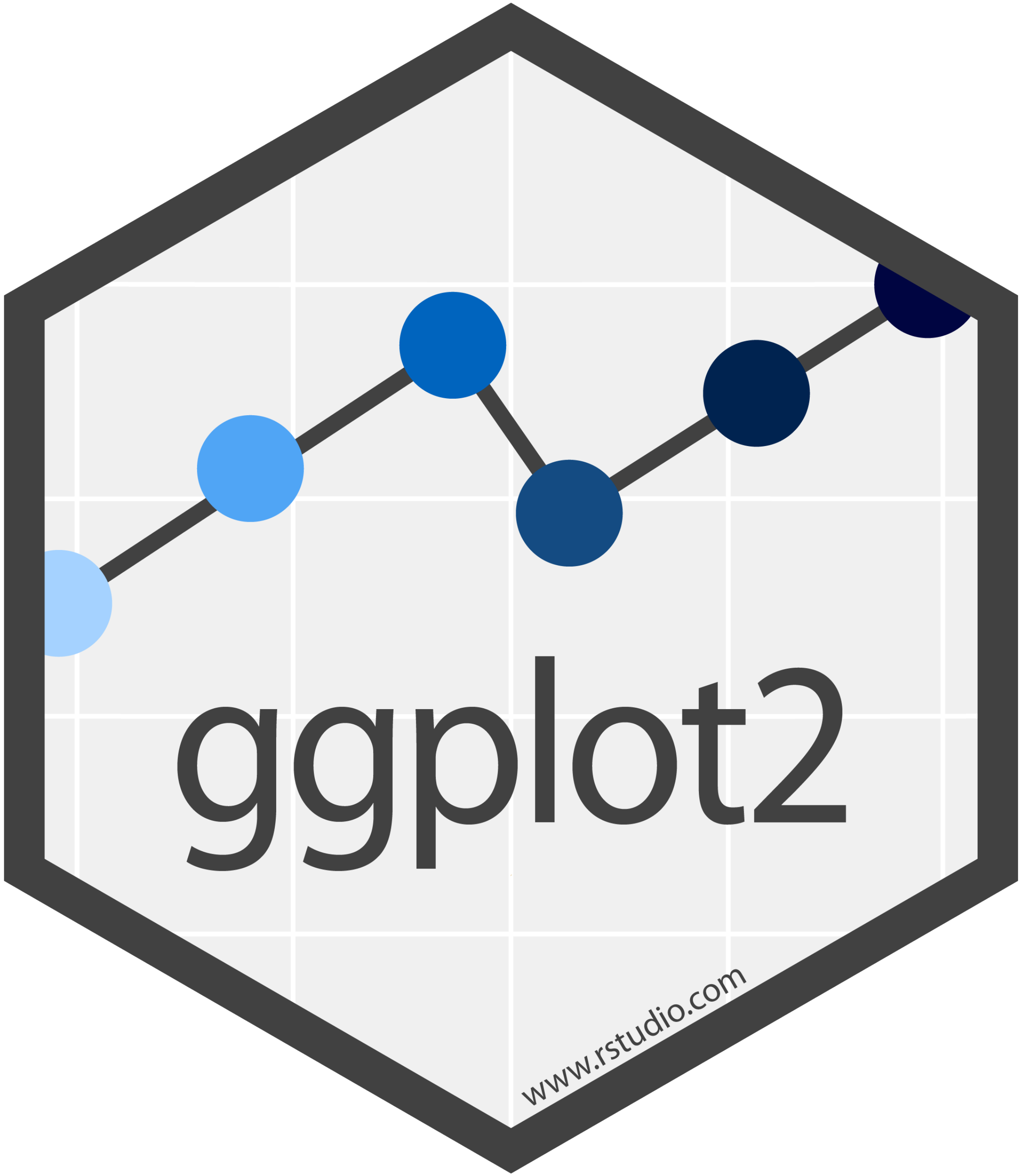
Lec 09 - Visualization with ggplot2

Lecture 09

Dr. Colin Rundel


ggplot2
www.rstudio.com

The Grammar of Graphics

- Visualization concept created by Leland Wilkinson (The Grammar of Graphics, 1999)
- attempt to taxonomize the basic elements of statistical graphics
- Adapted for R by Hadley Wickham (2009)
  - consistent and compact syntax to describe statistical graphics
  - highly modular as it breaks up graphs into semantic components
  - ggplot2 is not meant as a guide to which graph to use and how to best convey your data (more on that later), but it does have some strong opinions.

Terminology

A statistical graphic is a…

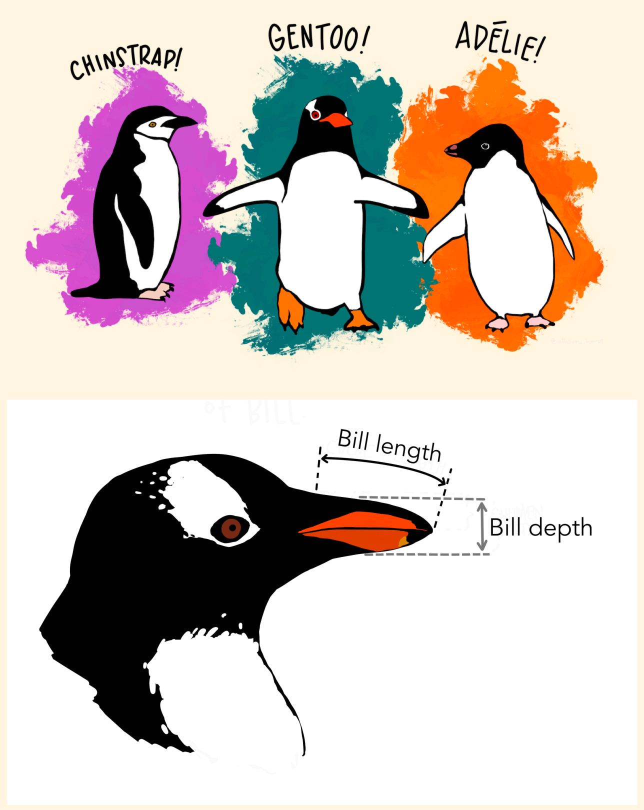
- mapping of **data**
- which may be **statistically transformed** (summarized, log-transformed, etc.)
- to **aesthetic attributes** (color, size, xy-position, etc.)
- using **geometric objects** (points, lines, bars, etc.)
- and mapped onto a specific **facet** and **coordinate system**

Anatomy of a ggplot call

```
ggplot(
  data = [dataframe],
  mapping = aes(
    x = [var x], y = [var y],
    color = [var color],
    shape = [var shape],
    ...
  )
) +
  geom_[some geom](
    mapping = aes(
      color = [var geom color],
      ...
    )
  ) +
  ... # other geometries
  scale_[some axis]_[some scale]() +
  facet_[some facet]([formula]) +
  ... # other options
```

Data - Palmer Penguins

Measurements for penguin species, island in Palmer Archipelago, size (flipper length, body mass, bill dimensions), and sex.



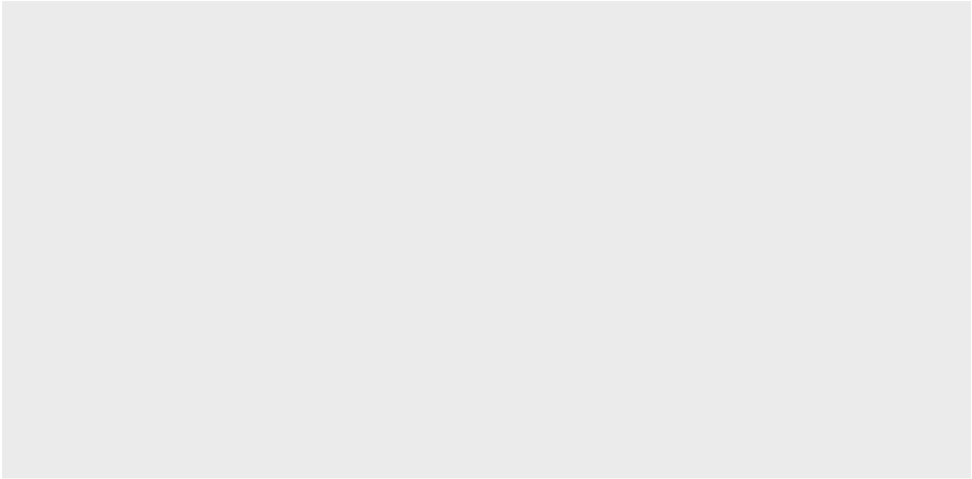
```r
library(palmerpenguins)
penguins
```

```
# A tibble: 344 × 8
   species island    bill_length_mm bill_depth_mm
   <fct>   <fct>              <dbl>         <dbl>
 1 Adelie  Torgersen           39.1          18.7
 2 Adelie  Torgersen           39.5          17.4
 3 Adelie  Torgersen           40.3          18  
 4 Adelie  Torgersen           NA            NA  
 5 Adelie  Torgersen           36.7          19.3
 6 Adelie  Torgersen           39.3          20.6
 7 Adelie  Torgersen           38.9          17.8
 8 Adelie  Torgersen           39.2          19.6
 9 Adelie  Torgersen           34.1          18.1
10 Adelie  Torgersen           42            20.2
# ℹ 334 more rows
# ℹ 4 more variables: flipper_length_mm <int>,
#   body_mass_g <int>, sex <fct>, year <int>
```

Text <-> Plot

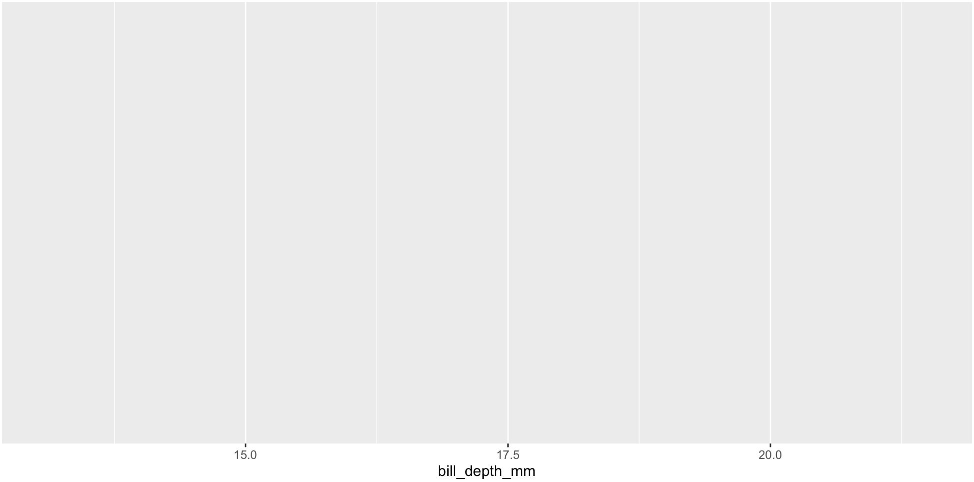
> *Start with the `penguins` data frame*

```r
ggplot(data = penguins)
```



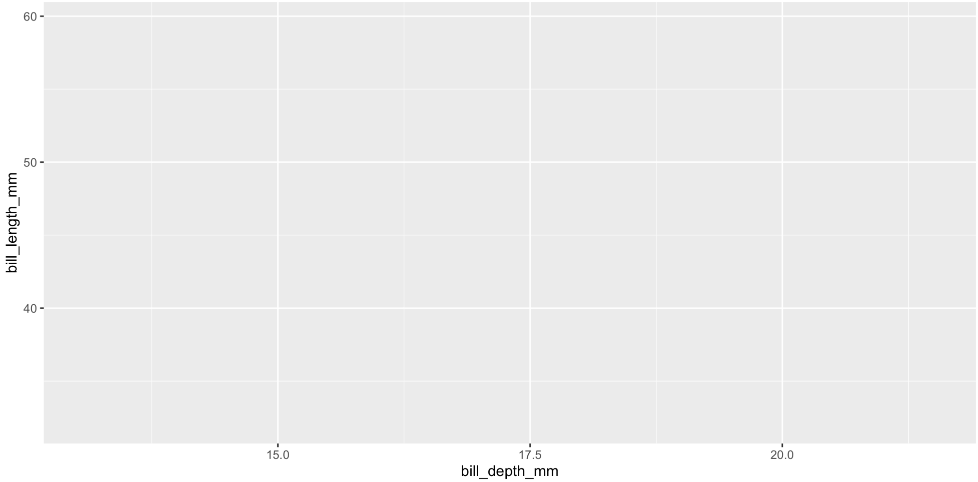
> Start with the `penguins` data frame, *map bill depth to the x-axis*

```
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm
  )
)
```



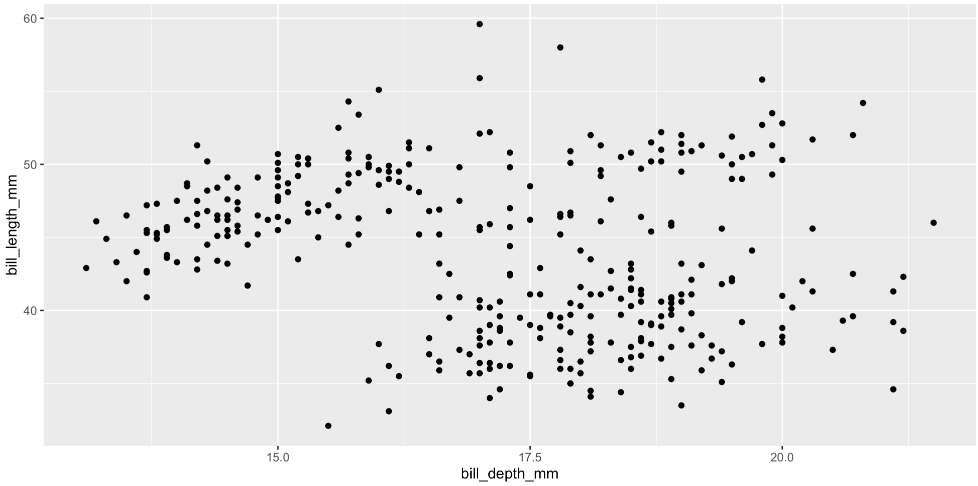
> Start with the `penguins` data frame, map bill depth to the x-axis *and map bill length to the y-axis.*

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



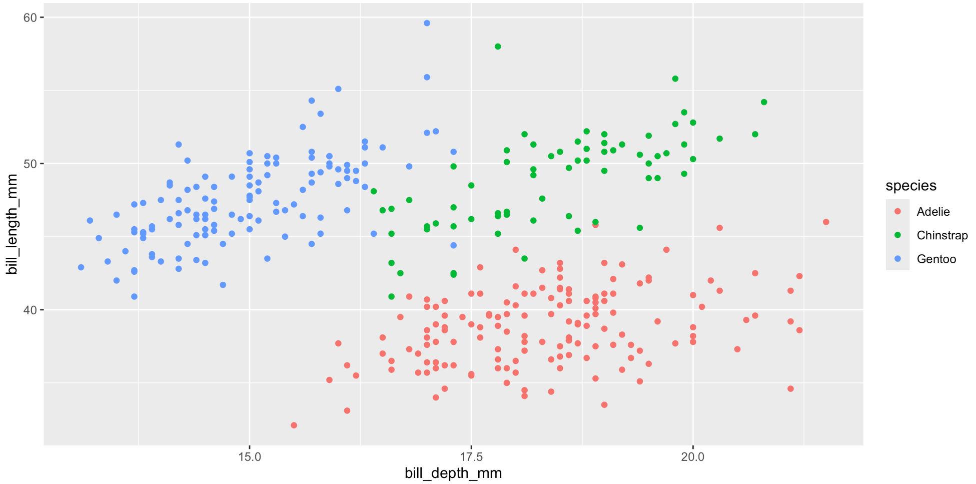
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. *Represent each observation with a point*

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



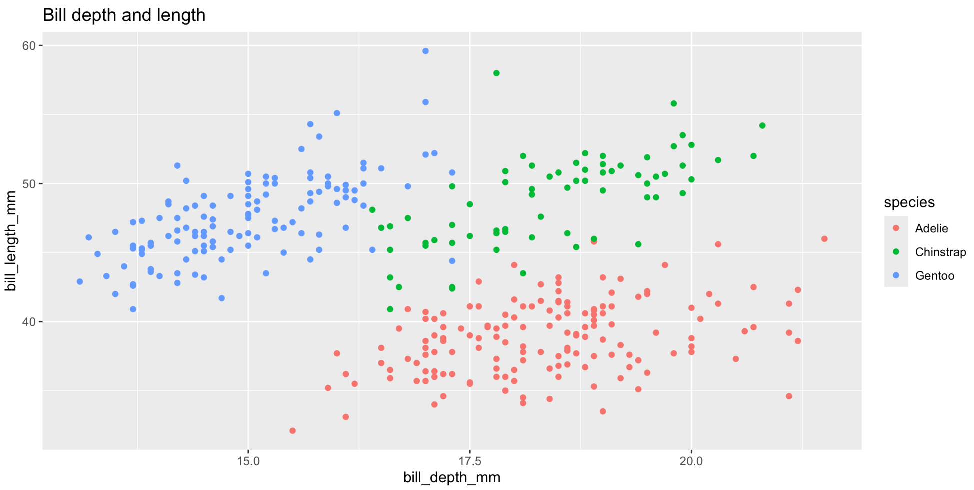
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point *and map species to the color of each point.*

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



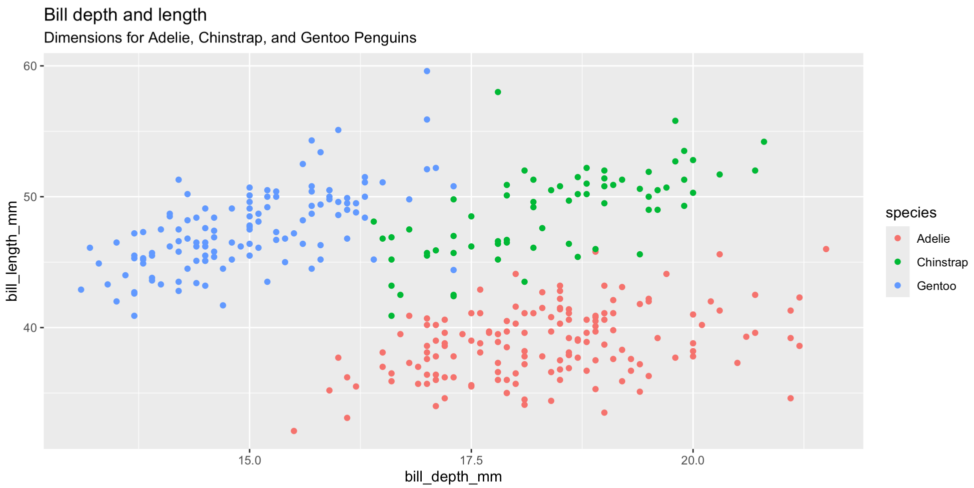
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. *Title the plot “Bill depth and length”*

```r
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm,
    y = bill_length_mm
  )
) +
  geom_point(
    mapping = aes(color = species)
  ) +
  labs(title = "Bill depth and length")
```



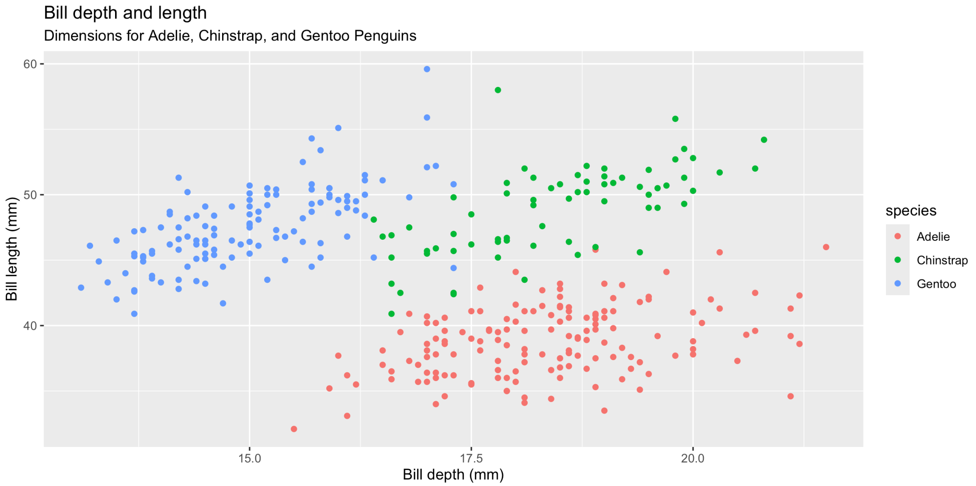
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. Title the plot “Bill depth and length”, *add the subtitle “Dimensions for Adelie, Chinstrap, and Gentoo Penguins”*

```r
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm,
    y = bill_length_mm
  )
) +
  geom_point(
    mapping = aes(color = species)
  ) +
  labs(
    title = "Bill depth and length",
    subtitle = paste("Dimensions for Adelie
                     "Chinstrap, and Gentoo
                     "Penguins")
  )
```



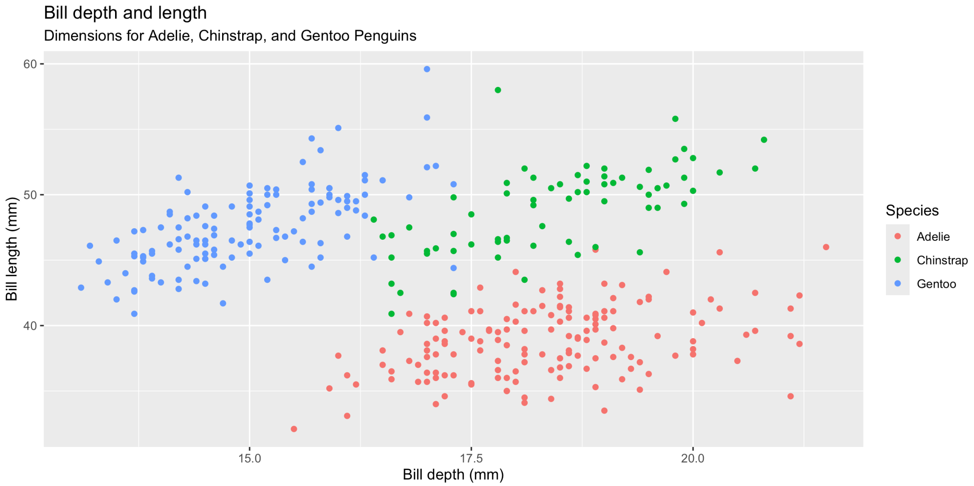
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. Title the plot “Bill depth and length”, add the subtitle “Dimensions for Adelie, Chinstrap, and Gentoo Penguins”, *label the x and y axes as “Bill depth (mm)” and “Bill length (mm)”, respectively*

```r
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm,
    y = bill_length_mm
  )
) +
  geom_point(
    mapping = aes(color = species)
  ) +
  labs(
    title = "Bill depth and length",
    subtitle = paste("Dimensions for Adelie
                     "Chinstrap, and Gentoo
                     "Penguins"),
    x = "Bill depth (mm)",
    y = "Bill length (mm)"
  )
```



