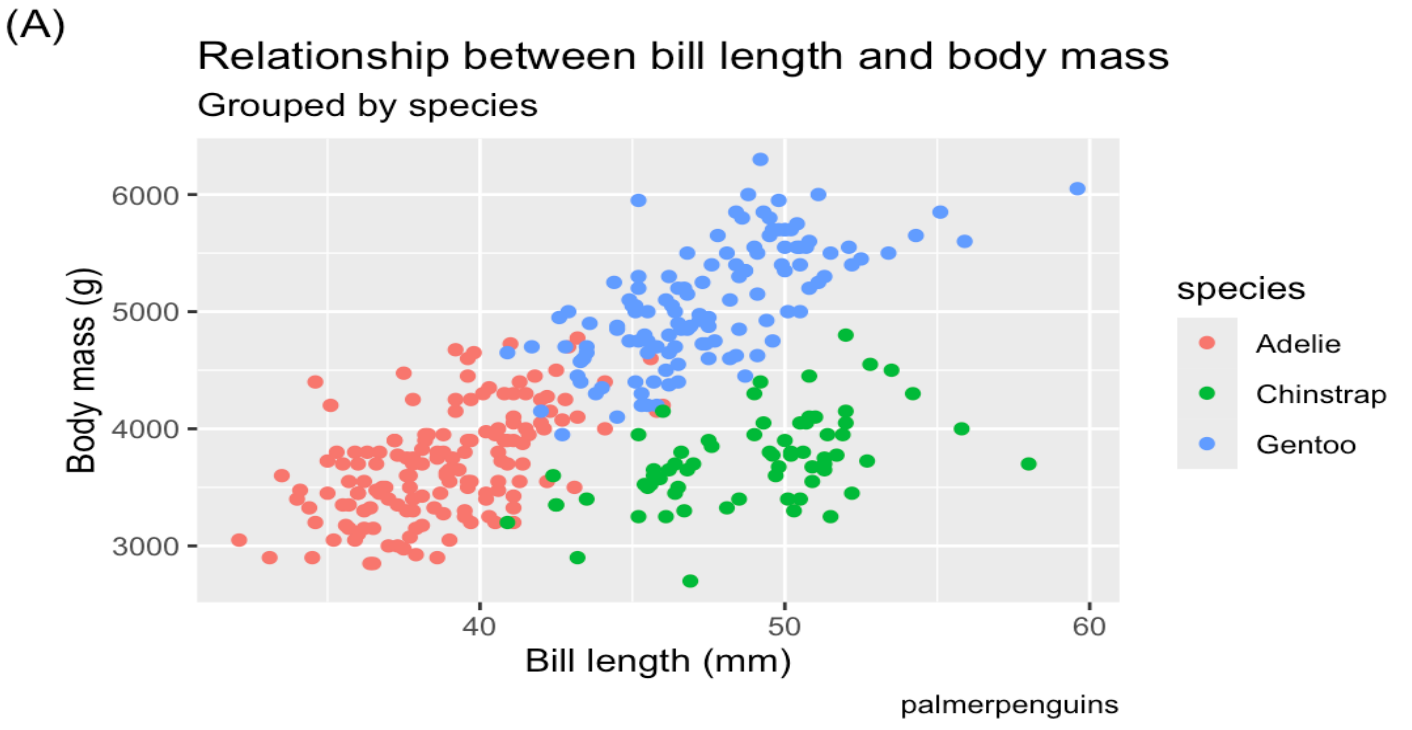
geom_function()

additional layers

Labels

Tag

```
…
 labs(x="Bill length (mm)",
       y="Body mass (g)",
       title = "Relationship between bill
                length and body mass",
       subtitle="Grouped by species",
       caption = "palmerpenguins",
       tag = "(A)")
```



data

ggplot(aes(mapping))