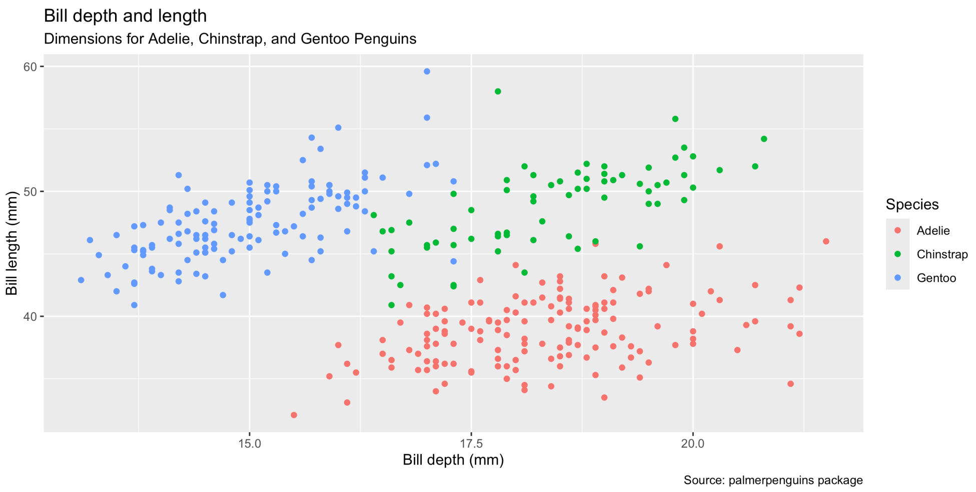
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. Title the plot “Bill depth and length”, add the subtitle “Dimensions for Adelie, Chinstrap, and Gentoo Penguins”, label the x and y axes as “Bill depth (mm)” and “Bill length (mm)”, respectively, *label the legend “Species”*

```r
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm,
    y = bill_length_mm
  )
) +
  geom_point(
    mapping = aes(color = species)
  ) +
  labs(
    title = "Bill depth and length",
    subtitle = paste("Dimensions for Adelie
                     "Chinstrap, and Gentoo"
                     "Penguins"),
    x = "Bill depth (mm)",
    y = "Bill length (mm)",
    color = "Species"
  )
```



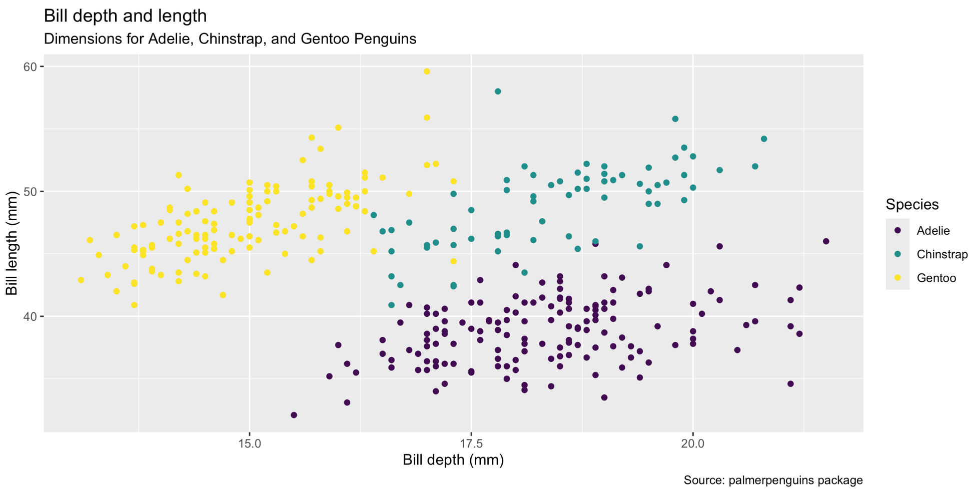
> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. Title the plot “Bill depth and length”, add the subtitle “Dimensions for Adelie, Chinstrap, and Gentoo Penguins”, label the x and y axes as “Bill depth (mm)” and “Bill length (mm)”, respectively, label the legend “Species”, *and add a caption for the data source.*

```r
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_depth_mm,
    y = bill_length_mm
  )
) +
  geom_point(
    mapping = aes(color = species)
  ) +
  labs(
    title = "Bill depth and length",
    subtitle = paste("Dimensions for Adelie
                     "Chinstrap, and Gentoo
                     "Penguins"),
    x = "Bill depth (mm)",
    y = "Bill length (mm)",
    color = "Species",
    caption = "Source: palmerpenguins packa
  )
```



> Start with the `penguins` data frame, map bill depth to the x-axis and map bill length to the y-axis. Represent each observation with a point and map species to the color of each point. Title the plot “Bill depth and length”, add the subtitle “Dimensions for Adelie, Chinstrap, and Gentoo Penguins”, label the x and y axes as “Bill depth (mm)” and “Bill length (mm)”, respectively, label the legend “Species”, and add a caption for the data source. *Finally, use the viridis color palette for all points.*

```r
ggplot(
    data = penguins,
    mapping = aes(
      x = bill_depth_mm,
      y = bill_length_mm
    )
  ) +
    geom_point(
      mapping = aes(color = species)
    ) +
    labs(
      title = "Bill depth and length",
      subtitle = paste("Dimensions for Adel
                        "Chinstrap, and Gent
                        "Penguins"),
      x = "Bill depth (mm)",
      y = "Bill length (mm)",
      color = "Species",
      caption = "Source: palmerpenguins pac
    ) +
    scale_color_viridis_d()
```



Aesthetics

Aesthetics options

Commonly used characteristics of plotting geometries that can be **mapped to a specific variable** in the data, examples include:

- position (`x`, `y`)
- `color`
- `shape`
- `size`
- `alpha` (transparency)

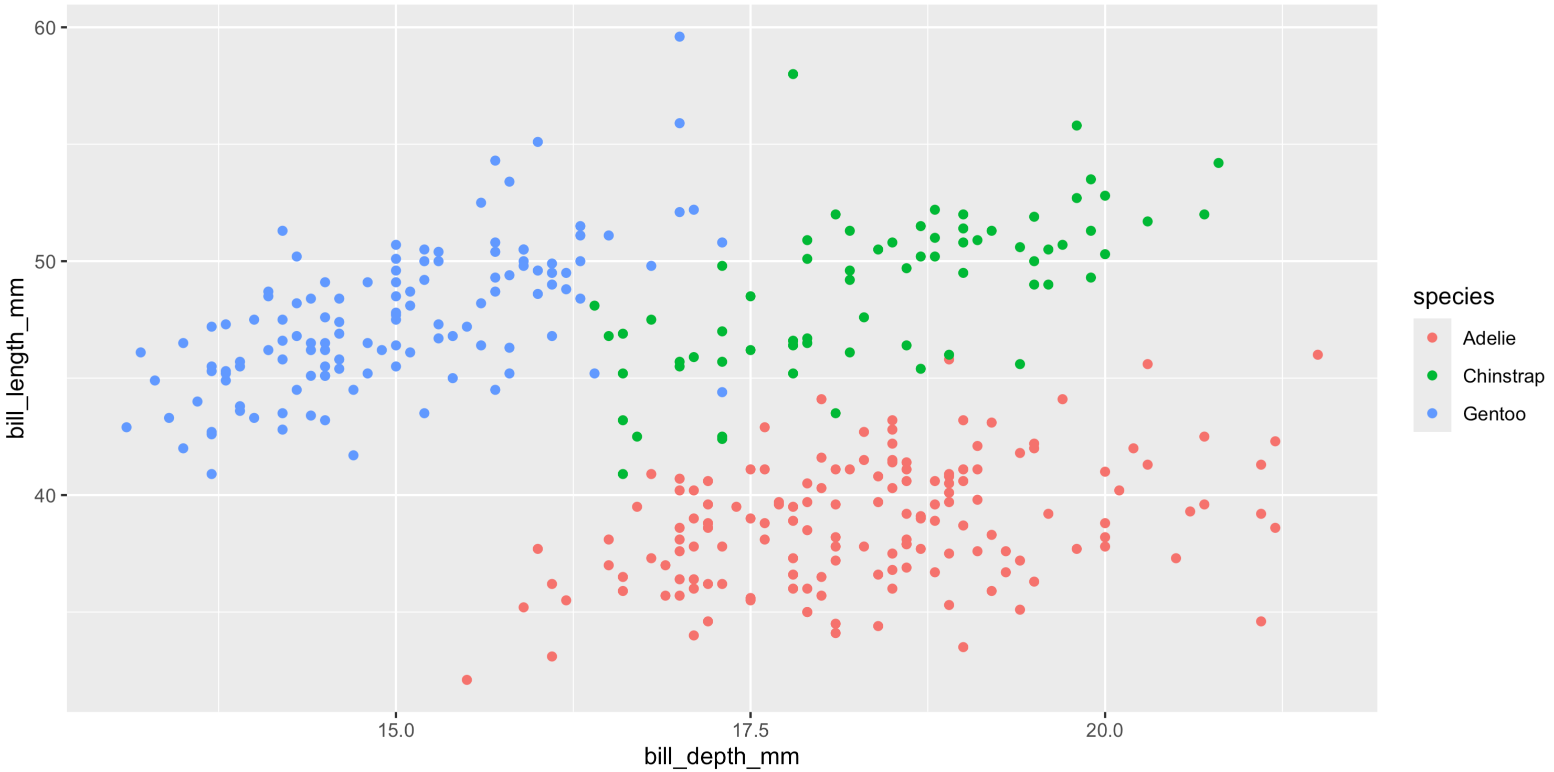
Different geometries have different aesthetics available - see the ggplot2 geoms help files for listings.

- Aesthetics given in `ggplot()` apply to all `geom`s.
- Aesthetics for a specific `geom_*()` can be overridden via `mapping` or as an argument.

color

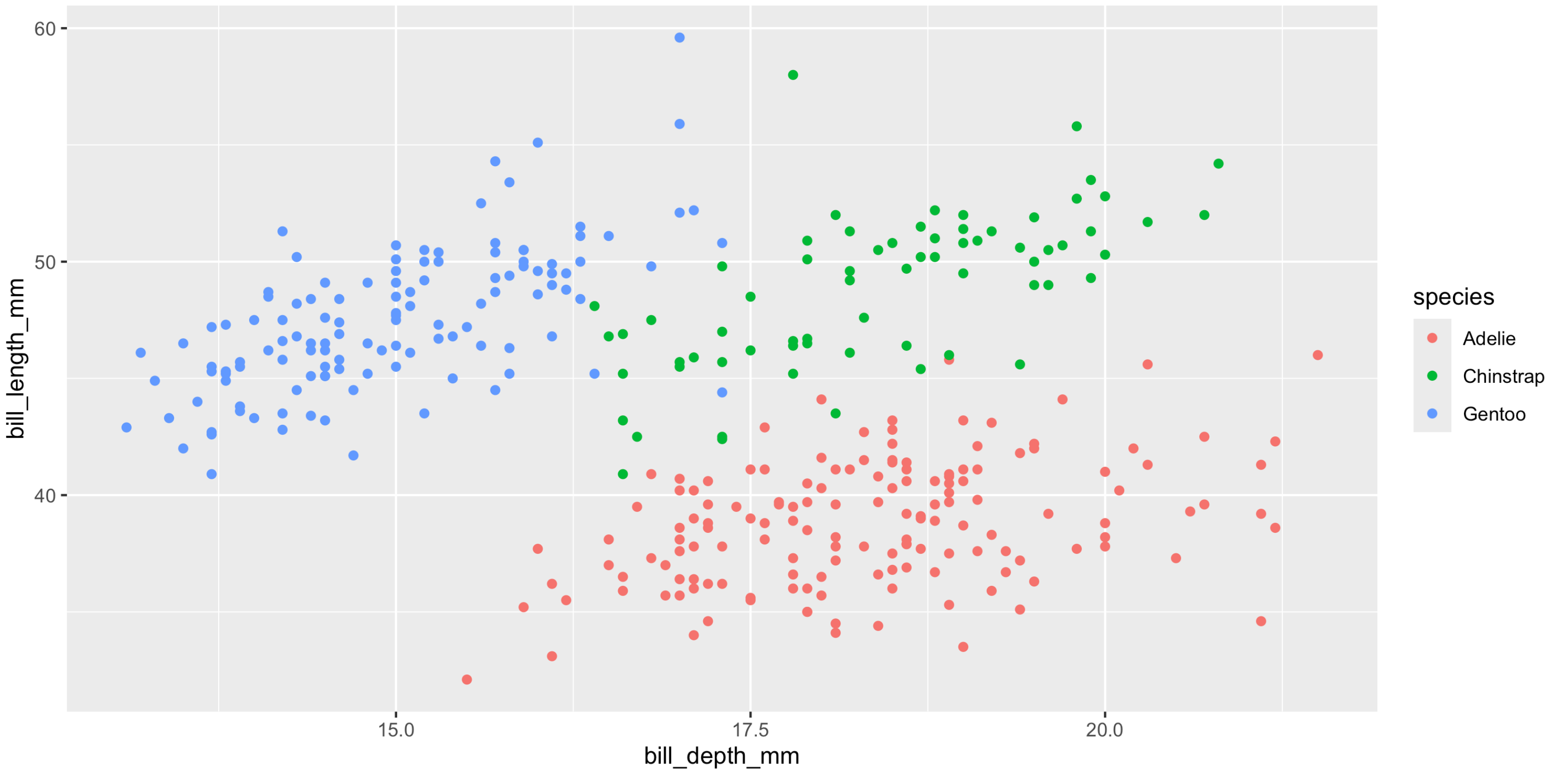
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species)
  )
```

```
Warning: Removed 2 rows containing missing values or values outside the
scale range (`geom_point()`).
```



Stop the warning

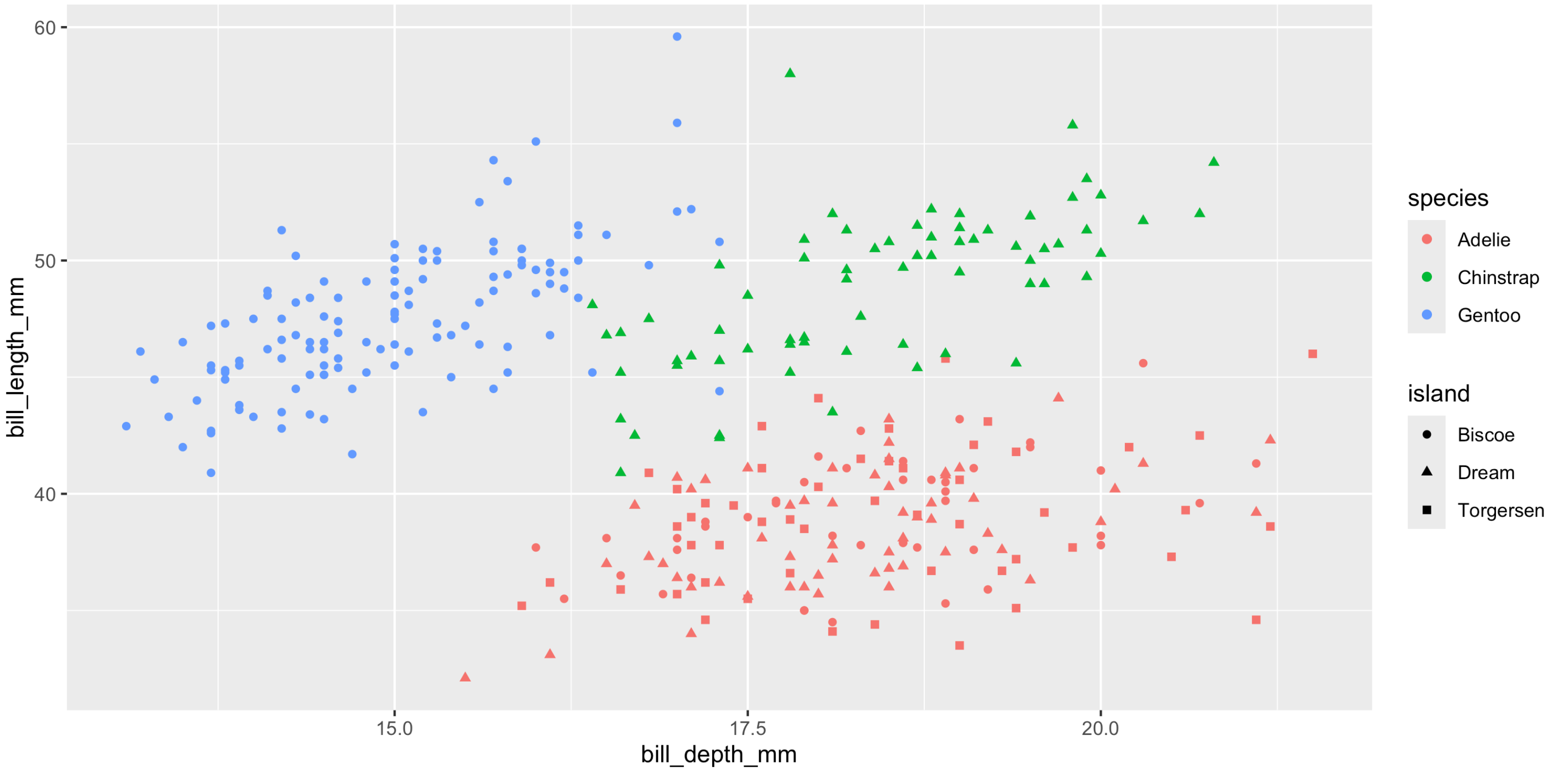
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species), na.rm=TRUE
  )
```



Shape

Mapped to a different variable than `color`

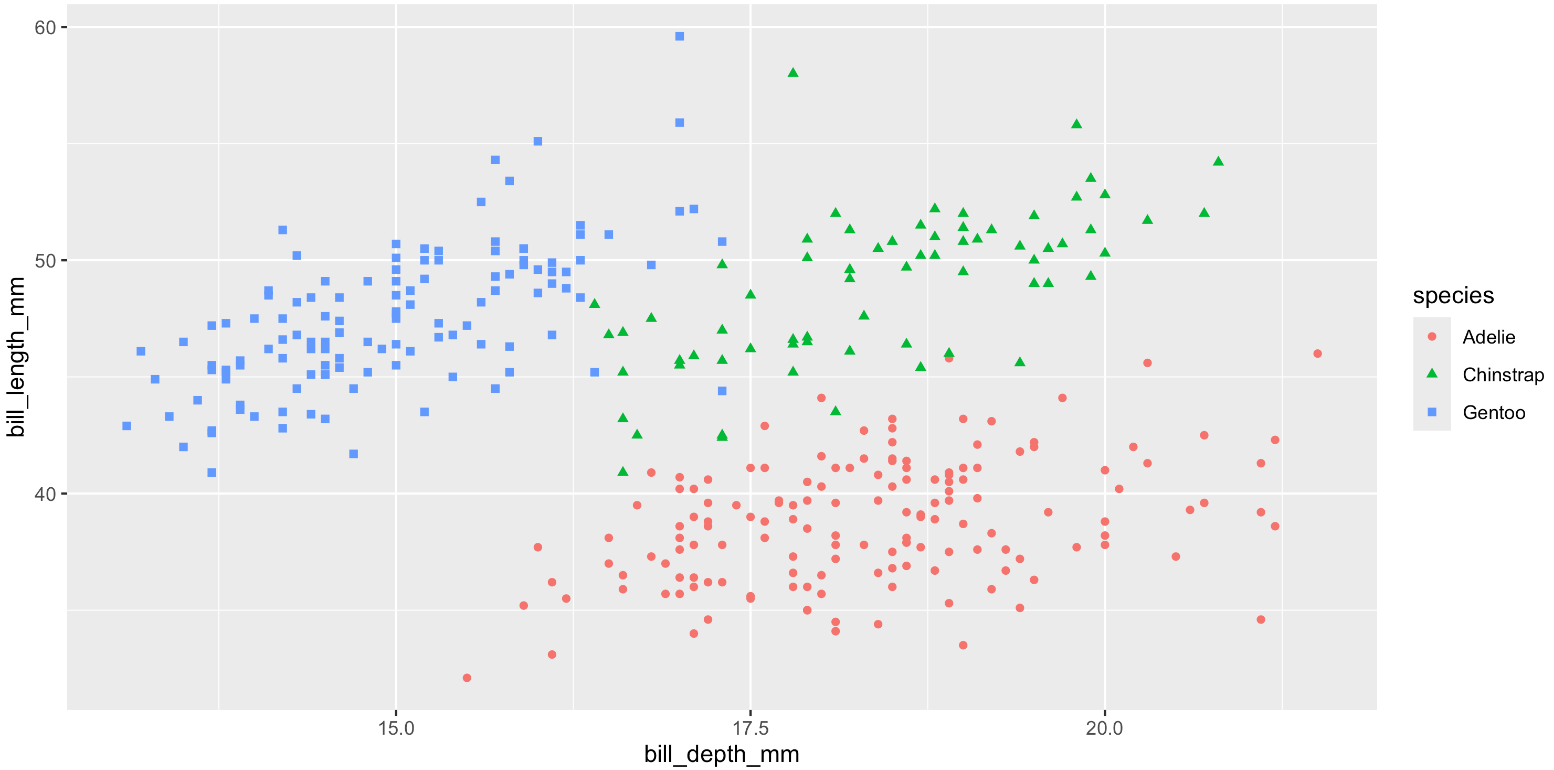
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species, shape = island), na.rm = TRUE
  )
```



Shape

Mapped to same variable as `color`

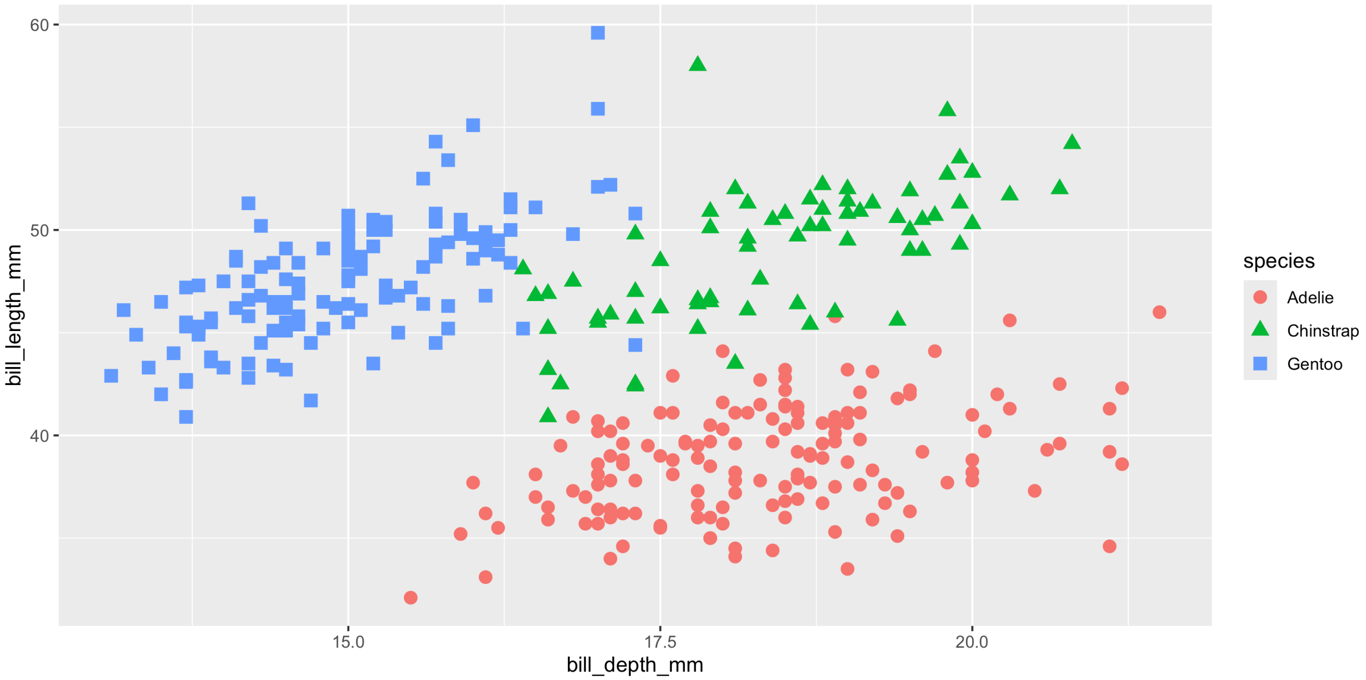
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species, shape = species), na.rm = TRUE
  )
```



Size

Using a fixed value - note that this value is outside of the `aes` call

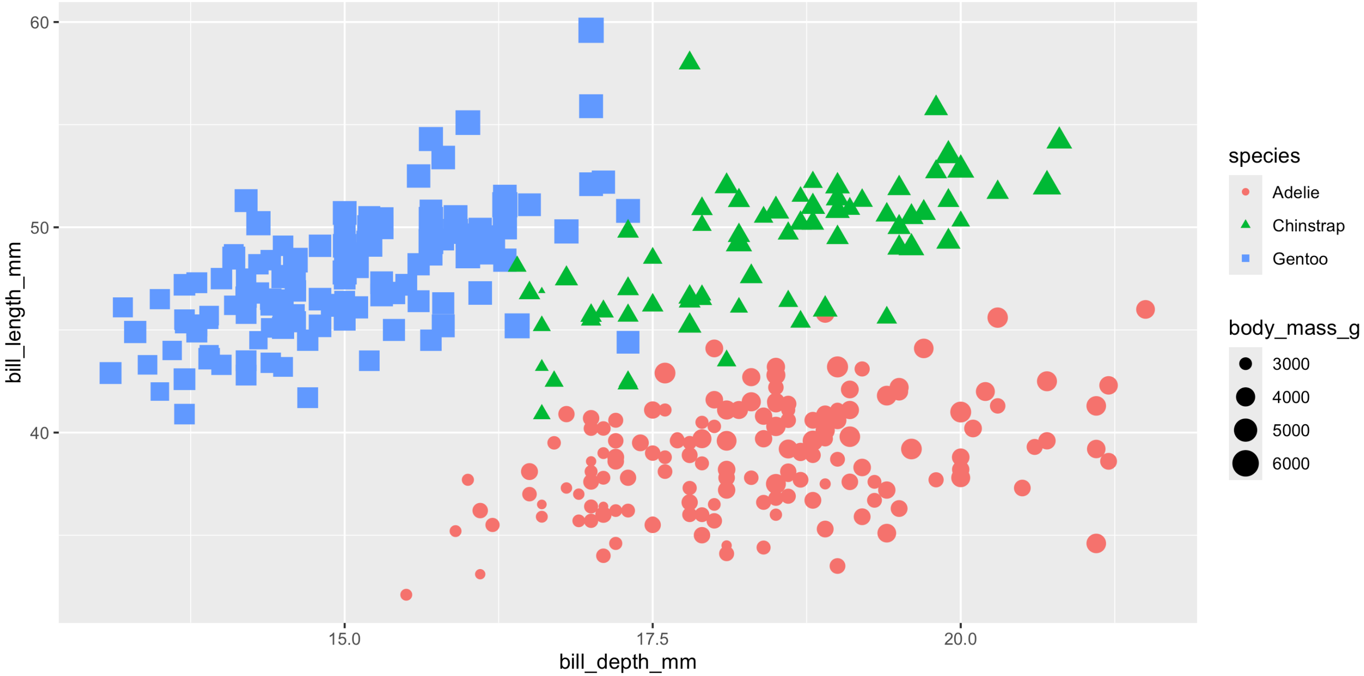
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species, shape = species), na.rm = TRUE,
    size = 3
  )
```



Size

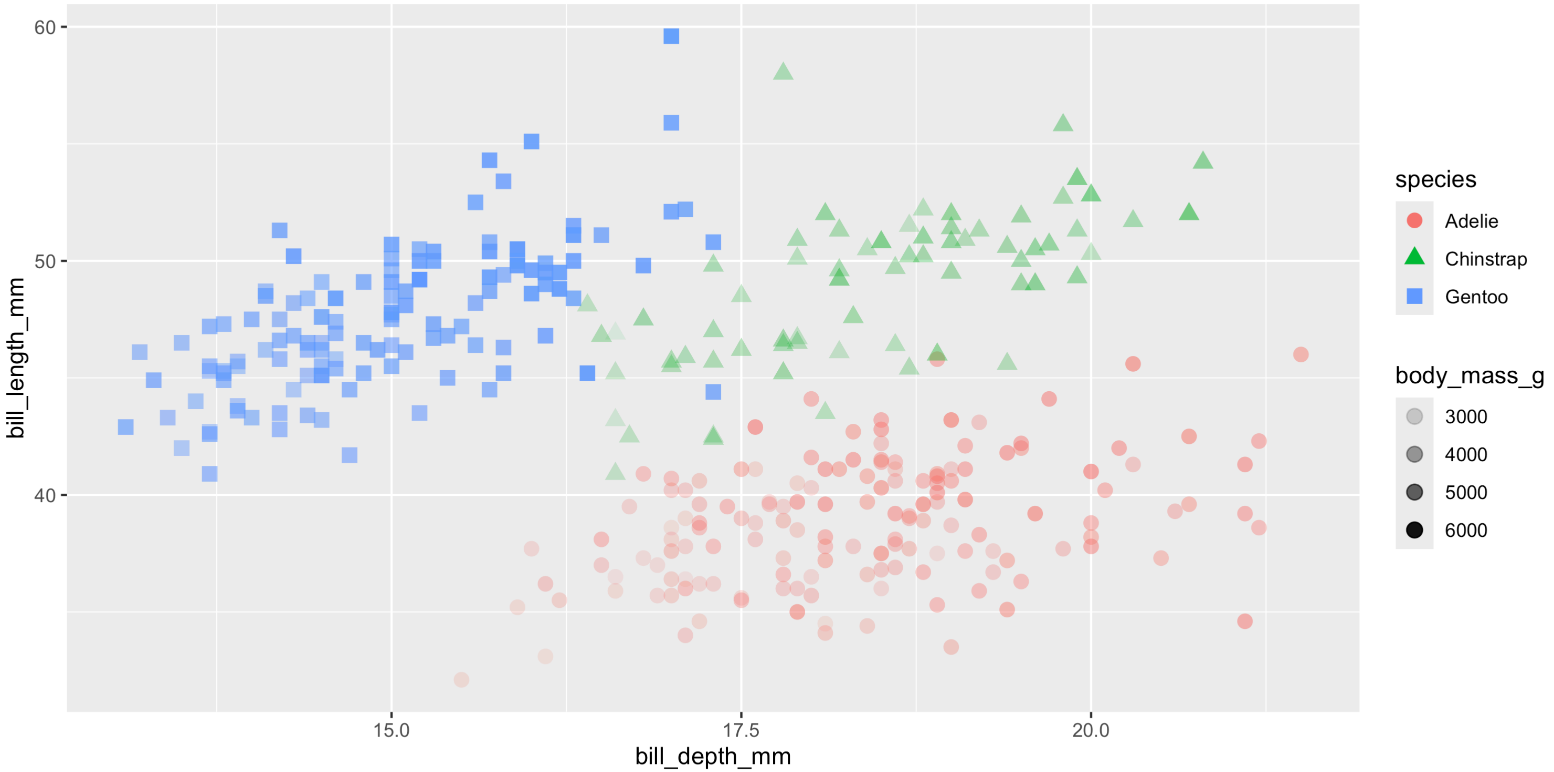
Mapped to a variable

```
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species, shape = species, size = body_mass_g), na.rm = TRUE
  )
```



Alpha

```r
ggplot(
  penguins,
  aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species, shape = species, alpha = body_mass_g), na.rm = TRUE,
    size = 3
  )
```



Mapping vs settings

- **Mapping** - Determine an aesthetic (the size, alpha, etc.) of a geom based on the values of a variable in the data
    - wrapped by `aes()` and pass as `mapping` argument to `ggplot()` or `geom_*()`.

- **Setting** - Determine an aesthetic (the size, alpha, etc.) of a geom using a constant value not directly from the data.
    - passed directly into `geom_*()` as an argument.

From the previous slide `color`, `shape`, and `alpha` are all **aesthetics** while `size` was a **setting**.

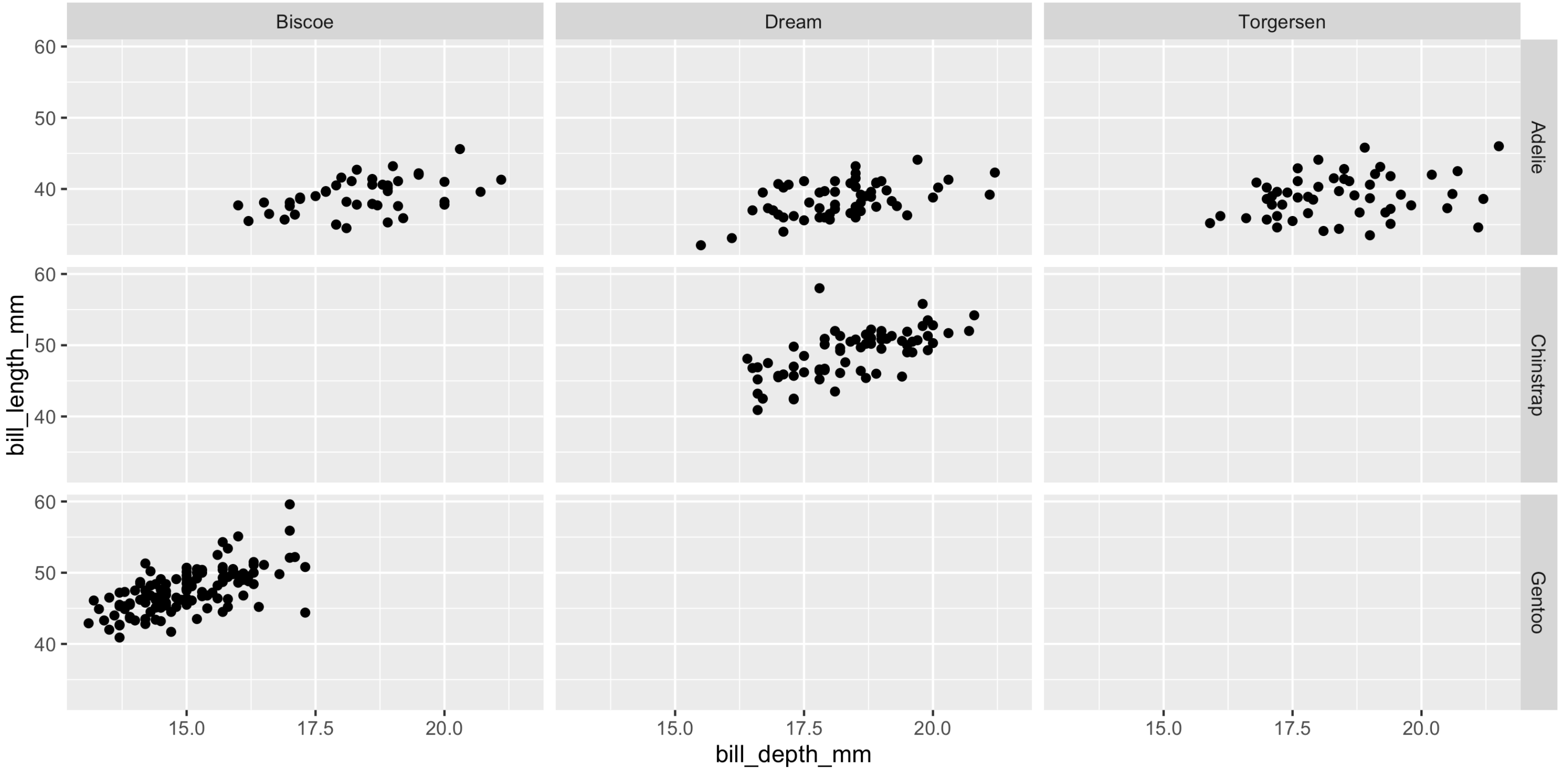
Faceting

Faceting

- Smaller plots that display different subsets of the data
- Useful for exploring conditional relationships and large data
- Sometimes referred to as “small multiples”

facet_grid

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(na.rm = TRUE) +
  facet_grid(
    species ~ island
  )
```

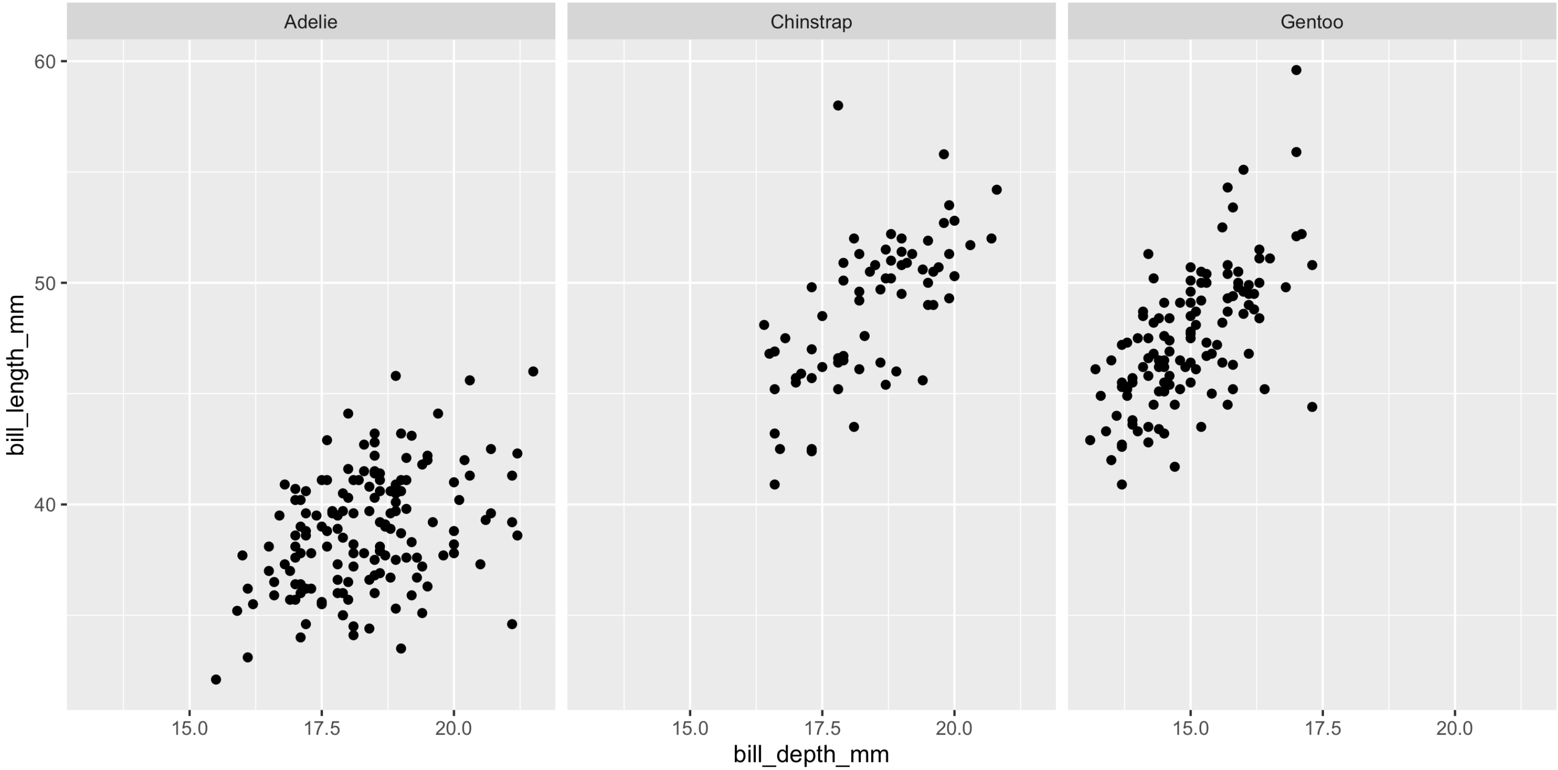


Compare with …

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(
    aes(color = species,shape = island), na.rm = TRUE, size = 3
  )
```

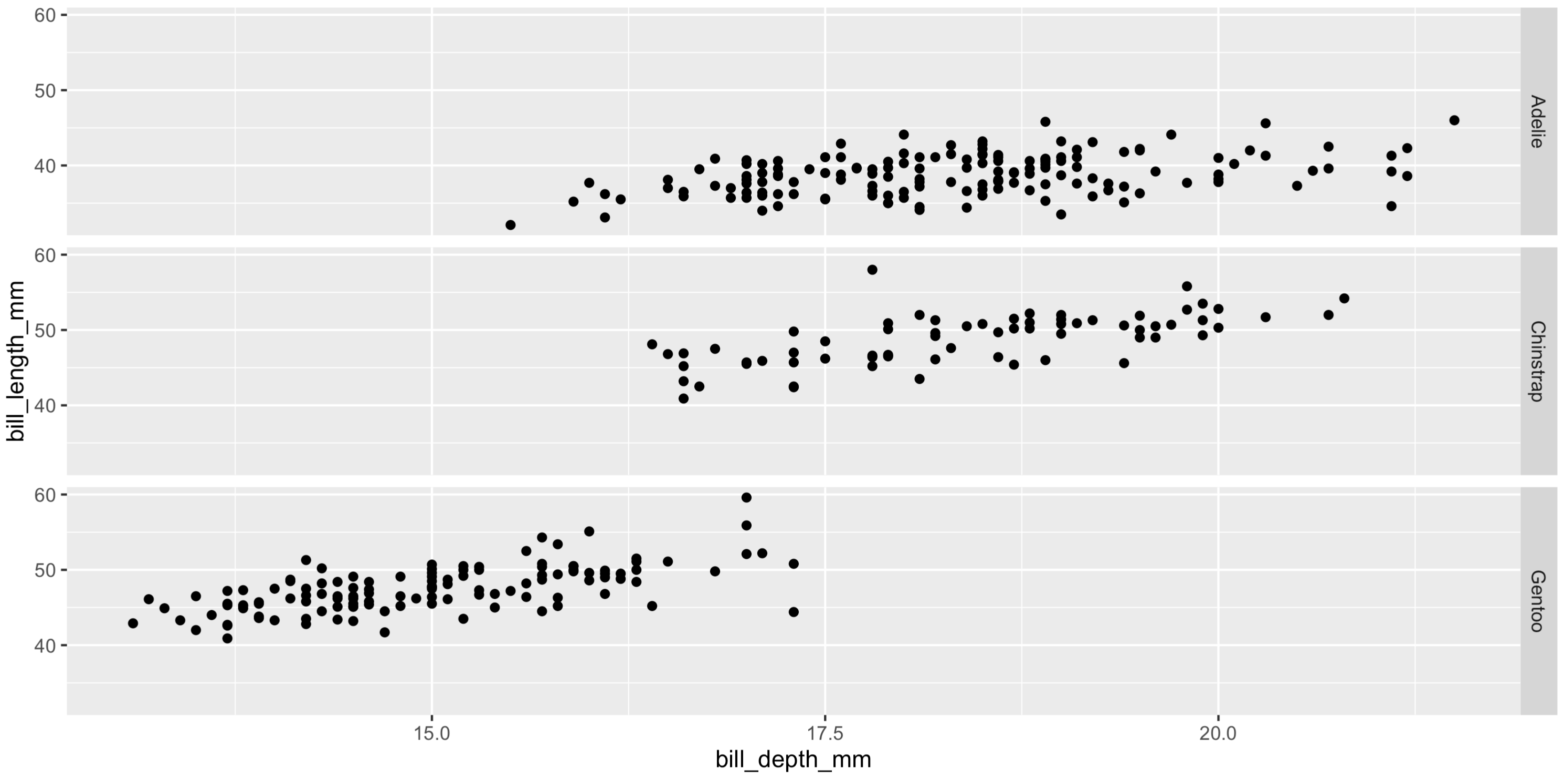
facet_grid (cols)

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(na.rm = TRUE) +
  facet_grid(
    ~ species
  )
```