layers

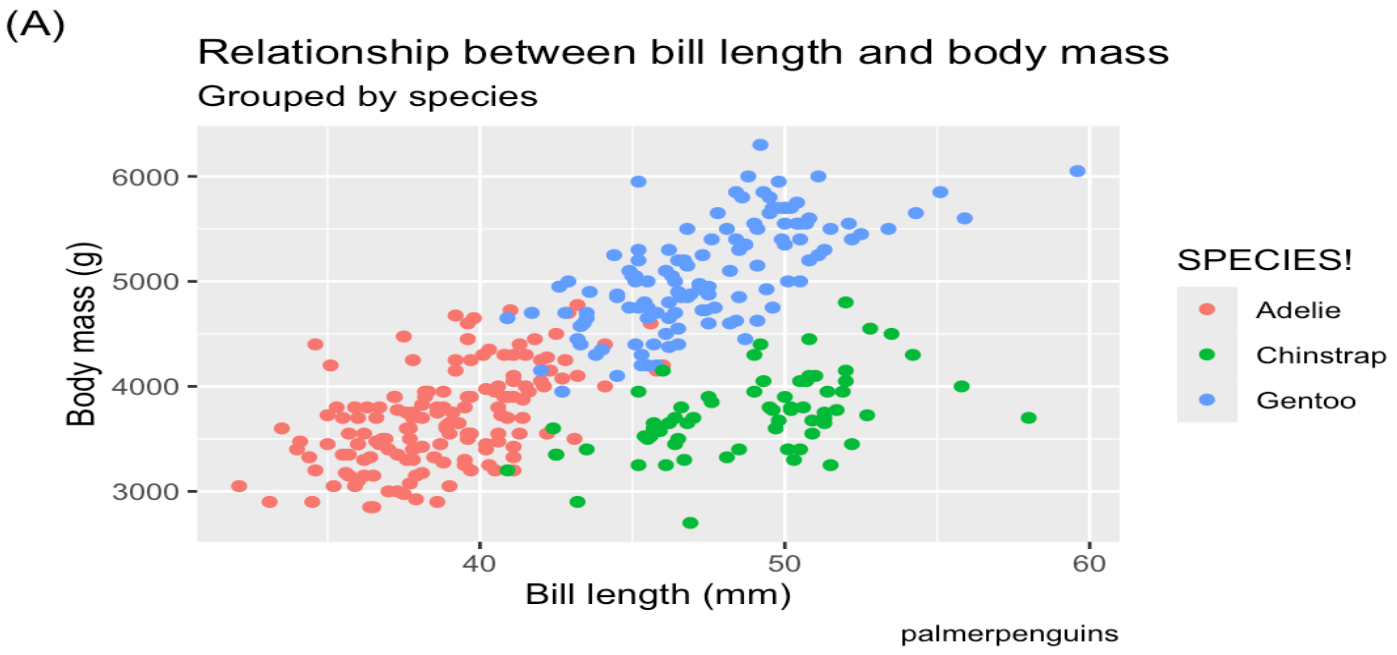
geom_function()

additional layers

Labels

Legend (one way)

```
…
 labs(x="Bill length (mm)",
       y="Body mass (g)",
       title = "Relationship between bill
                length and body mass",
       subtitle="Grouped by species",
       caption = "palmerpenguins",
       tag = "(A)",
       color = "Species!")
```



data

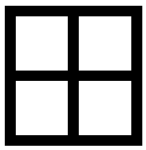
ggplot(aes(mapping))

layers

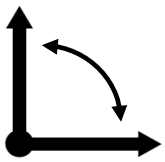
geom_function()

additional layers

additional layers



Facet wrap

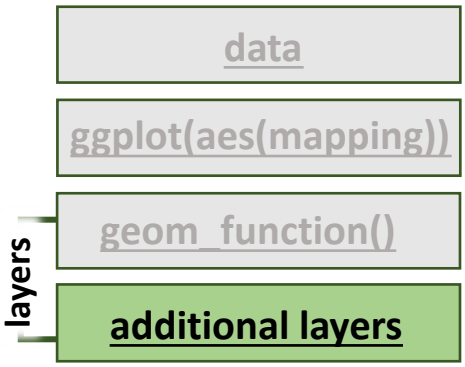


Coord flip

Scales

Labels

Themes



Themes

theme_*

- The default is theme_gray()
- There are a lot of built-in alternative in {ggplot2}
- My go-to is theme_minimal() because it is clean without a lot of unnecessary visuals
- Set theme globally: theme_set(theme_minimal()) at the start of your file (after loading libraries)

data

ggplot(aes(mapping))