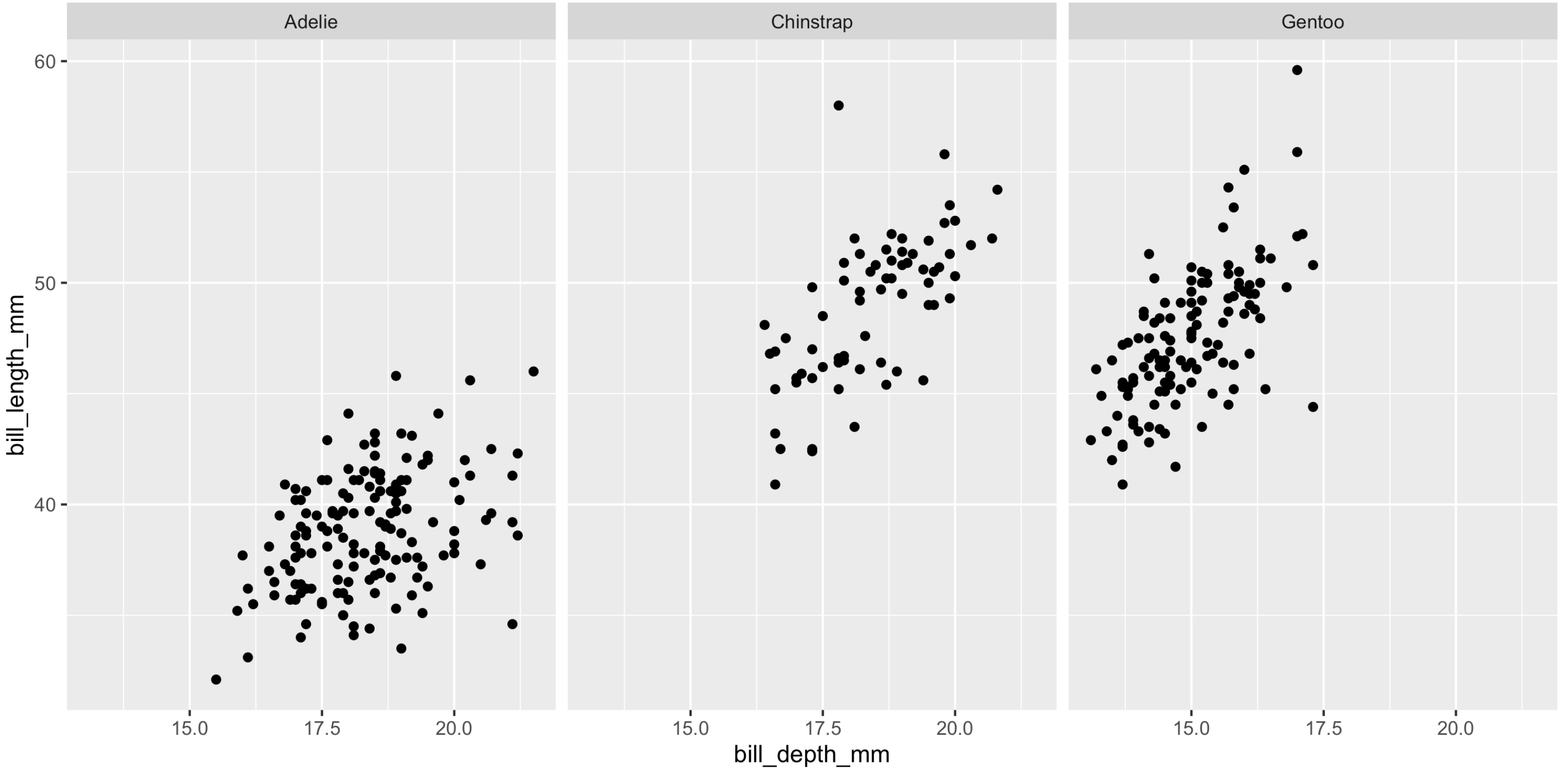
facet_grid (rows)

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(na.rm = TRUE) +
  facet_grid(
    species ~ .
  )
```



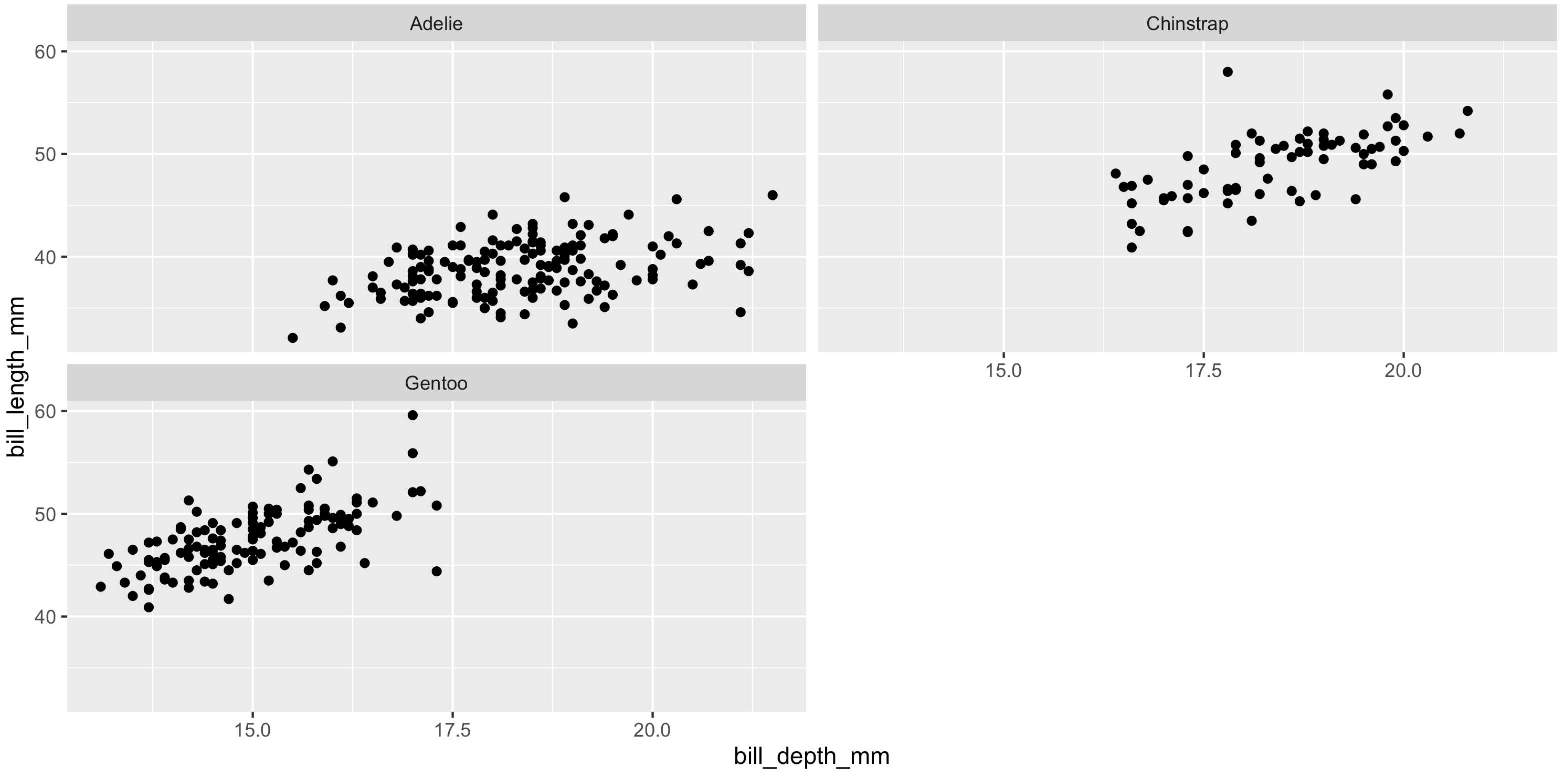
facet_wrap

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(na.rm = TRUE) +
  facet_wrap(
    ~ species
  )
```



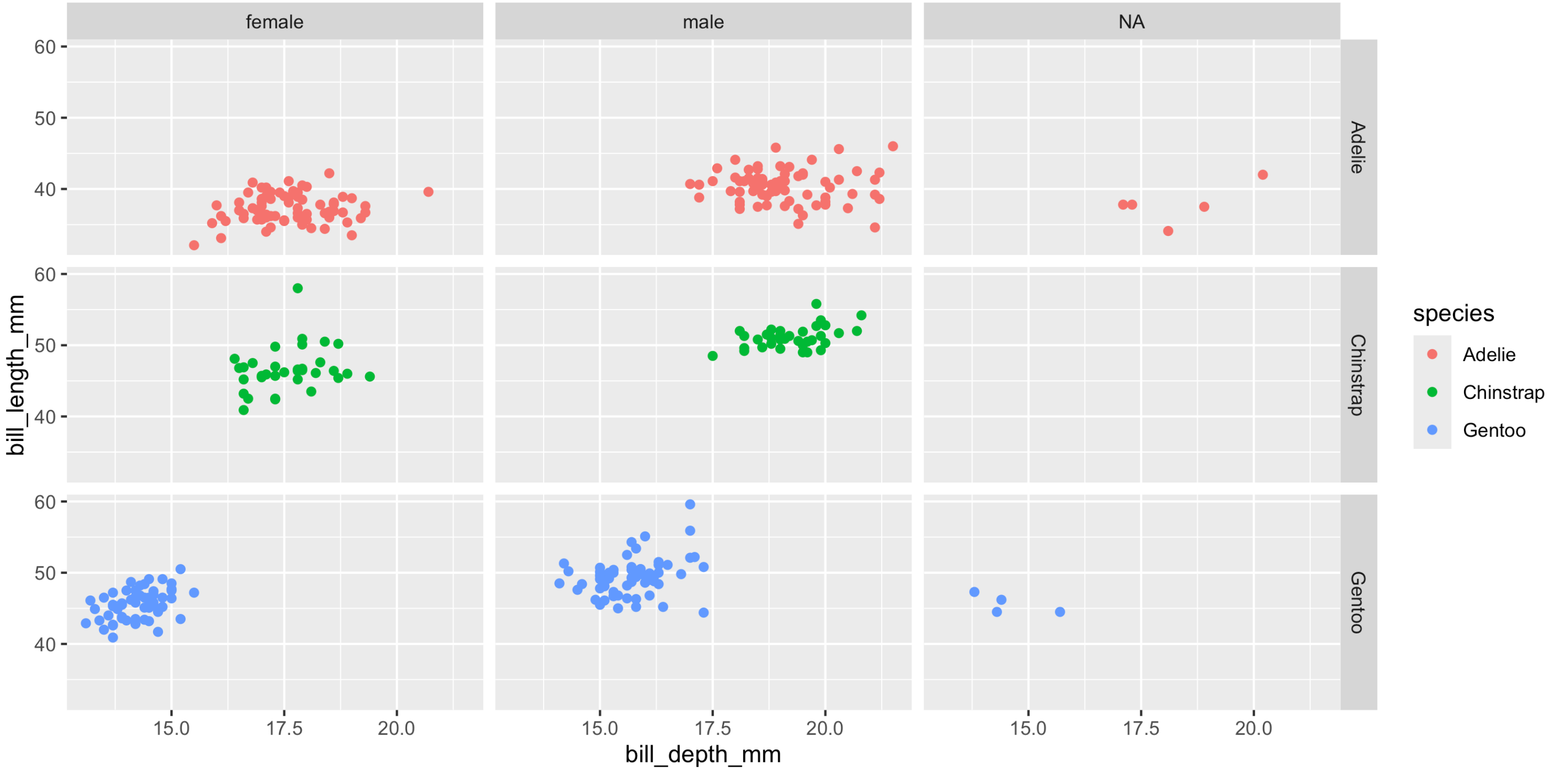
facet_wrap

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm)
) +
  geom_point(na.rm = TRUE) +
  facet_wrap(
    ~ species, ncol = 2
  )
```



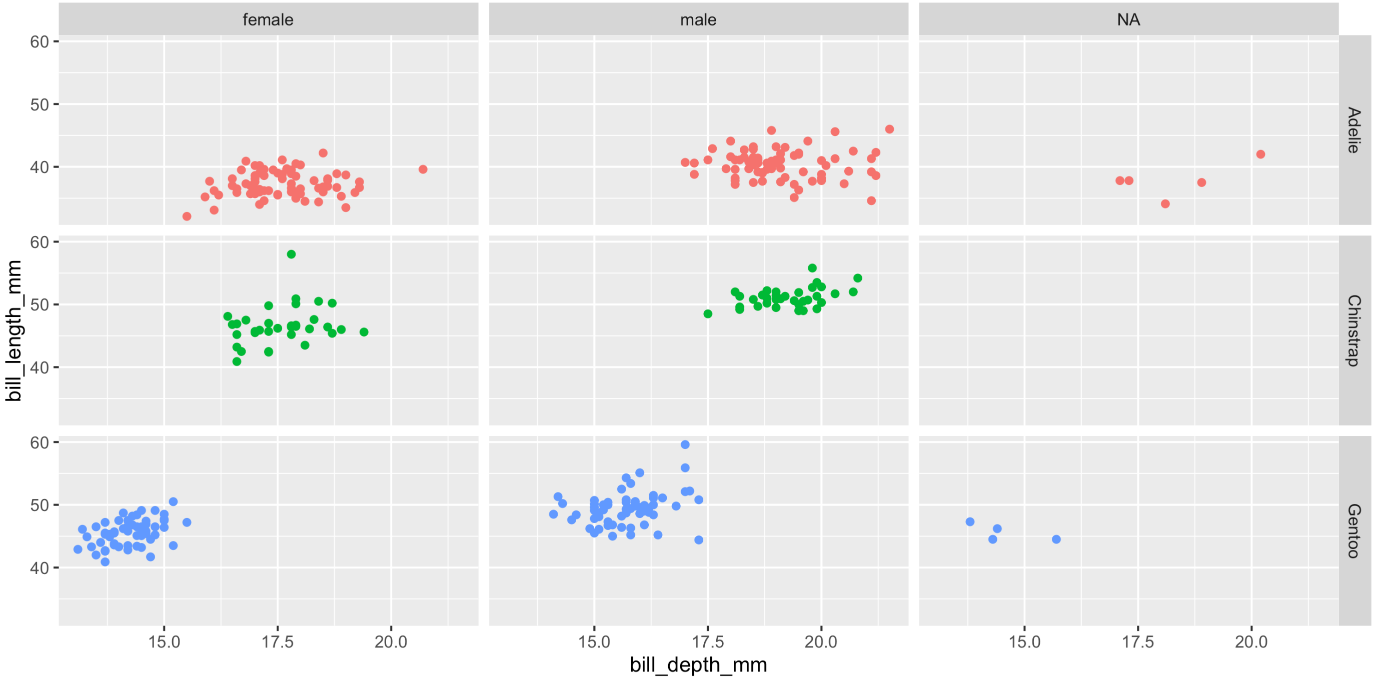
Faceting and color

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm, color = species)
) +
  geom_point(na.rm = TRUE) +
  facet_grid(
    species ~ sex
  )
```



Hiding legend elements

```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm, color = species)
) +
  geom_point(na.rm = TRUE) +
  facet_grid(
    species ~ sex
  ) +
  guides(color = "none")
```

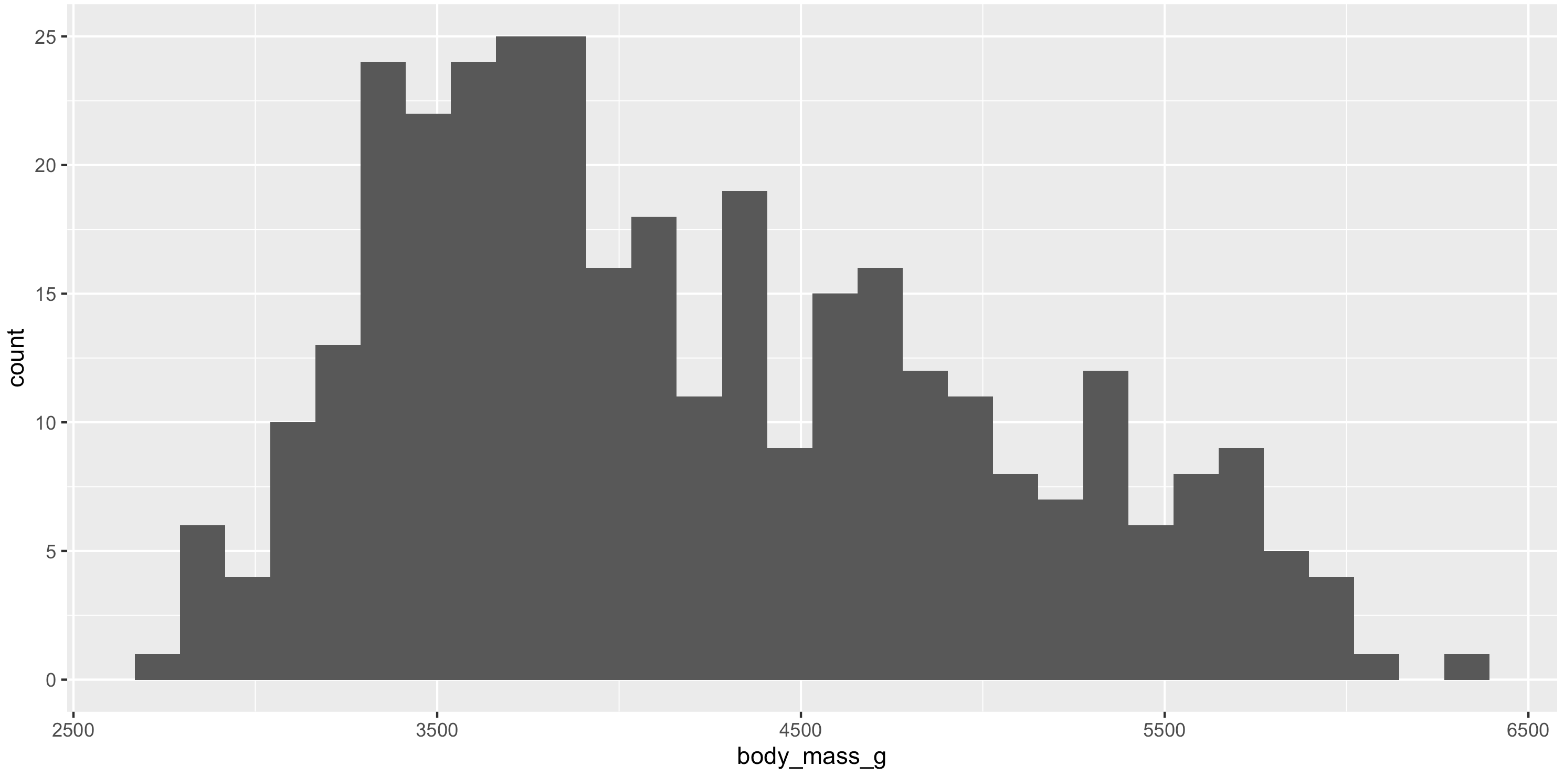


A brief plot tour of ggplot2 plots

Histograms

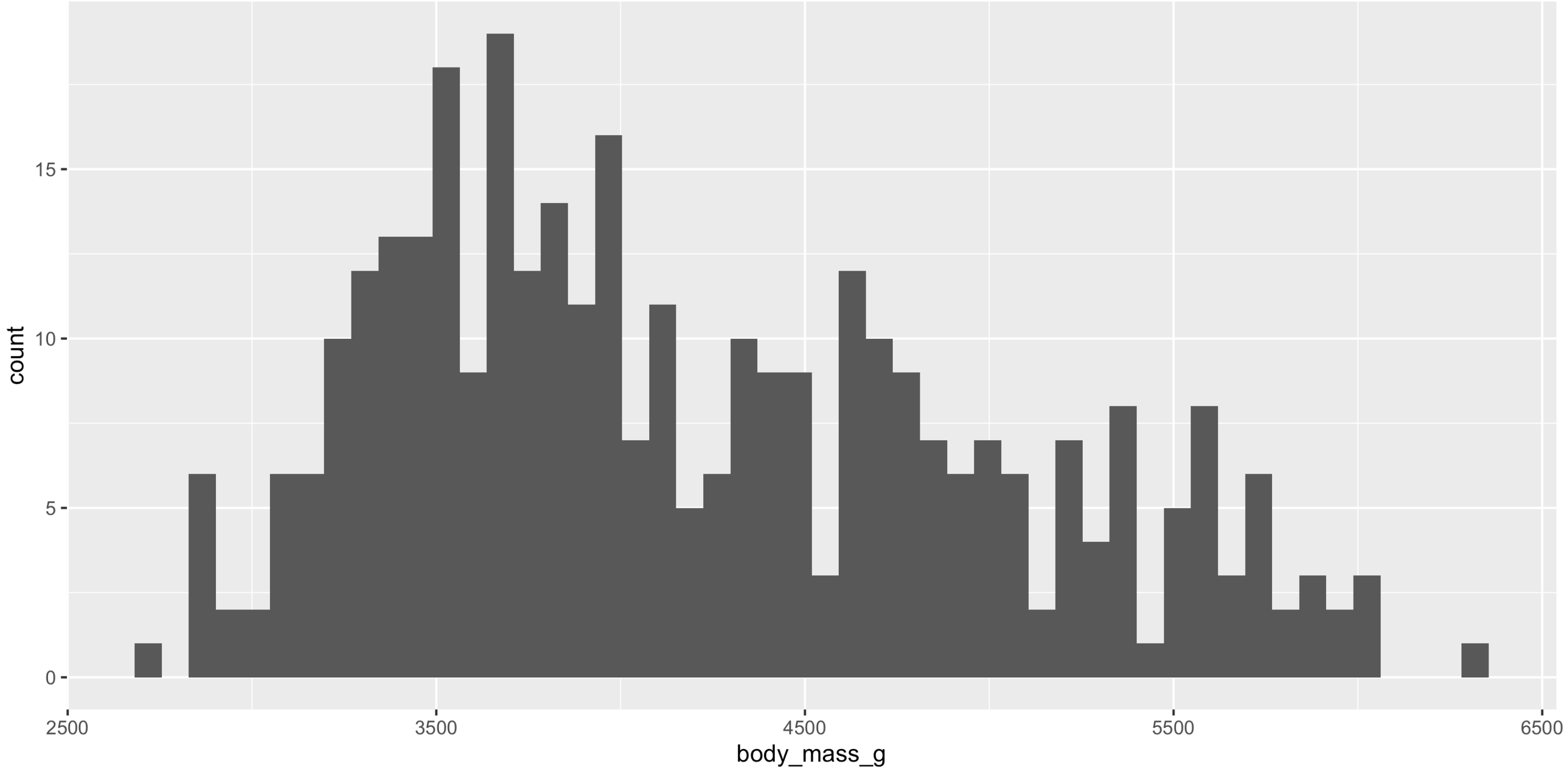
```
ggplot(
  penguins, aes(x = body_mass_g)
) +
  geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with
`binwidth`.
```



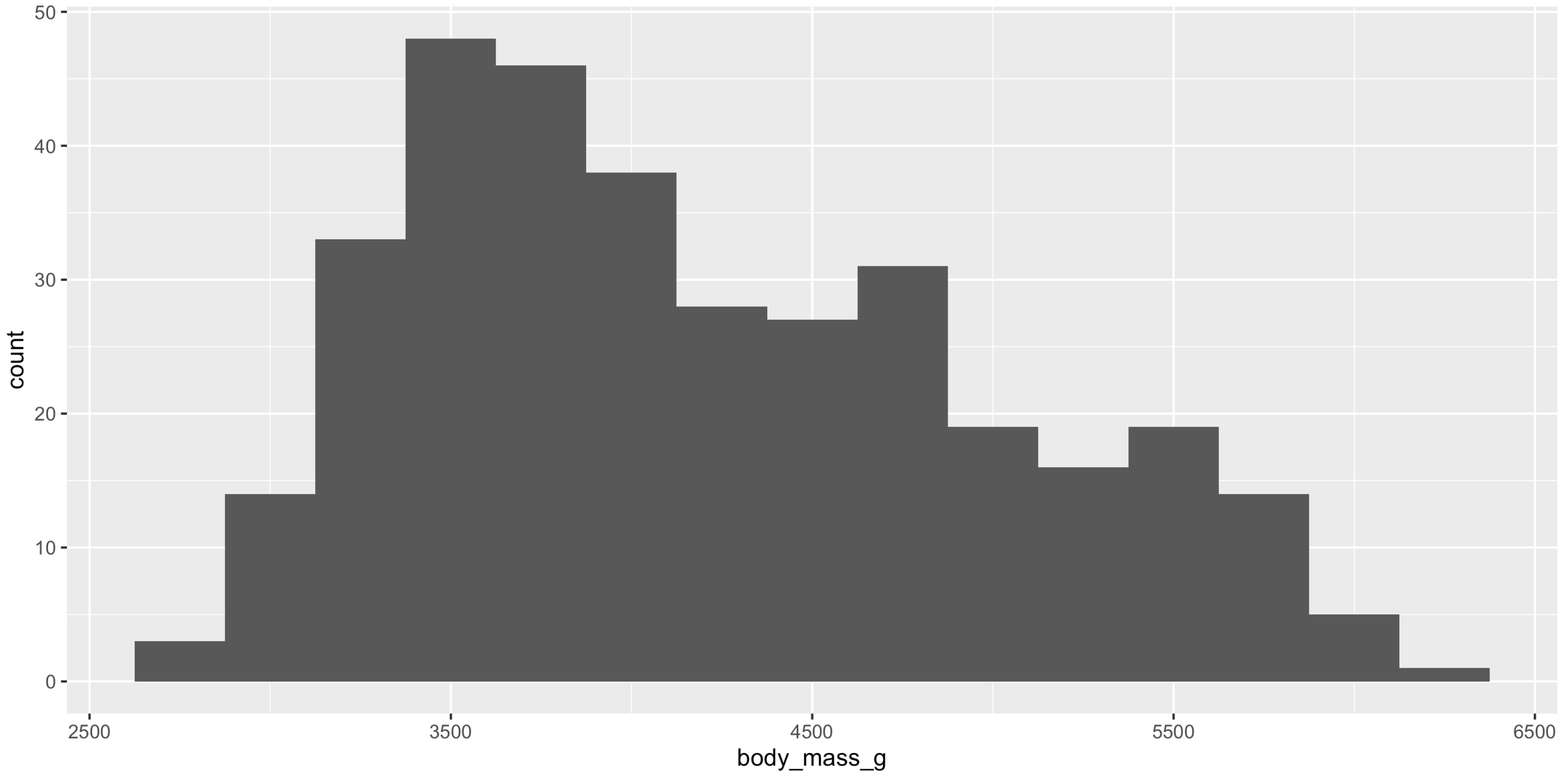
Histograms - bins

```r
ggplot(
  penguins, aes( x = body_mass_g )
) +
  geom_histogram(bins = 50)
```



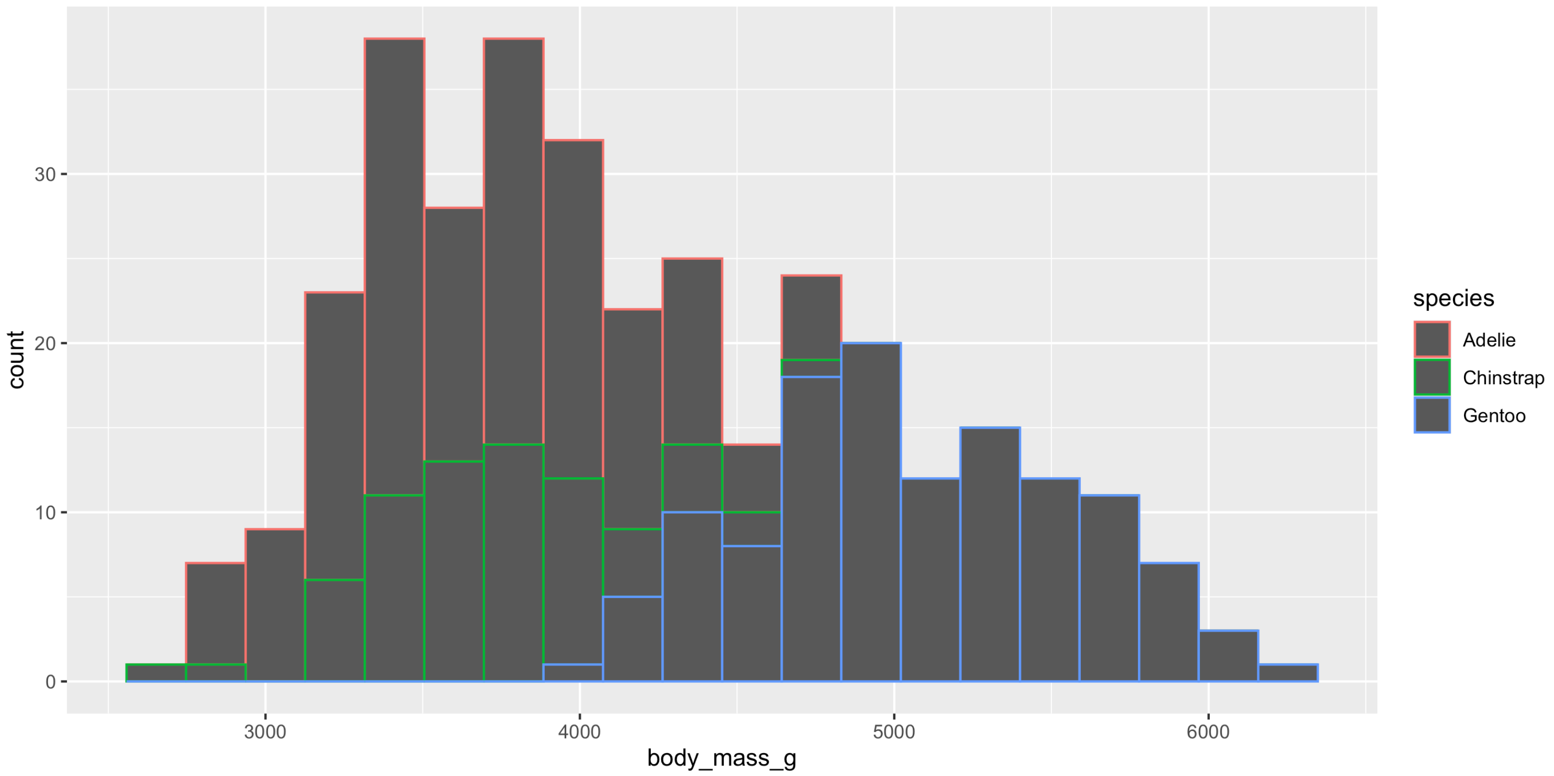
Histograms - binwidth

```
ggplot(
  penguins, aes( x = body_mass_g )
) +
  geom_histogram(binwidth = 250)
```



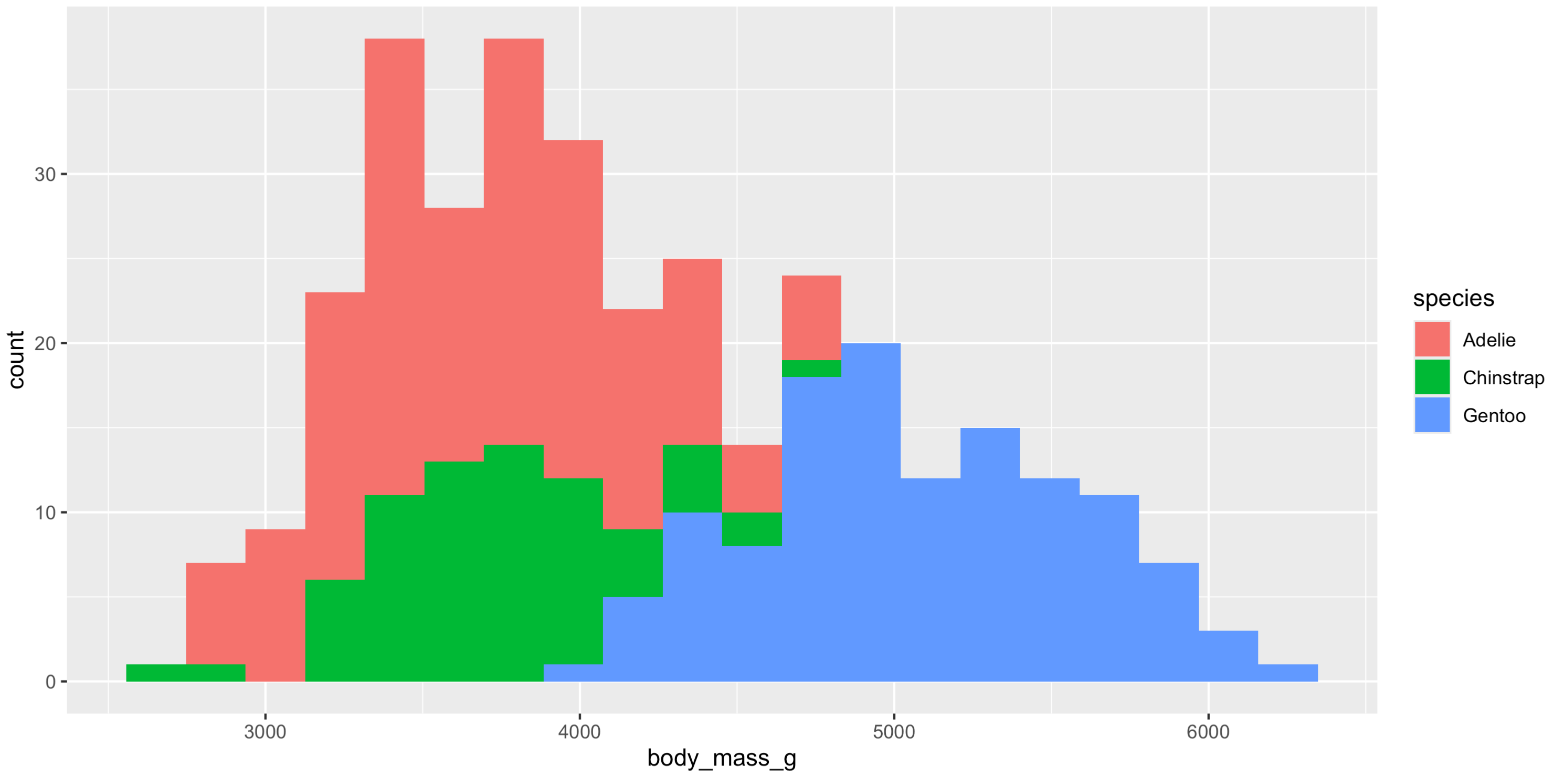
Histograms - color

```r
ggplot(
  penguins, aes(x = body_mass_g, color = species)
) +
  geom_histogram(bins = 20)
```



Histograms - fill

```r
ggplot(
  penguins, aes(x = body_mass_g, fill = species)
) +
  geom_histogram(bins = 20)
```

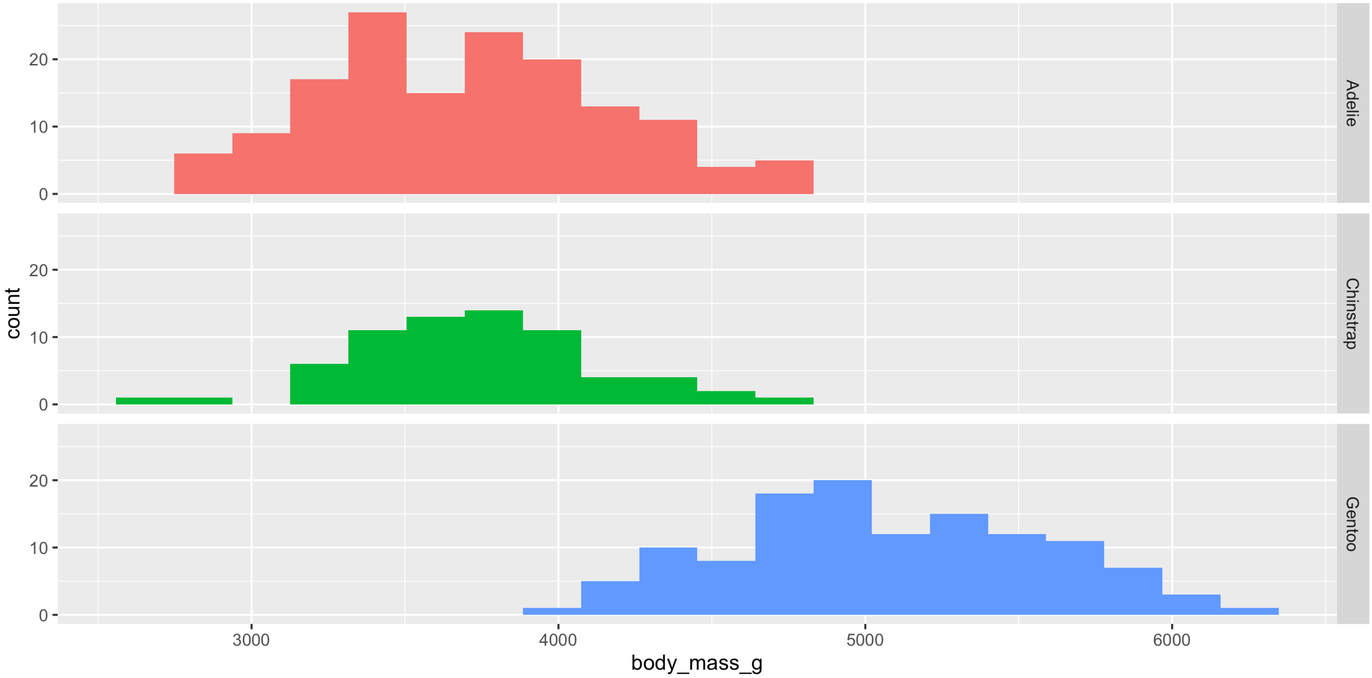


Histograms - position

```r
ggplot(
  penguins, aes(x = body_mass_g, fill = species)
) +
  geom_histogram(
    bins = 20, alpha = 0.5, position = "identity"
  )
```

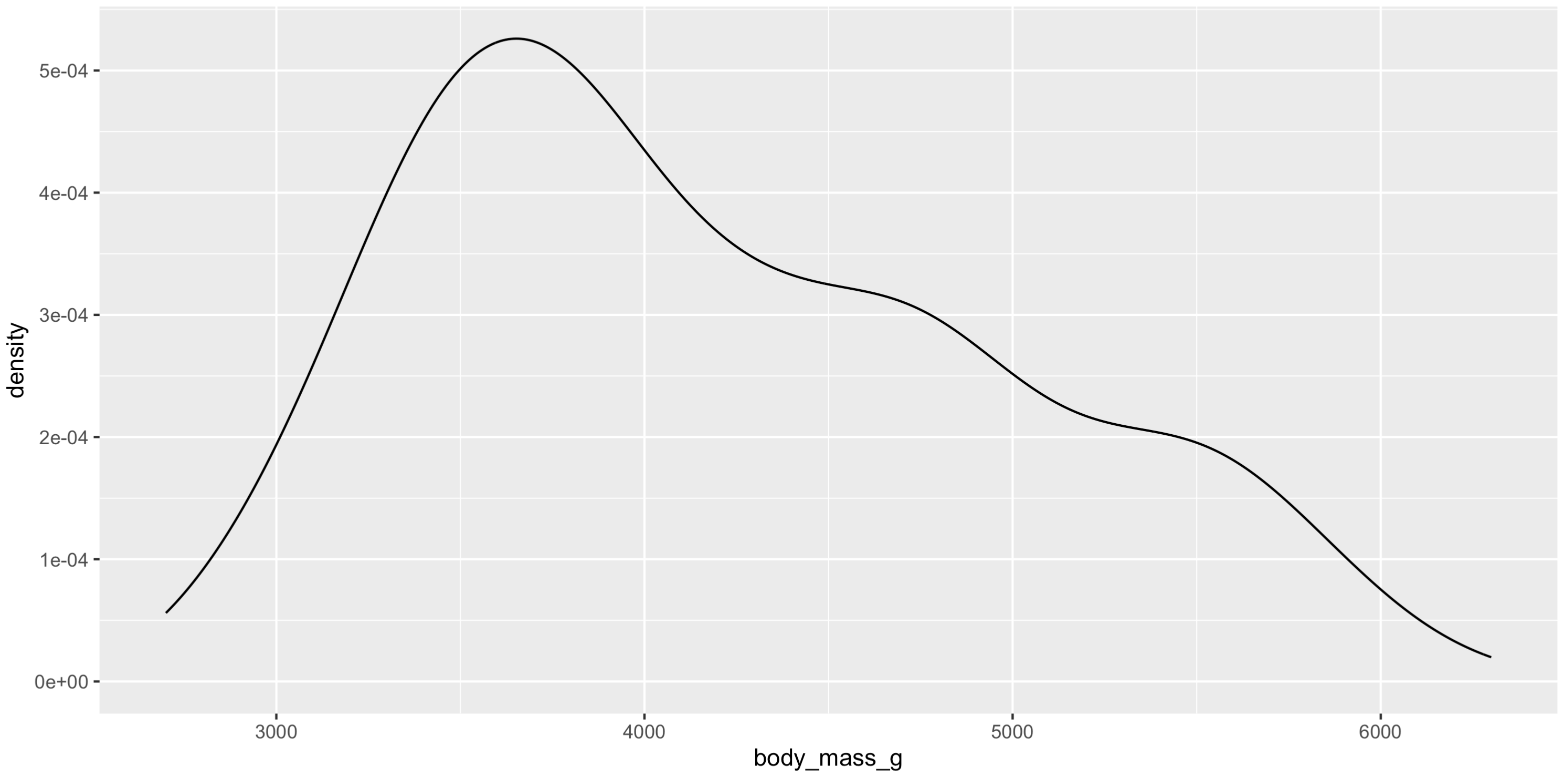
Histograms - facets

```r
ggplot(
  penguins, aes(x = body_mass_g, fill = species)
) +
  geom_histogram(bins = 20) +
  facet_grid(species ~ .) +
  guides(fill = "none")
```



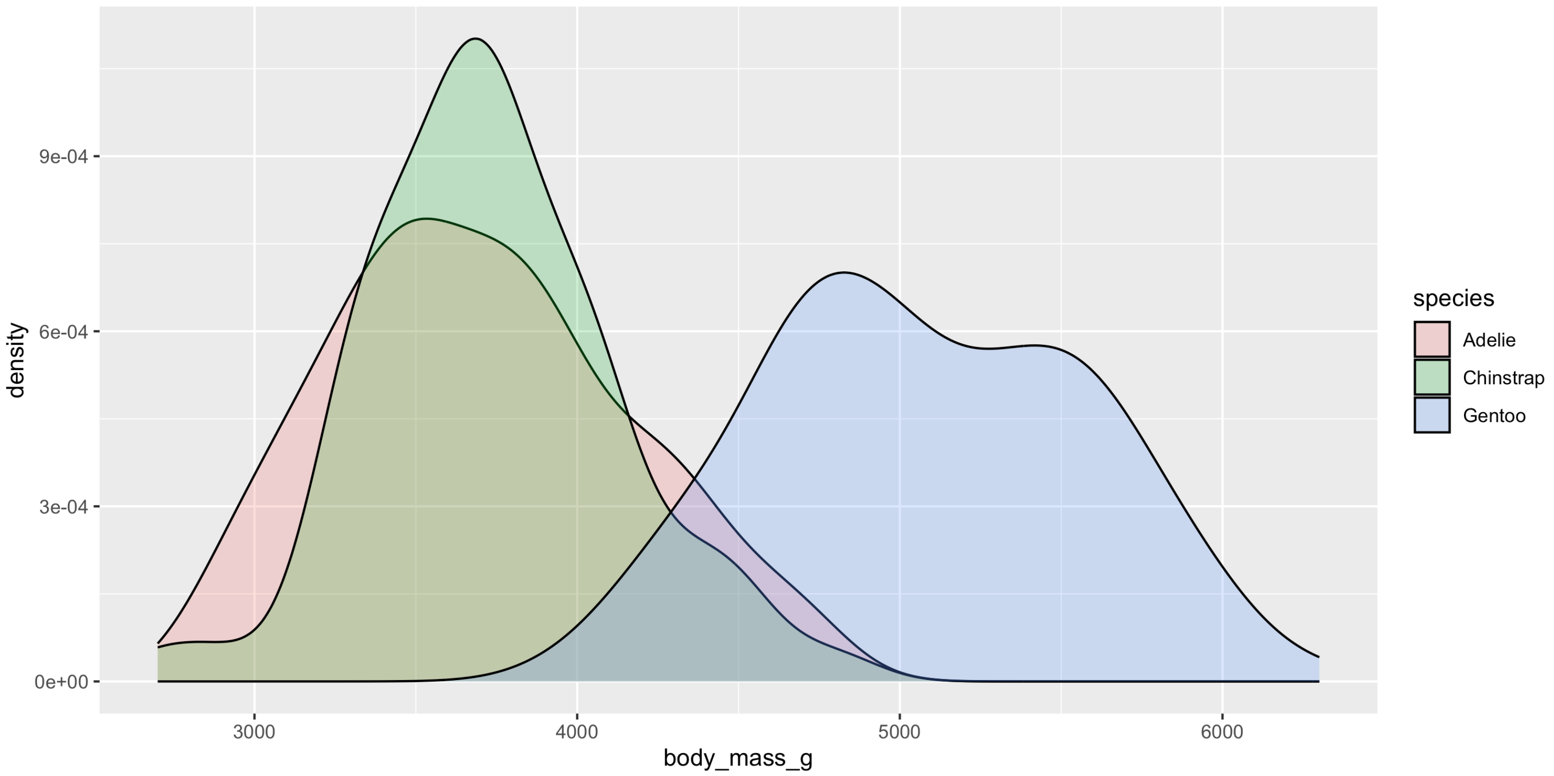
Density plot

```r
ggplot(
  penguins, aes(x = body_mass_g)
) +
  geom_density()
```