layers

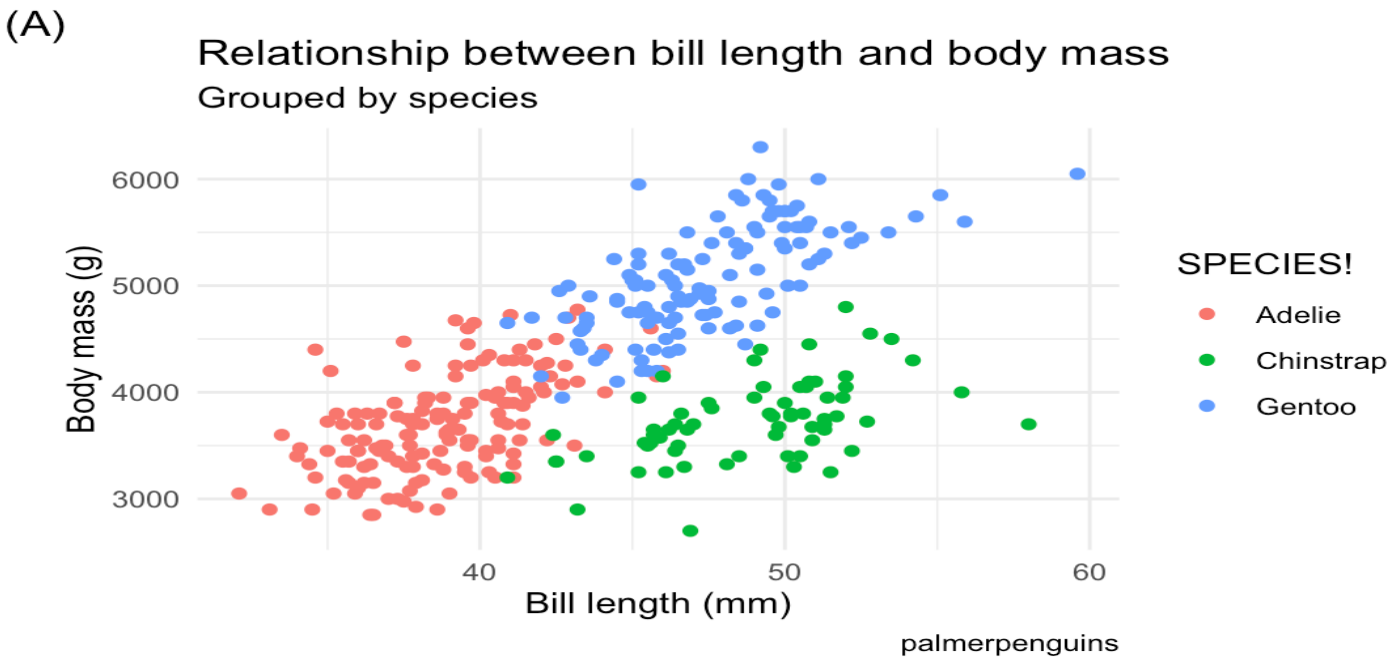
geom_function()

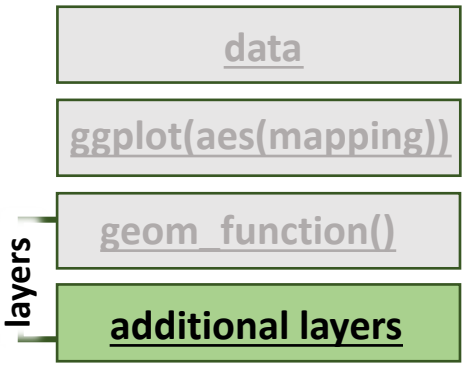
additional layers

Themes

theme_*

```
penguins |>
  ggplot(aes(bill_length_mm, body_mass_g)) +
  geom_point(aes(color = species)) +
  labs(x="Bill length (mm)",
       y="Body mass (g)",
       title = "Relationship between bill
  length and body mass",
       subtitle="Grouped by species",
       caption = "palmerpenguins",
       tag = "(A)",
       color="SPECIES!") +
  theme_minimal()
```





Themes

theme_*

- Other packages:
  - {ggthemes}
  - {ggthemeassist}
  - {hrbrthemes}
  - Build your own theme

data

ggplot(aes(mapping))