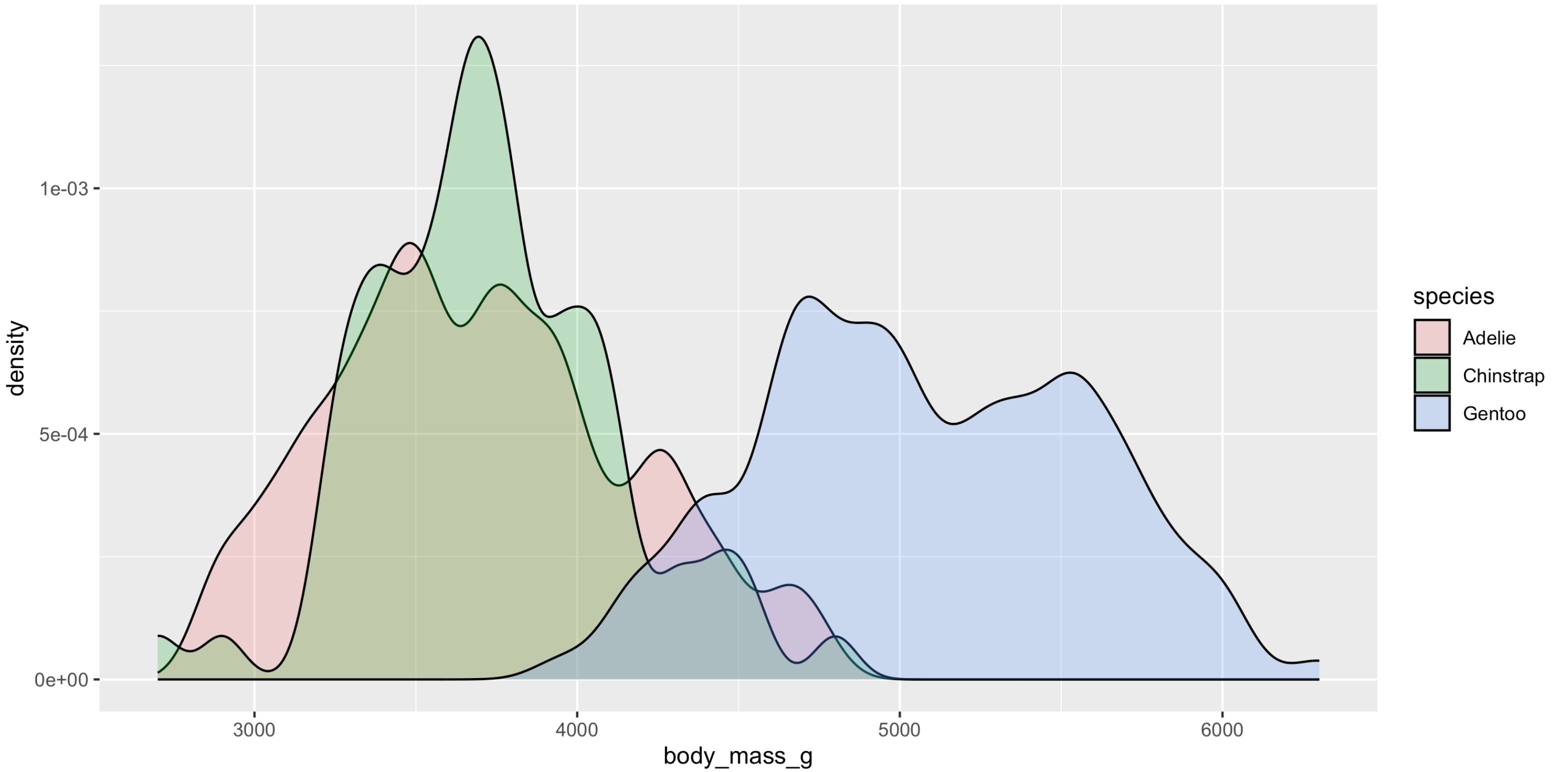
Density plot - fill

```r
ggplot(
  penguins, aes(x = body_mass_g, fill = species)
) +
  geom_density(alpha = 0.25)
```



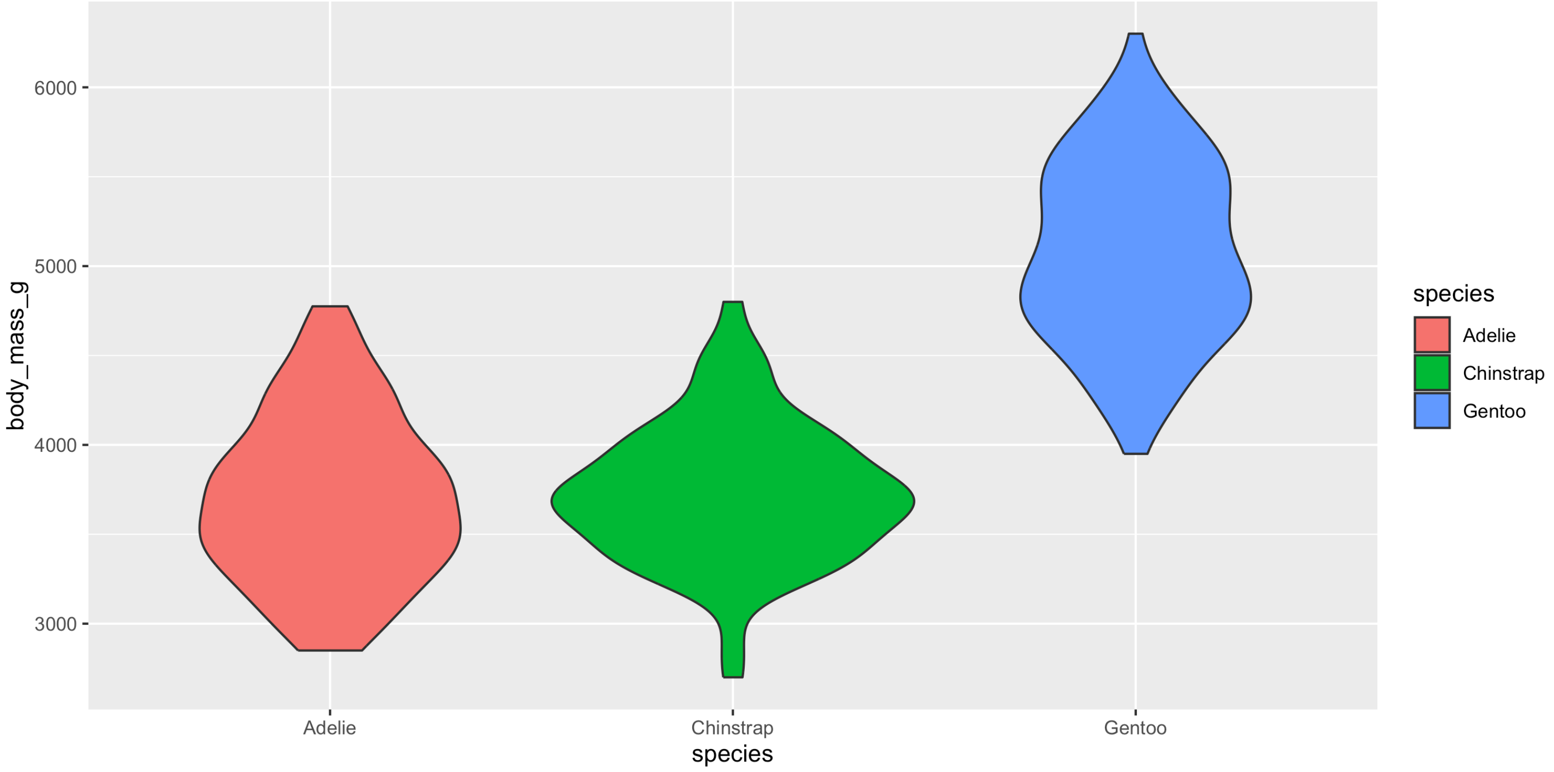
Density plot - adjust

```
ggplot(
  penguins, aes(x = body_mass_g, fill = species)
) +
  geom_density(
    adjust = 0.5, alpha = 0.25
  )
```



Violin plot

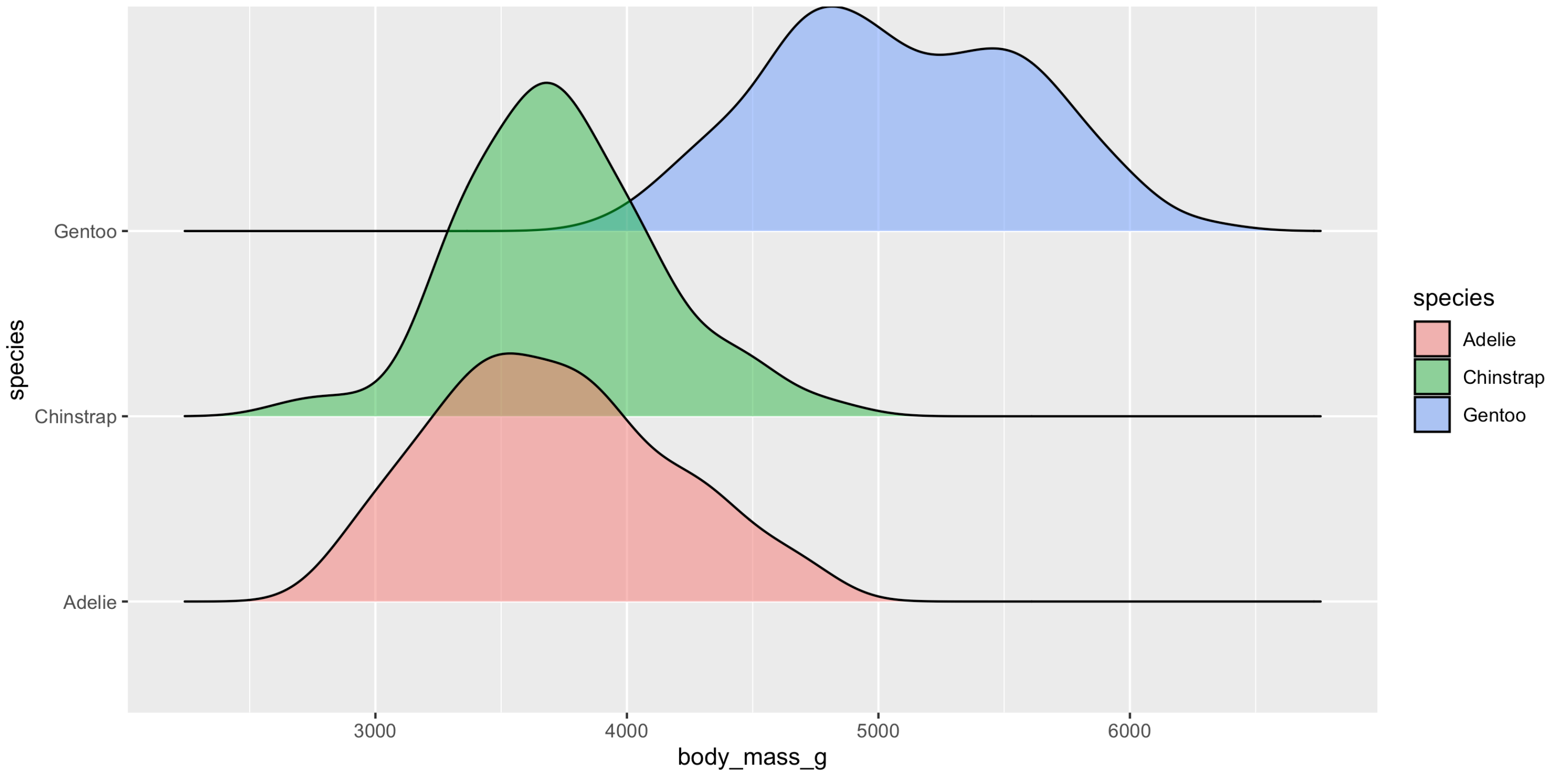
```r
ggplot(
  penguins, aes(x = species, fill = species, y = body_mass_g )
) +
  geom_violin()
```



Ridge plot

```r
ggplot(
  penguins, aes(x = body_mass_g, y = species, fill = species)
) +
  ggridges::geom_density_ridges(alpha = 0.5)
```

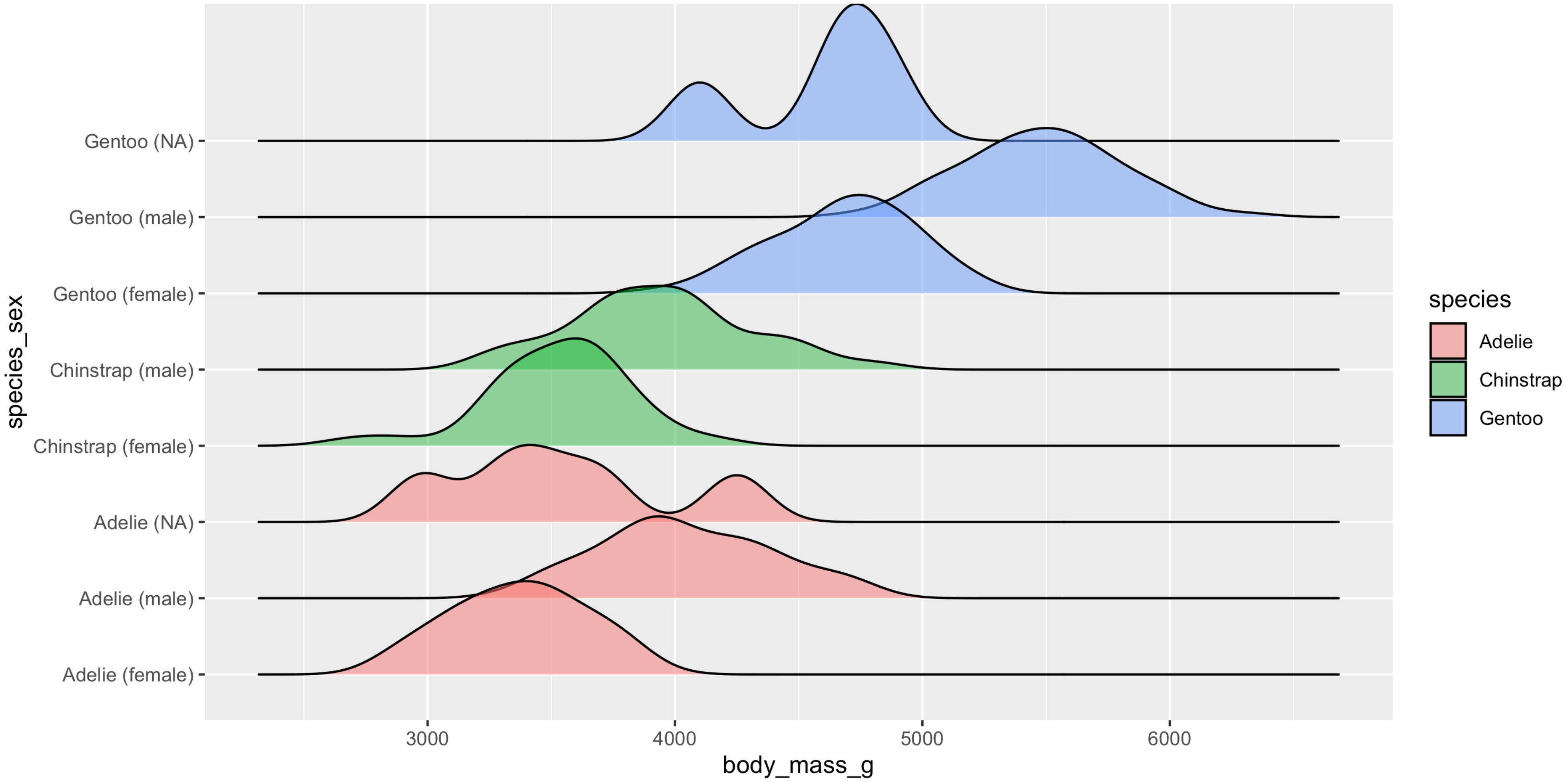
```
Picking joint bandwidth of 153
```



Ridge plot - more categories + dplyr

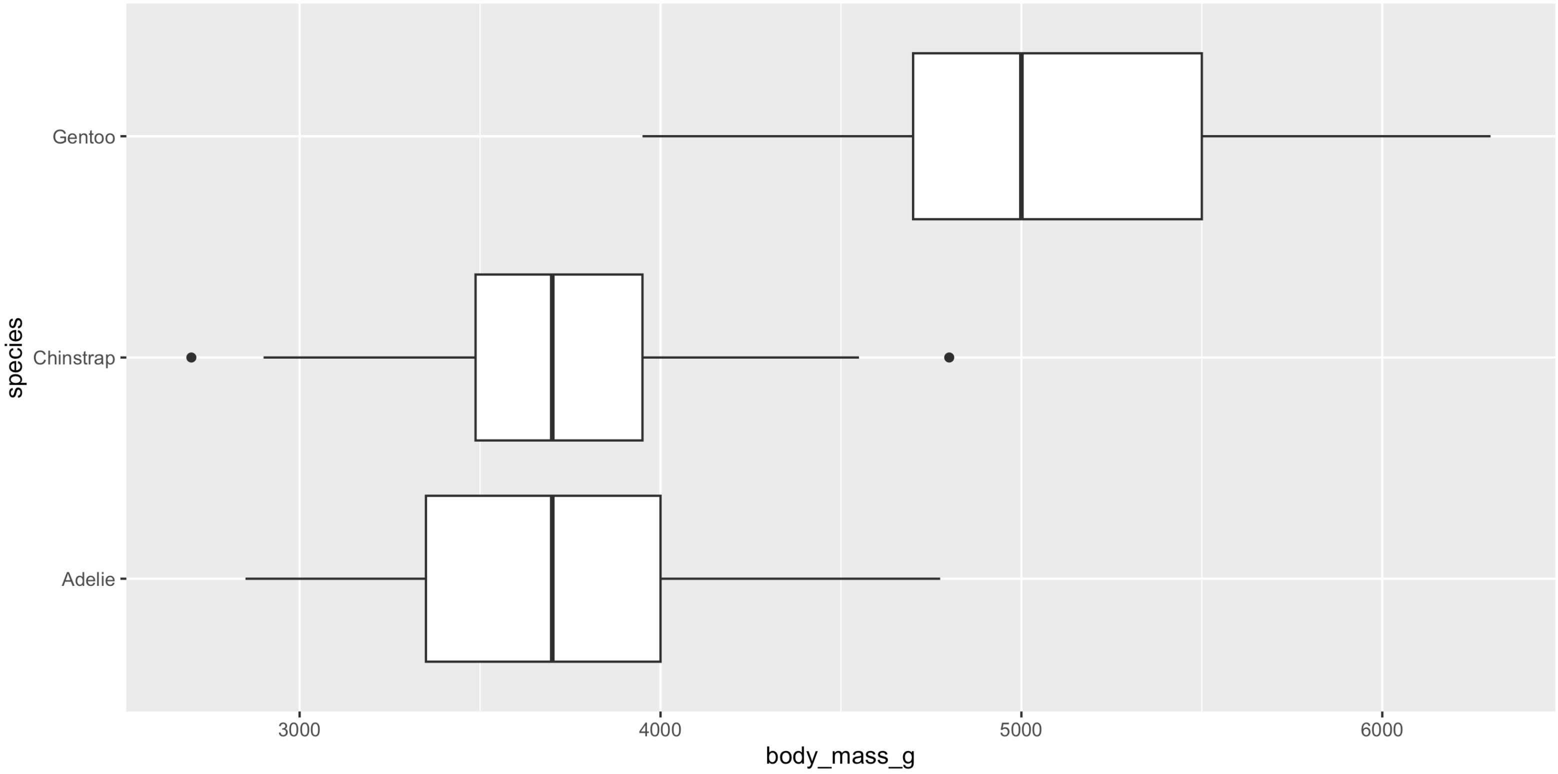
```r
penguins |>
  mutate( species_sex = paste0( species, " (", sex, ")") ) |>
  ggplot( aes(x = body_mass_g, y = species_sex, fill = species) ) +
  ggridges::geom_density_ridges( alpha = 0.5 )
```

```
Picking joint bandwidth of 127
```


species_sex
Gentoo (NA)
Gentoo (male)
Gentoo (female)
Chinstrap (male)
Chinstrap (female)
Adelie (NA)
Adelie (male)
Adelie (female)
3000
4000
5000
6000
body_mass_g
species
Adelie
Chinstrap
Gentoo

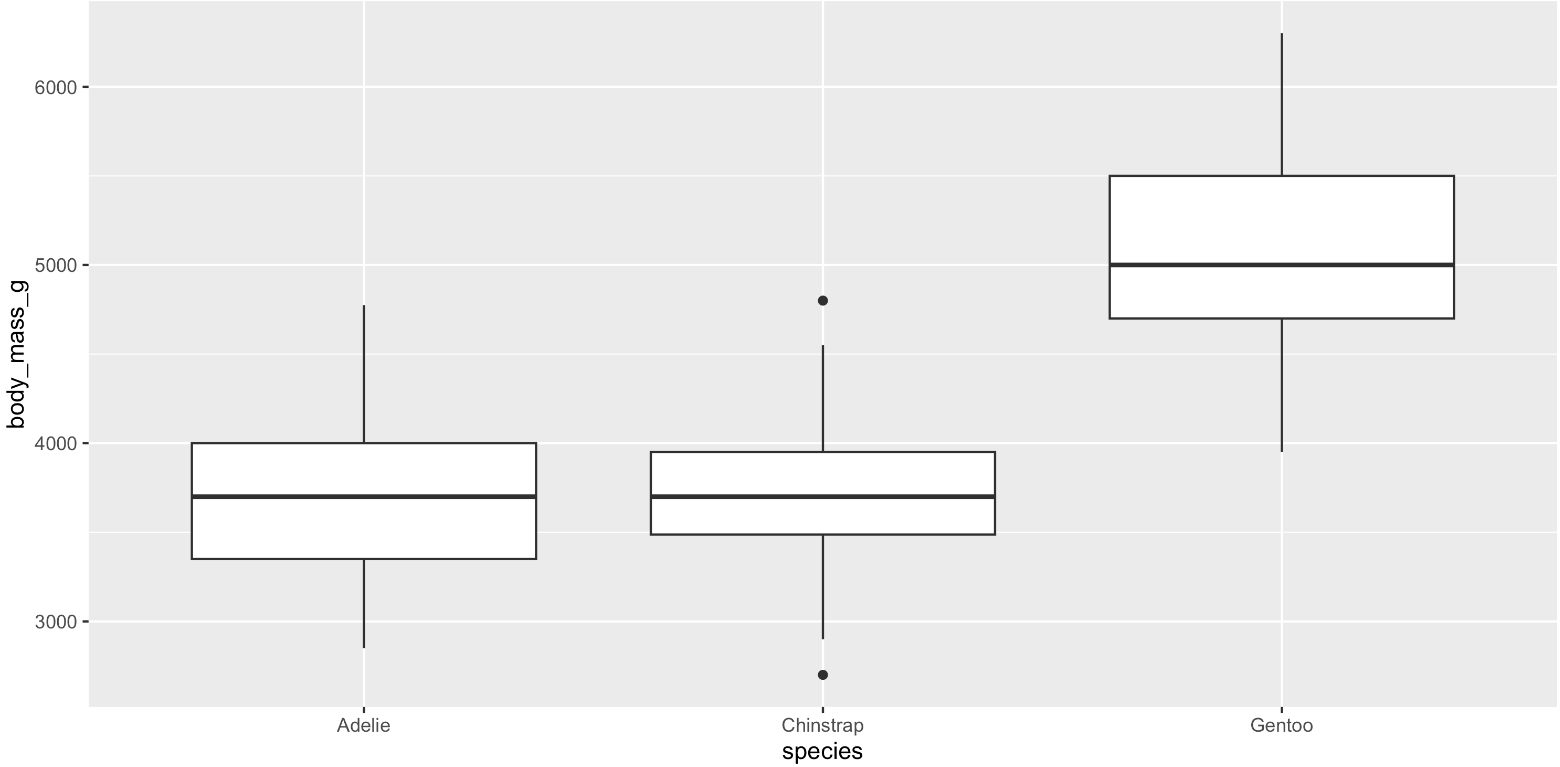
Box plot

```r
ggplot(
  penguins, aes(x = body_mass_g, y = species)
) +
  geom_boxplot()
```