layers

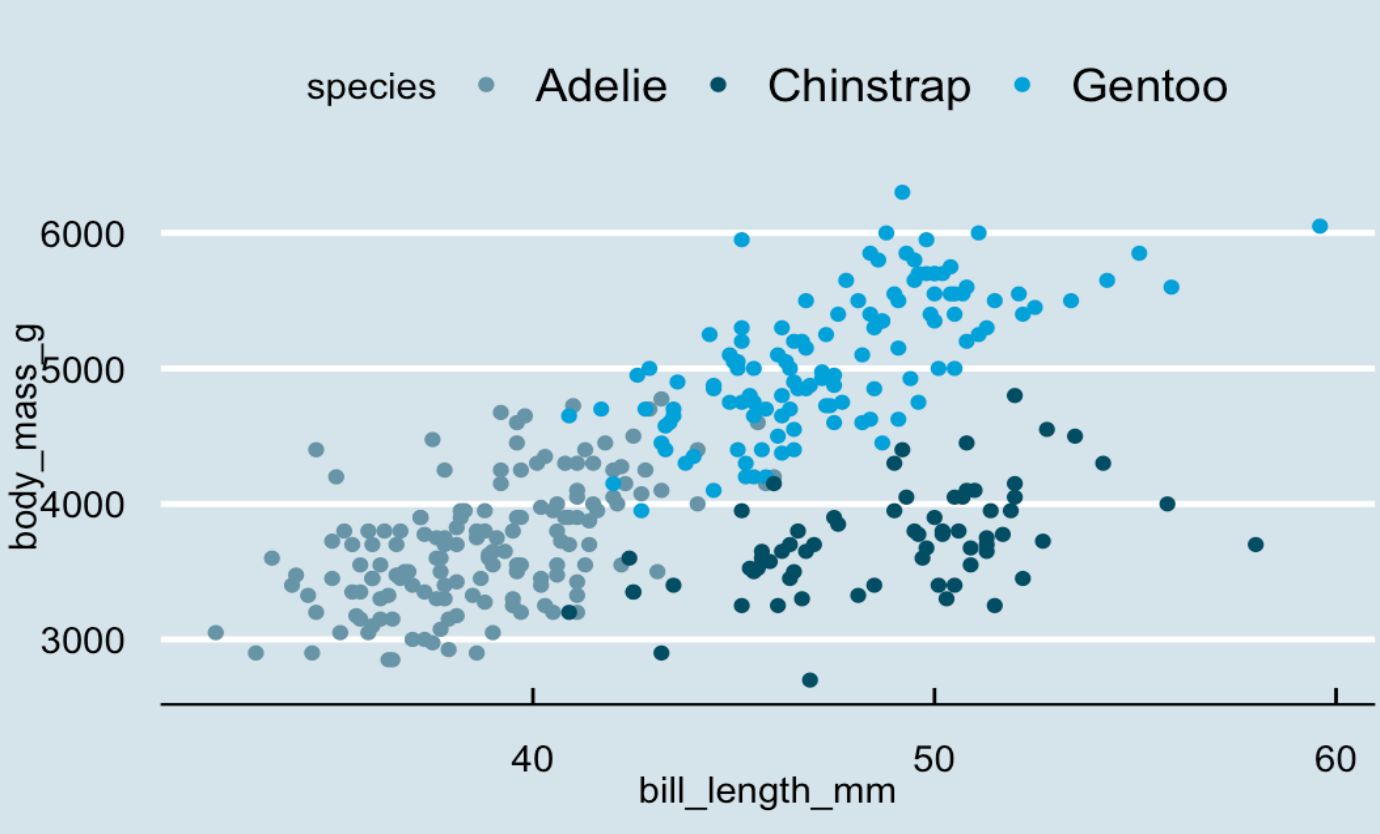
geom_function()

additional layers

Themes

theme_*

```
penguins |>
  ggplot(aes(bill_length_mm,
  body_mass_g)) +
  geom_point() +
  ggthemes::theme_economist()+
  ggthemes::scale_color_economist()
```



data

ggplot(aes(mapping))

layers

geom_function()

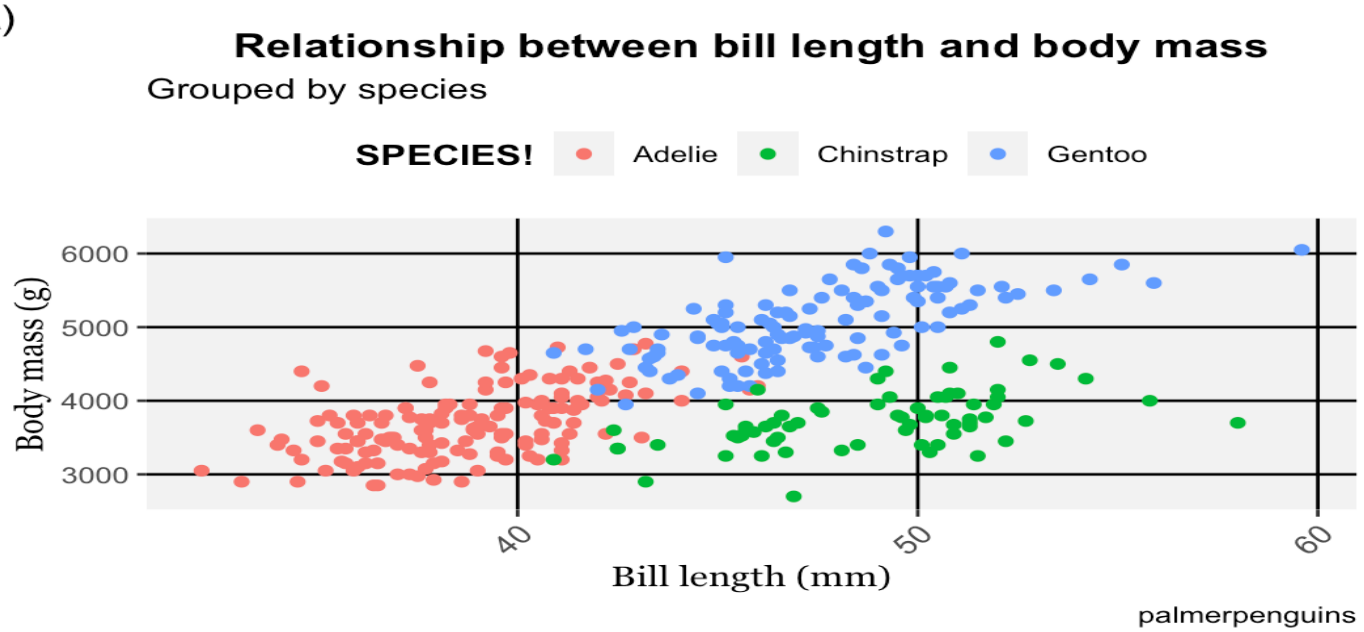
additional layers

Themes

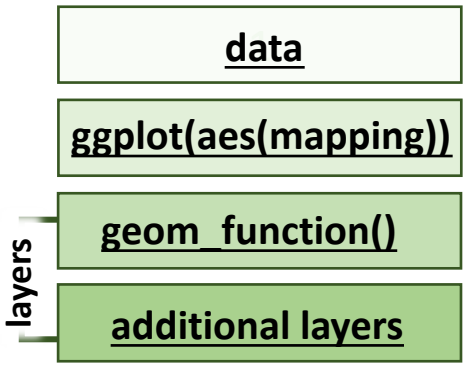
theme()

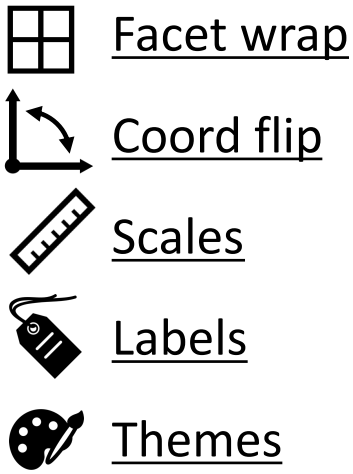
```
…
theme(plot.title         = element_text(size=16, face="bold",
                                        hjust =0.5),
      axis.title.x       = element_text(size=12,face="bold"),
      axis.text.x        = element_text(size=10,angle=45,
                                        hjust=1),
      panel.background   = element_rect(fill="white"),
      plot.background    = element_rect(fill = "grey"),
      panel.grid.major   = element_line(color="black"),
      panel.grid.minor   = element_blank(),
      legend.position    = "top",
      legend.title       = element_text(face="bold"),
      legend.background= element_rect(fill="transparent"))
```

(A)



Review/Index





		Only 1 Variable	Variable 2 Continuous	Variable 2 Discrete
Variable 1 Continuous		geom_histogram geom_density	geom_point geom_smooth geom_line	geom_density_ridges *(from {ggridges})* geom_boxplot geom_violin geom_col
Variable 1 Discrete		geom_bar	X	geom_count

Heatmaps: geom_tile

Other Topics

- color vs fill
- vars
- conditional vs fixed
- transparency (alpha)
- global vs local

UNIVERSITY OF OREGON

Practice

1. Get to know the data - `str(mpg)` or `head(mpg)`
2. What is the overall distribution of city fuel efficiency (mpg) across car models?
3. How does the distribution vary by drivetrain type (e.g., front-, rear-, 4-wheel drive)?
4. What is the raw relationship between city and highway mpg?
5. Can we focus on/emphasize Audi's relationship?
6. Can we have larger points for clarity?
7. How are the city/hwy mpg relationships different by car class?
8. Too much clutter. Can we just see trends?
9. Still too much clutter. Better way to clearly see each trends?
10. Can we make it colorblind friendly?
11. Can we clarify axis and legend labels?
12. Can we polish the appearance with a theme?