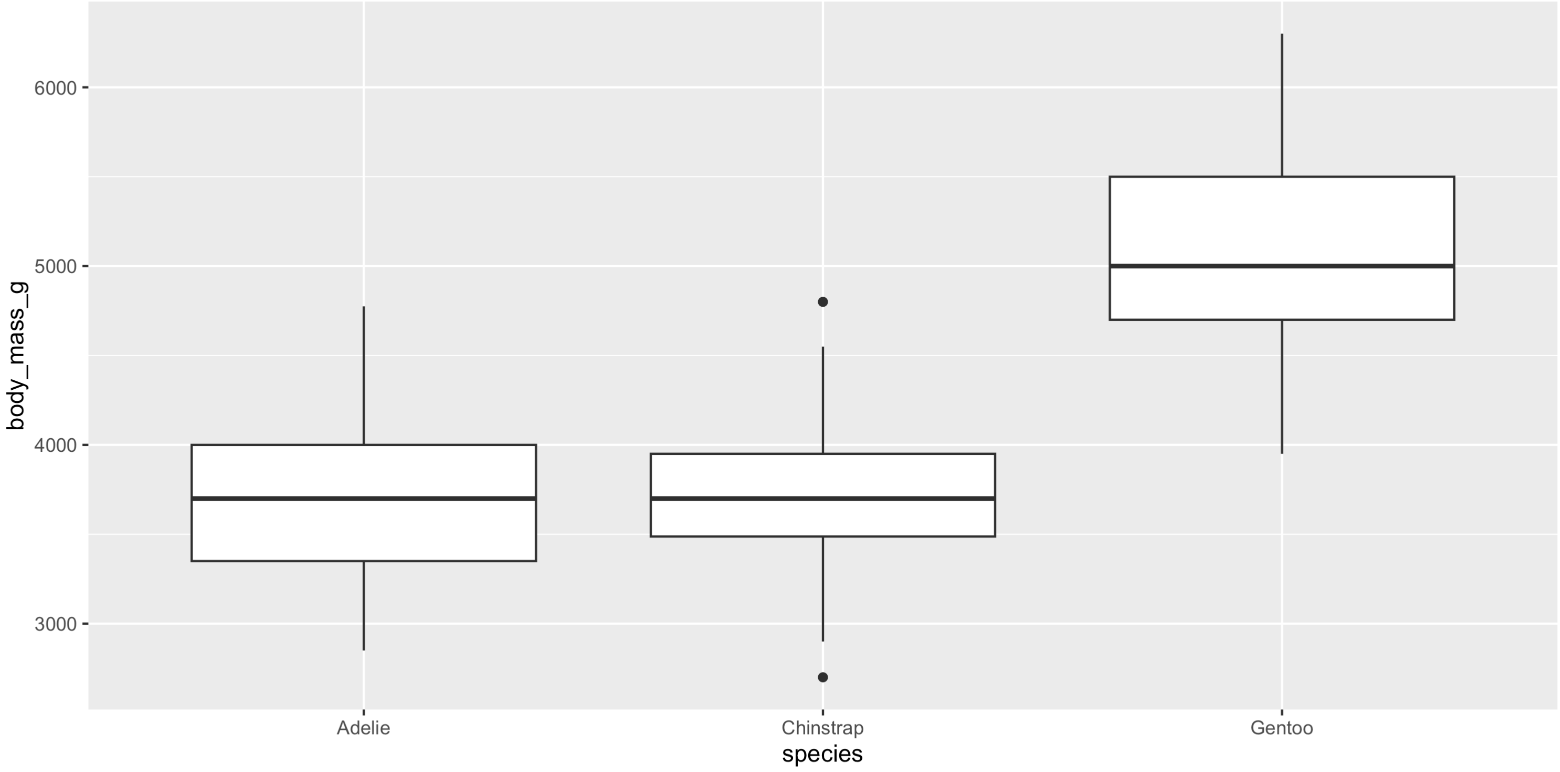
Box plot - coord_flip

```r
ggplot(
  penguins, aes(x = body_mass_g, y = species)
) +
  geom_boxplot() +
  coord_flip()
```



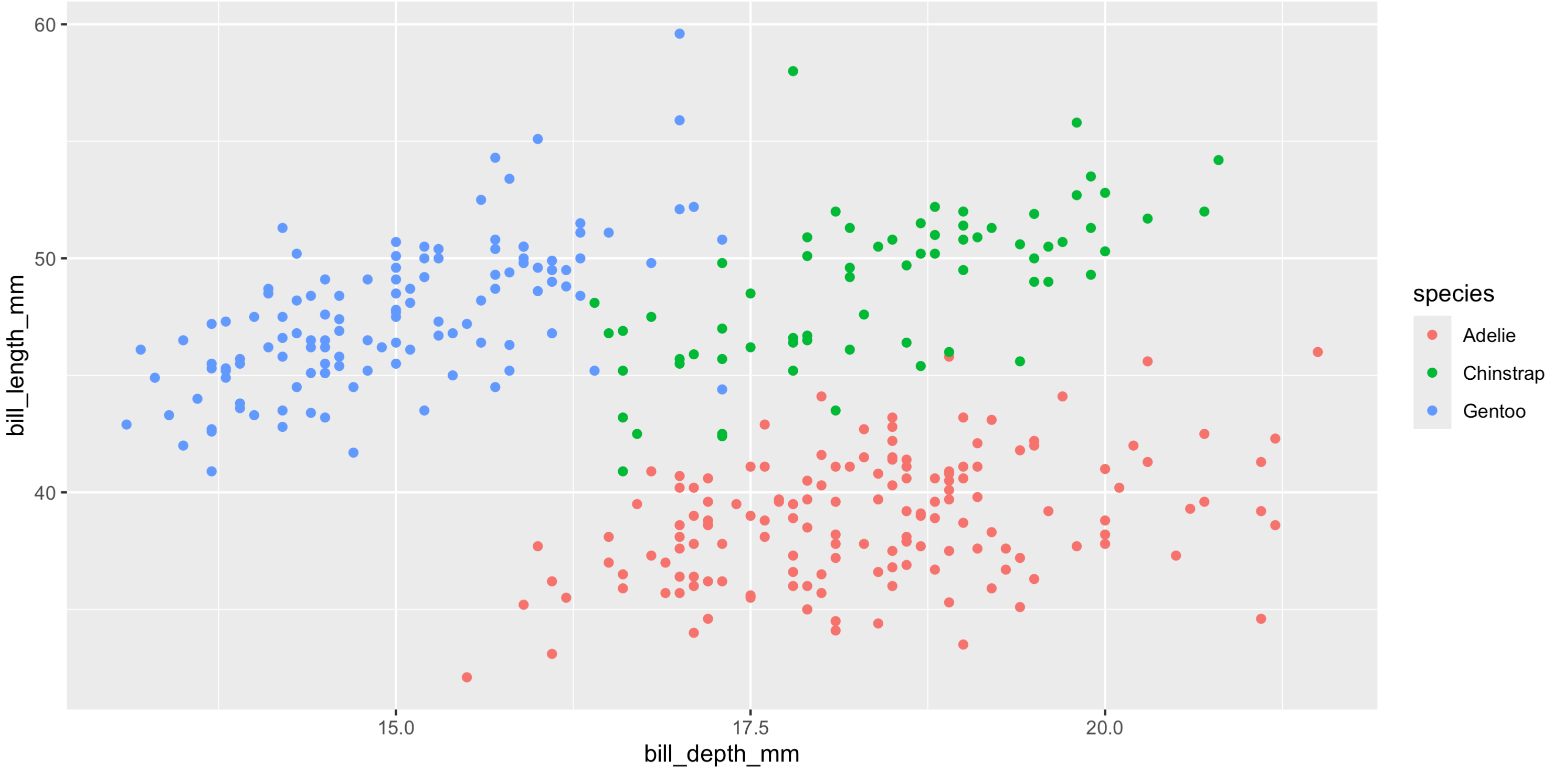
Box plot - swap coords

```r
ggplot(
  penguins, aes(x = species, y = body_mass_g)
) +
  geom_boxplot()
```



Scatter plot

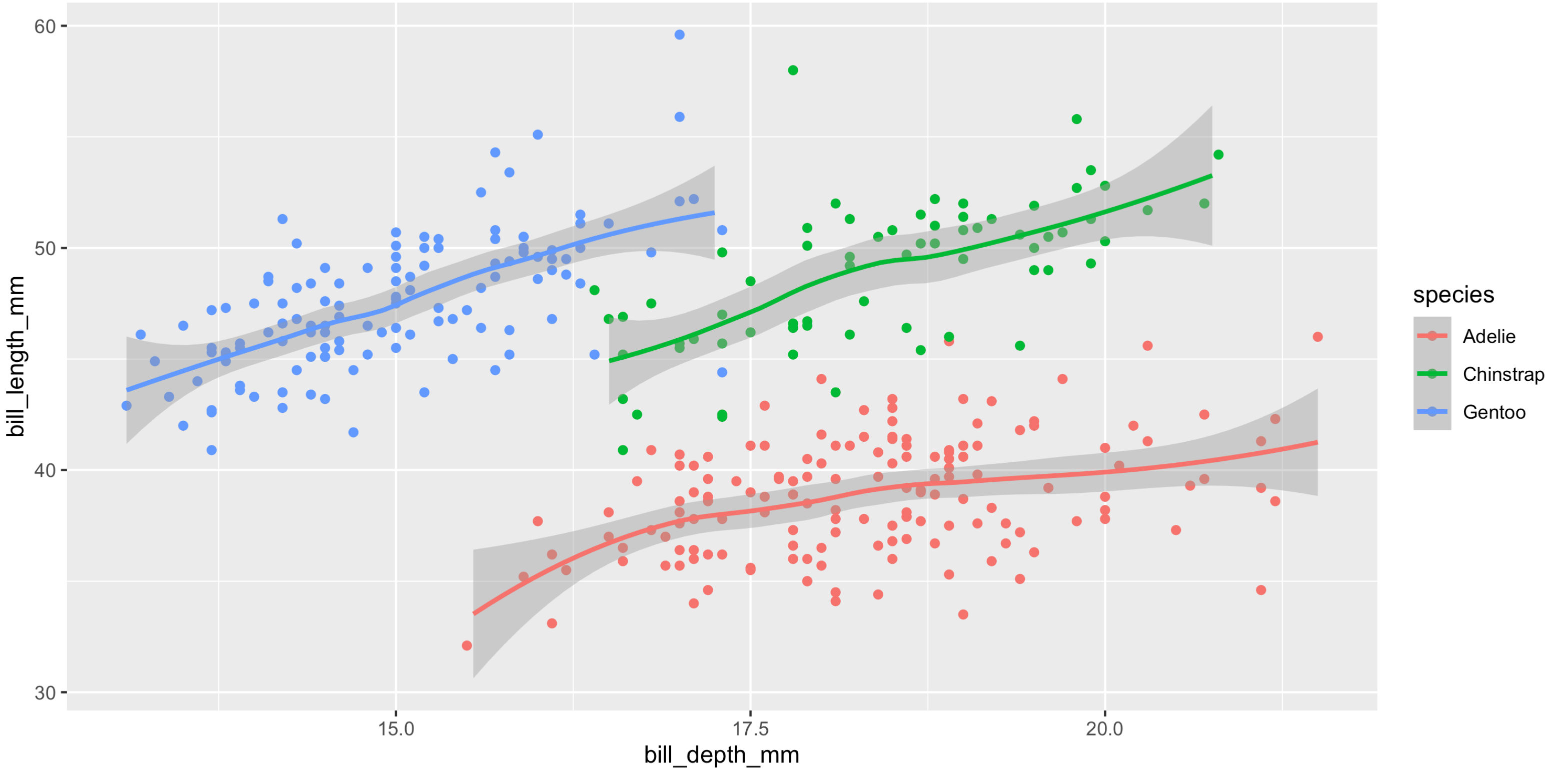
```
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm, color = species)
) +
  geom_point()
```



Scatter plot - geom_smooth

```r
ggplot(
  penguins,
  aes(x = bill_depth_mm, y = bill_length_mm, color = species)
) +
  geom_point() +
  geom_smooth( fullrange = TRUE )
```

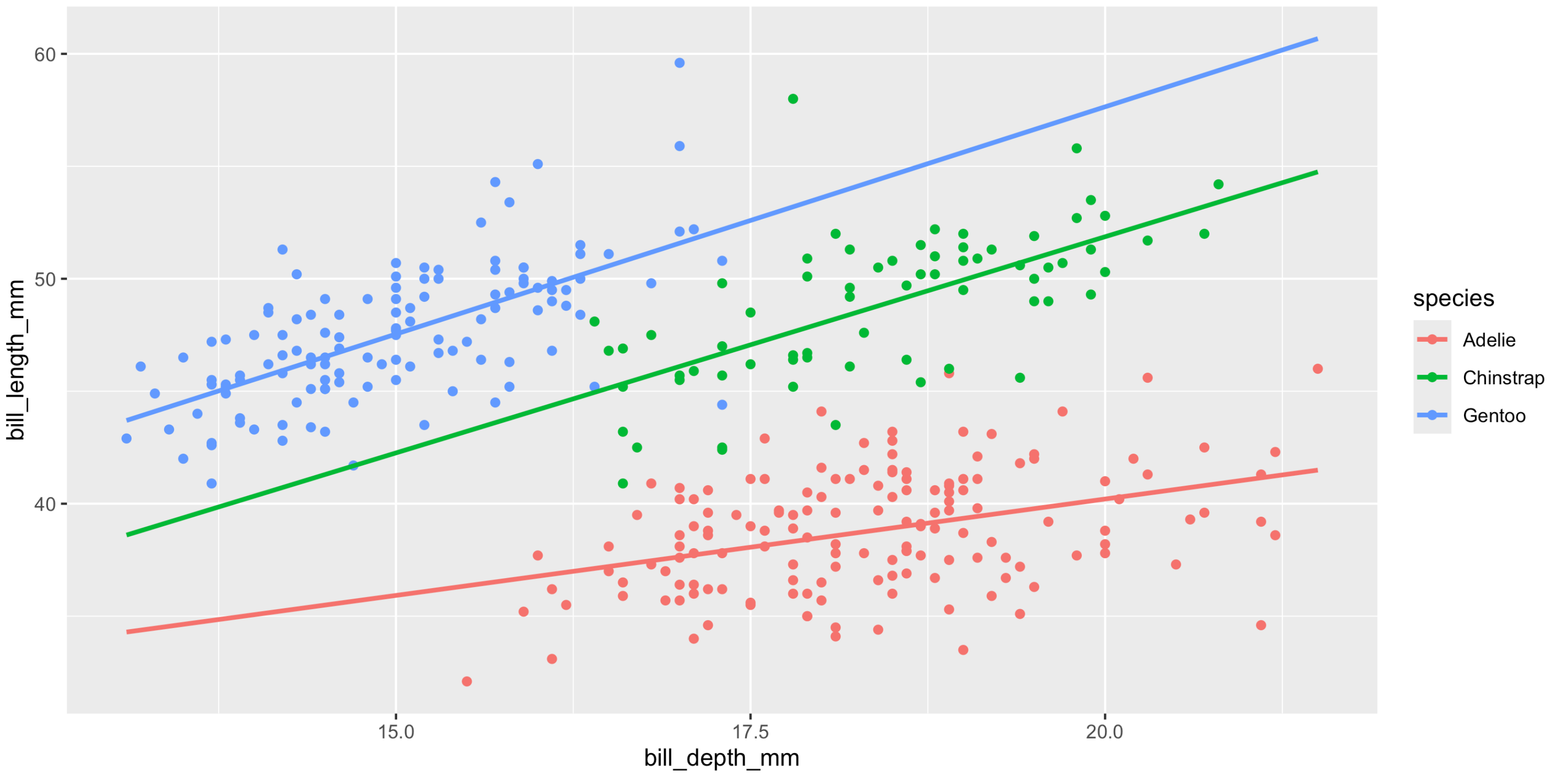
```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Scatter plot - geom_smooth w/ lm

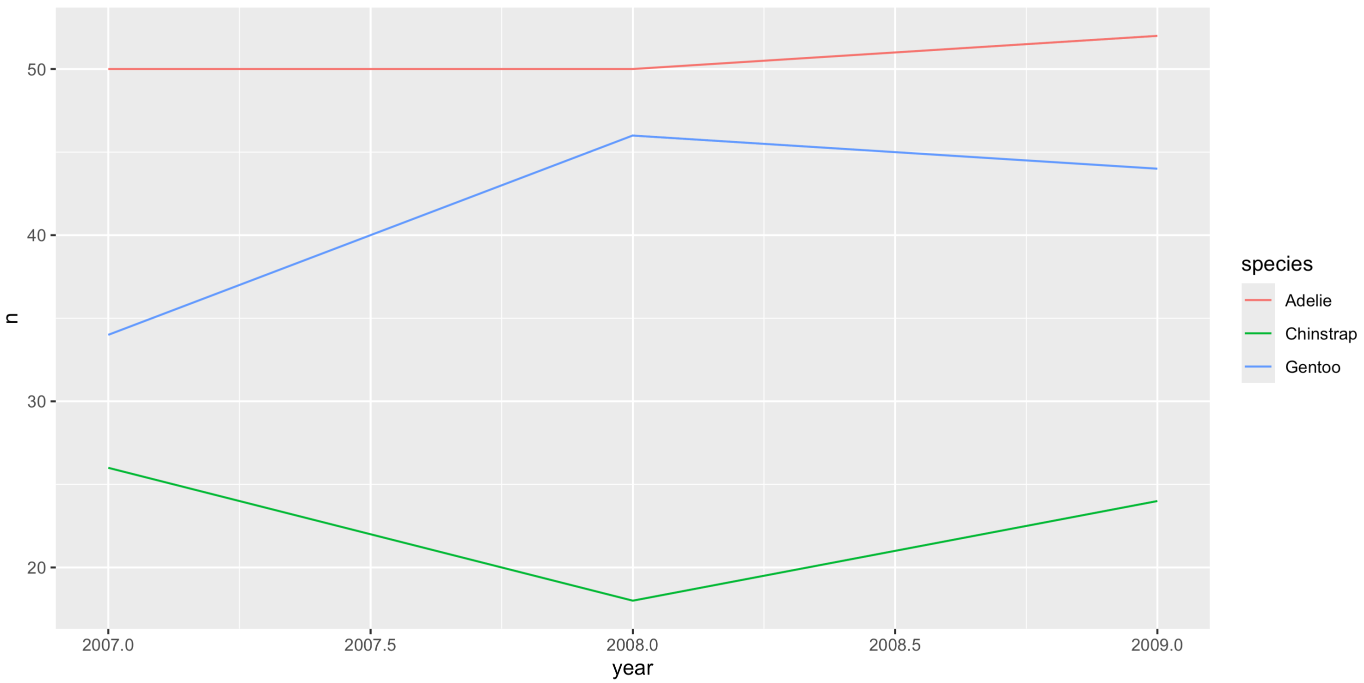
```r
ggplot(
  penguins, aes(x = bill_depth_mm, y = bill_length_mm, color = species)
) +
  geom_point() +
  geom_smooth(
    method = "lm", se = FALSE, fullrange = TRUE
  )
```

```
`geom_smooth()` using formula = 'y ~ x'
```


60
50
40
bill_length_mm
15.0
17.5
20.0
bill_depth_mm
species
Adelie
Chinstrap
Gentoo

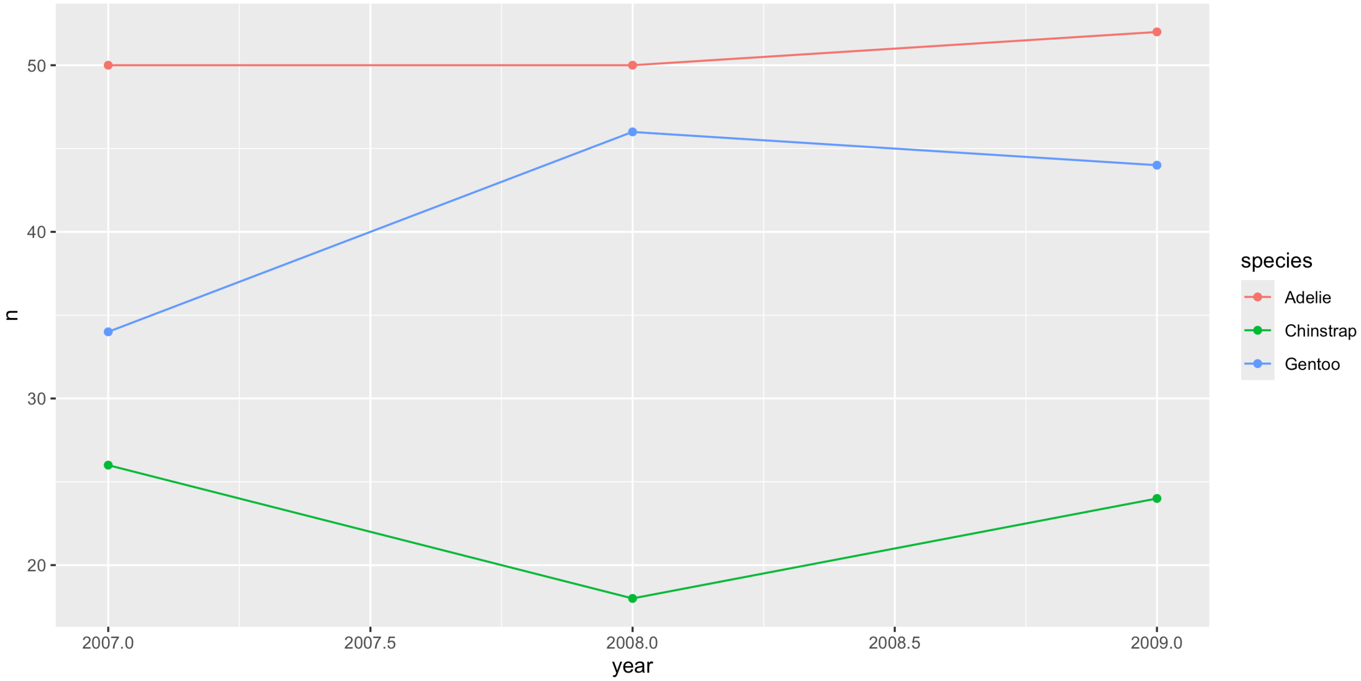
Line plot

```r
penguins |>
  count(species, year) |>
  ggplot(
    aes(x = year, y = n, color = species, group = species)
  ) +
  geom_line()
```



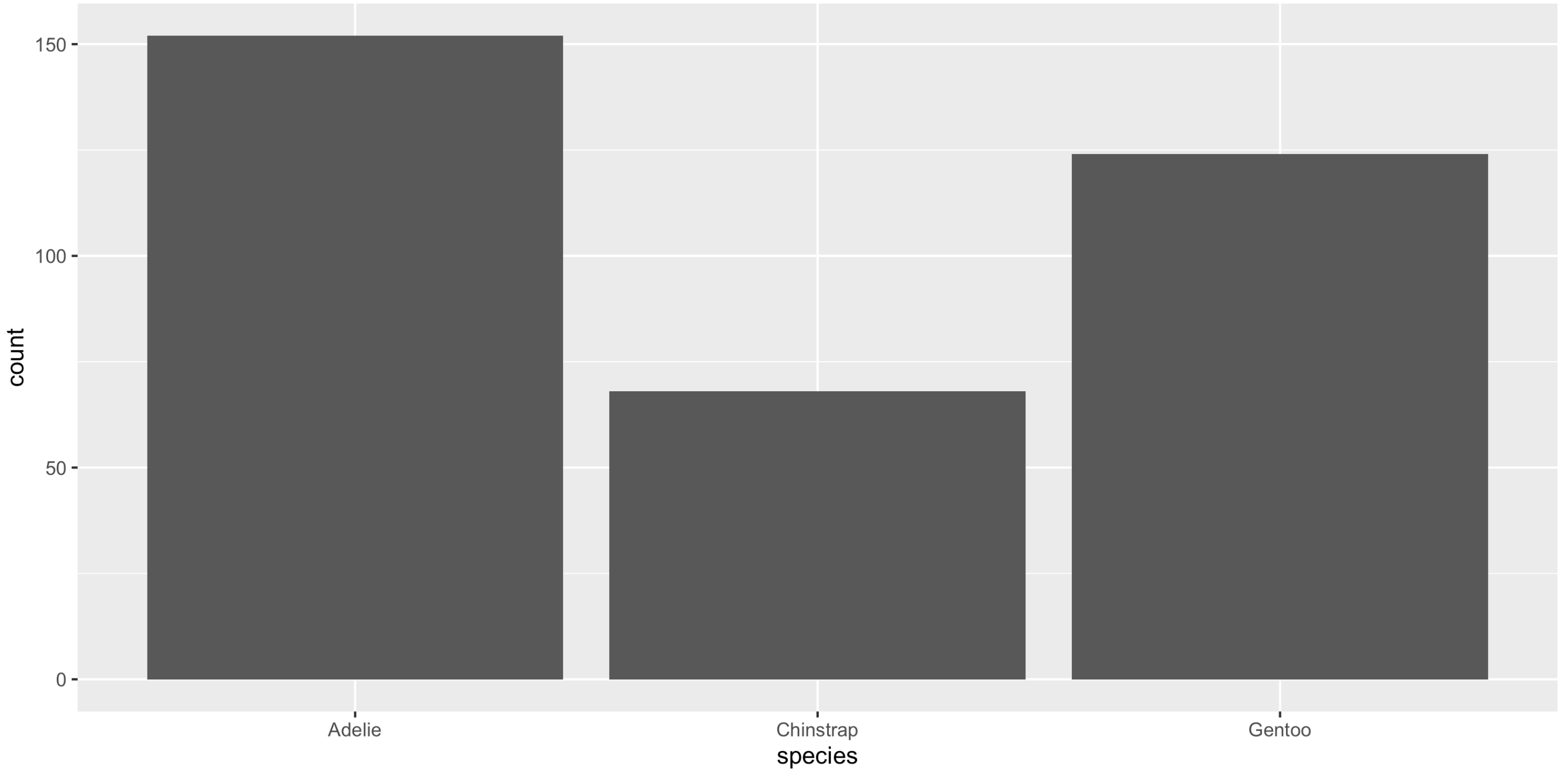
Line plot - with points

```r
penguins |>
  count(species, year) |>
  ggplot(
    aes(x = year, y = n, color = species, group = species)
  ) +
  geom_line() +
  geom_point()
```



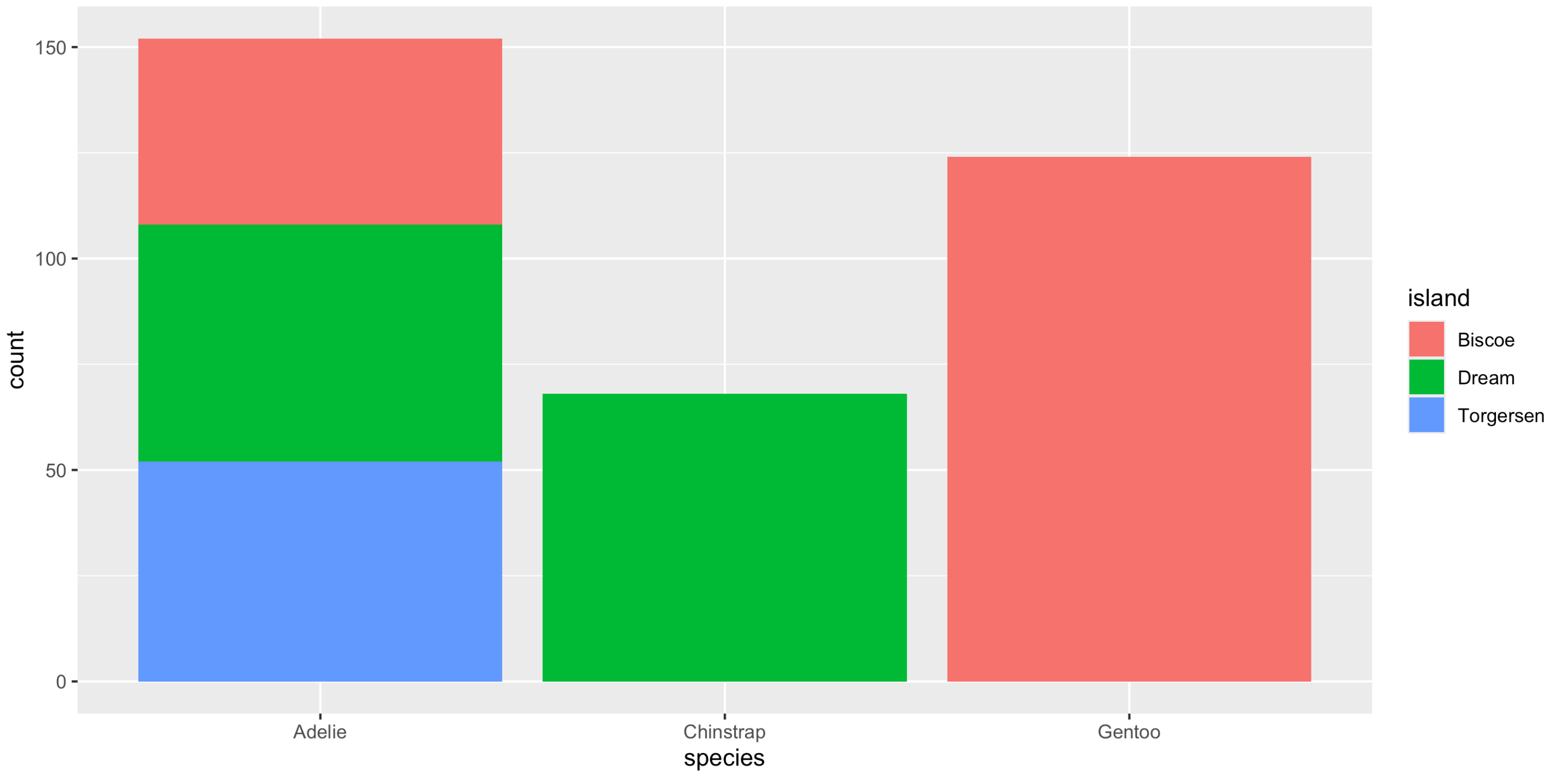
Bar plot

```r
ggplot(
  penguins, aes(x = species)
) +
  geom_bar()
```



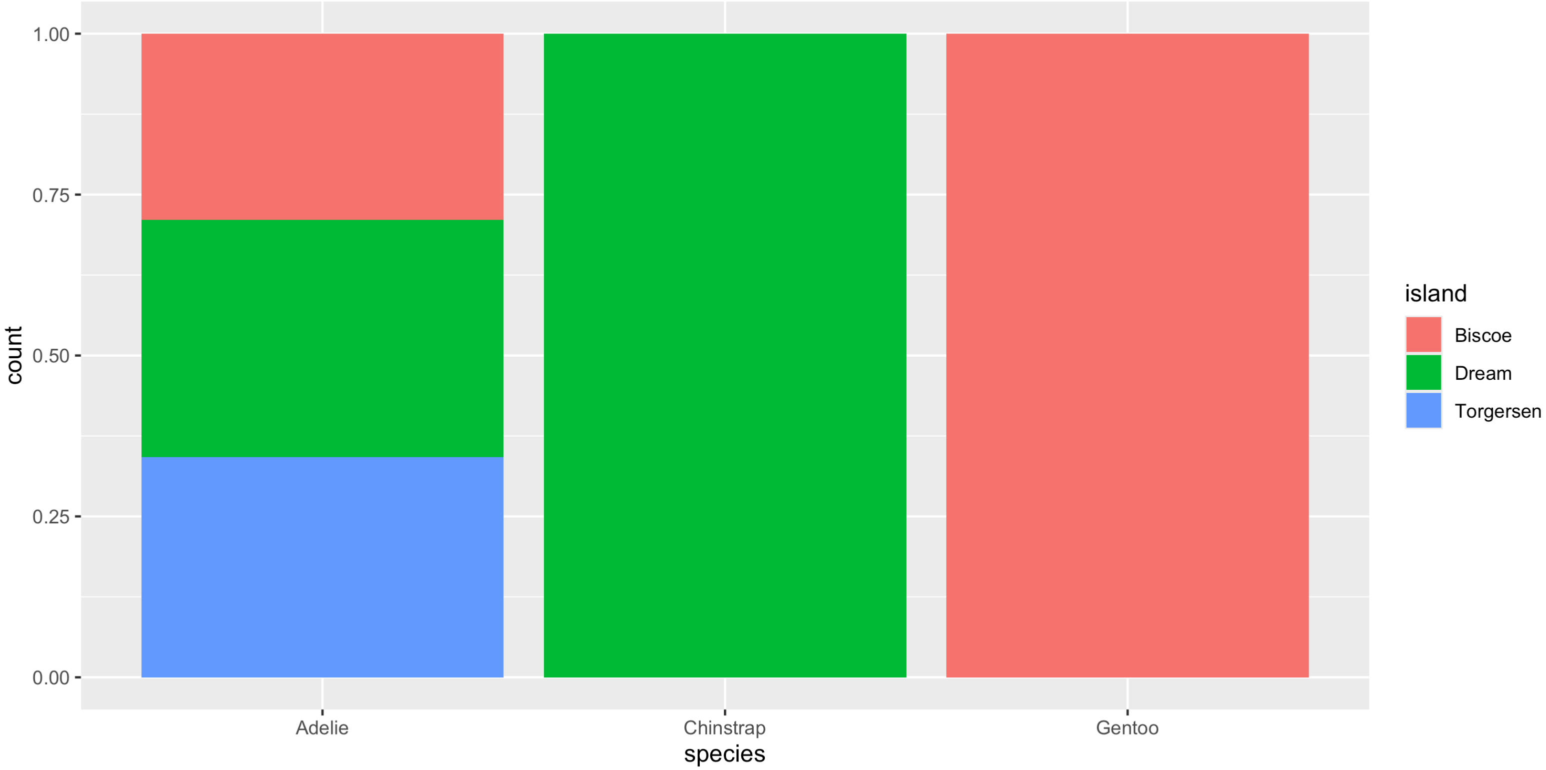
Stacked bar plot

```r
ggplot(
  penguins, aes(x = species, fill = island)
) +
  geom_bar()
```



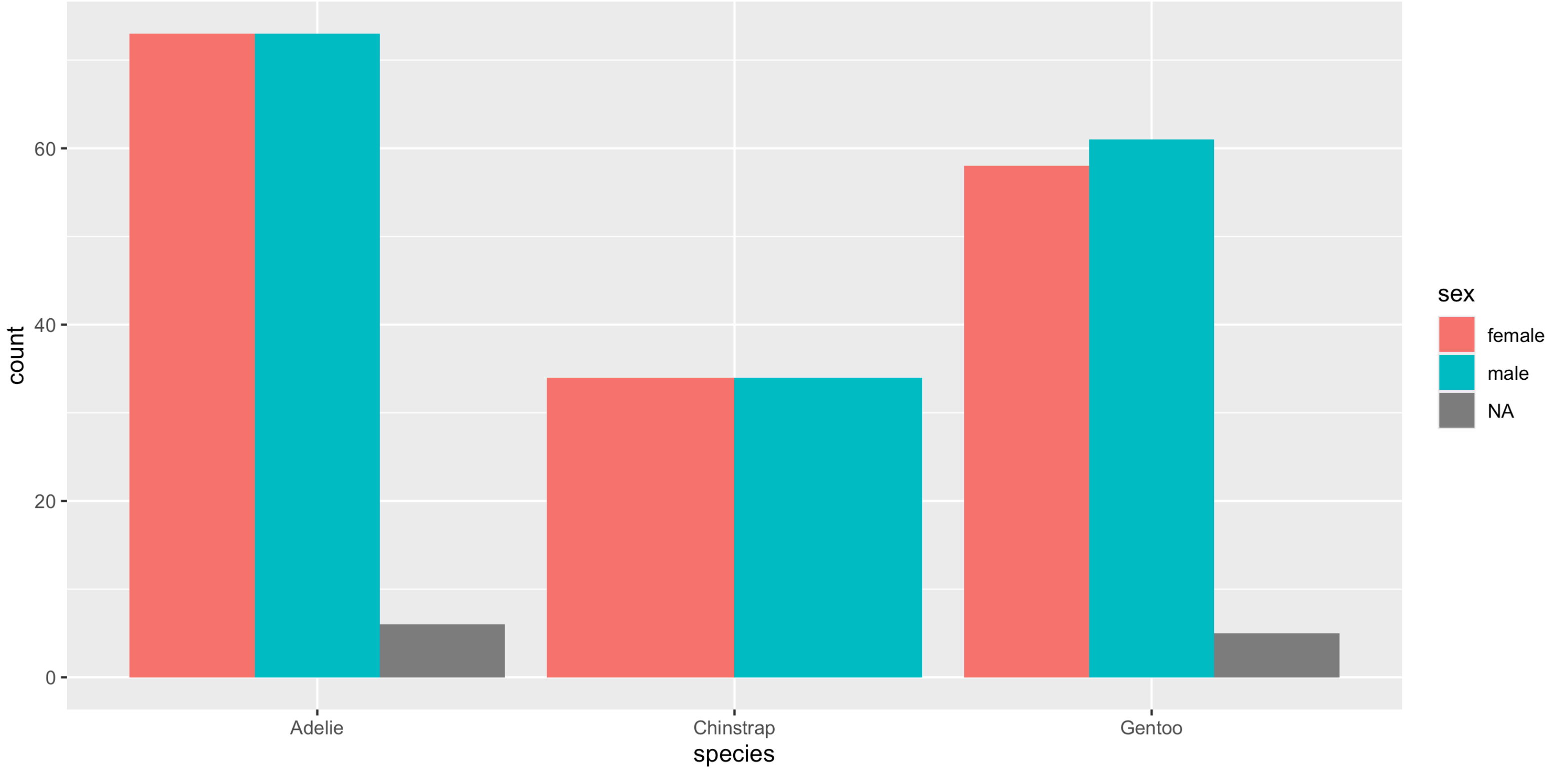
Stacked relative frequency bar plot

```
ggplot(
  penguins, aes(x = species, fill = island)
) +
  geom_bar(position = "fill")
```



Dodged bar plot

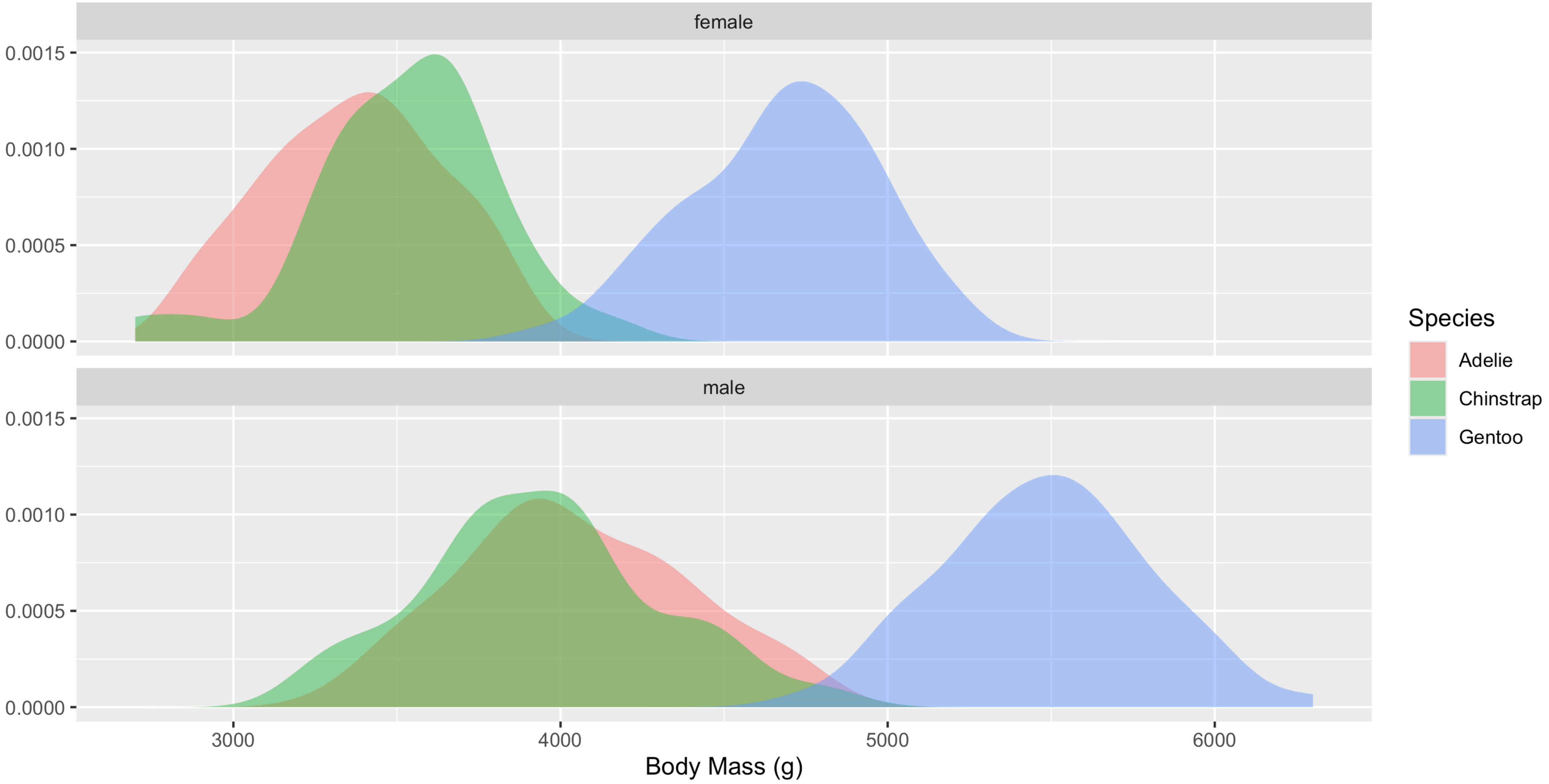
```r
ggplot(
  penguins, aes(x = species,fill = sex)
) +
  geom_bar(position = "dodge")
```



Exercises

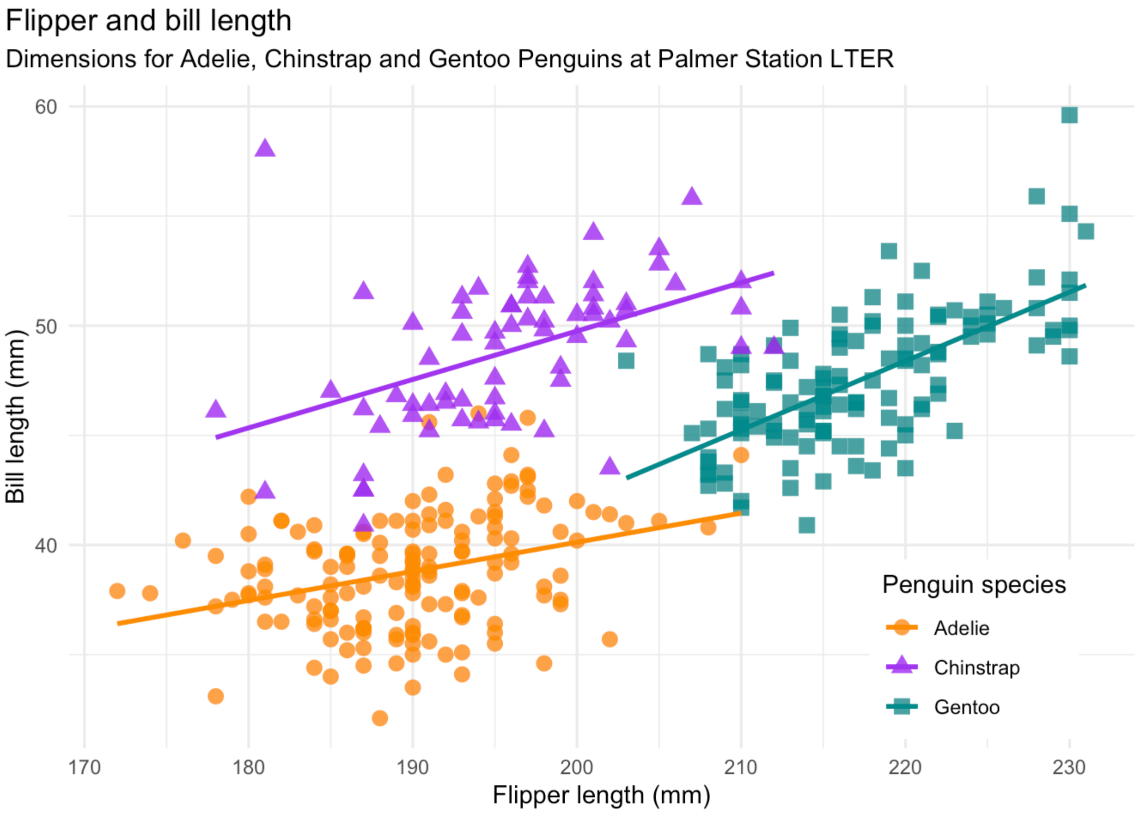
Exercise 1

Recreate, as faithfully as possible, the following plot using ggplot2 and the `penguins` data.



Exercise 2

Recreate, as faithfully as possible, the following plot from the `palmerpenguin` package README in ggplot2.



See the `palmerpenguins` pkgdown site