

Data Visualization

```
library(palmerpenguins)
```

```
##  
## Attaching package: 'palmerpenguins'  
  
## The following objects are masked from 'package:datasets':  
##  
##   penguins, penguins_raw
```

1. How many rows are in `penguins`? How many columns?
2. What does the `bill_depth_mm` variable in the `penguins` data frame describe? Read the help for `?penguins` to find out.
3. Make a scatterplot of `bill_depth_mm` vs. `bill_length_mm`. That is, make a scatterplot with `bill_depth_mm` on the y-axis and `bill_length_mm` on the x-axis. Describe the relationship between these two variables.
4. What happens if you make a scatterplot of `species` vs. `bill_depth_mm`? What might be a better choice of geom?
5. Why does the following give an error and how would you fix it?

```
ggplot(data = penguins) +  
  geom_point()
```

6. What does the `na.rm` argument do in `geom_point()`? What is the default value of the argument? Create a scatterplot where you successfully use this argument set to `TRUE`.
7. Add the following caption to the plot you made in the previous exercise: “Data come from the palmer-penguins package.” Hint: Take a look at the documentation for `labs()`.
8. Recreate the following visualization. What aesthetic should `bill_depth_mm` be mapped to? And should it be mapped at the global level or at the geom level?

g

5000 -

6000 -

9. Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions.

```
ggplot(  
  data = penguins,  
  mapping = aes(x = flipper_length_mm, y = body_mass_g, color = island)  
) +  
  geom_point() +  
  geom_smooth(se = FALSE)
```

10. Will these two graphs look different? Why/why not?

```
ggplot(  
  data = penguins,  
  mapping = aes(x = flipper_length_mm, y = body_mass_g)  
) +  
  geom_point() +  
  geom_smooth()  
  
ggplot() +  
  geom_point(  
    data = penguins,  
    mapping = aes(x = flipper_length_mm, y = body_mass_g)  
  ) +  
  geom_smooth(  
    data = penguins,  
    mapping = aes(x = flipper_length_mm, y = body_mass_g)  
  )
```

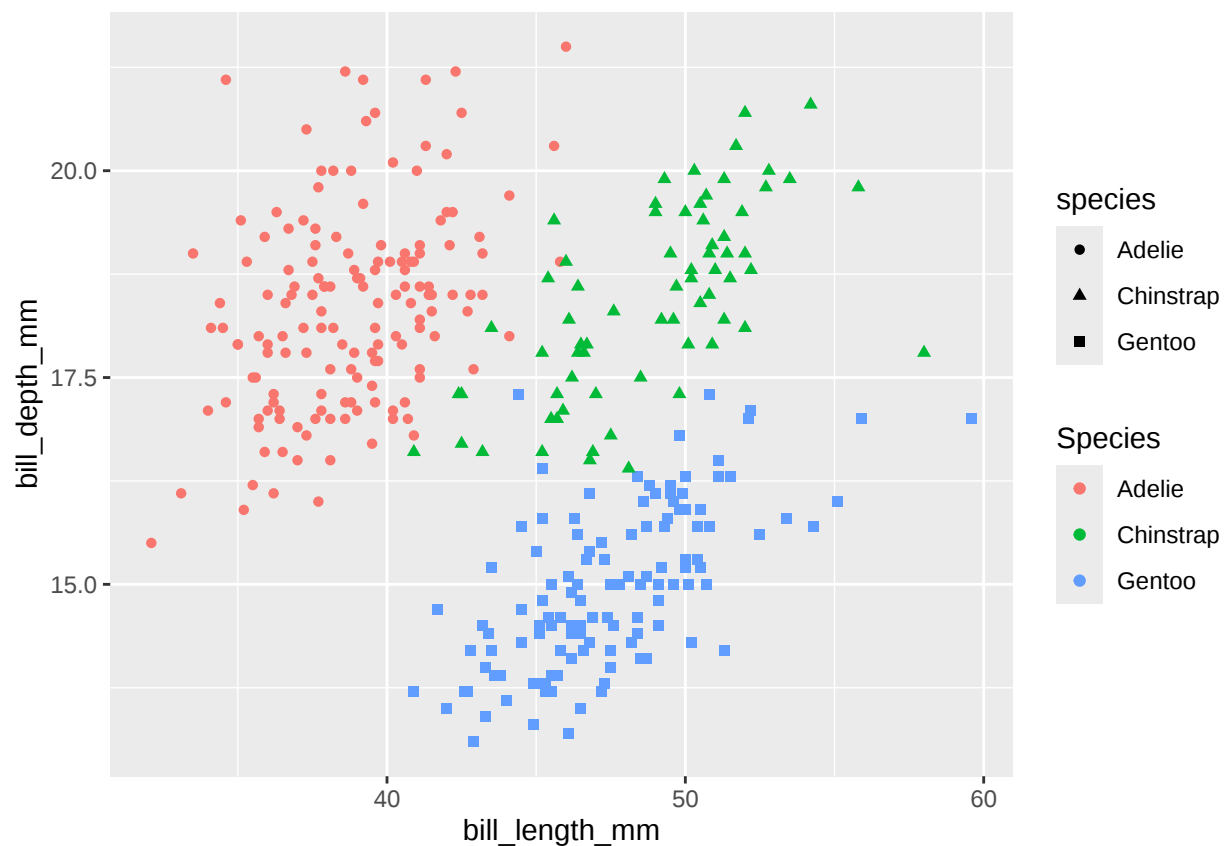
11. Make a bar plot of **species** of penguins, where you assign **species** to the y aesthetic. How is this plot different?
12. How are the following two plots different? Which aesthetic, **color** or **fill**, is more useful for changing the color of bars?

```
ggplot(penguins, aes(x = species)) +  
  geom_bar(color = "red")  
  
ggplot(penguins, aes(x = species)) +  
  geom_bar(fill = "red")
```

13. What does the **bins** argument in **geom_histogram()** do?
14. Make a histogram of the **carat** variable in the **diamonds** dataset that is available when you load the tidyverse package. Experiment with different binwidths. What binwidth reveals the most interesting patterns?
15. The **mpg** data frame that is bundled with the **ggplot2** package contains 234 observations collected by the US Environmental Protection Agency on 38 car models. Which variables in **mpg** are categorical? Which variables are numerical? (Hint: Type **?mpg** to read the documentation for the dataset.) How can you see this information when you run **mpg**?

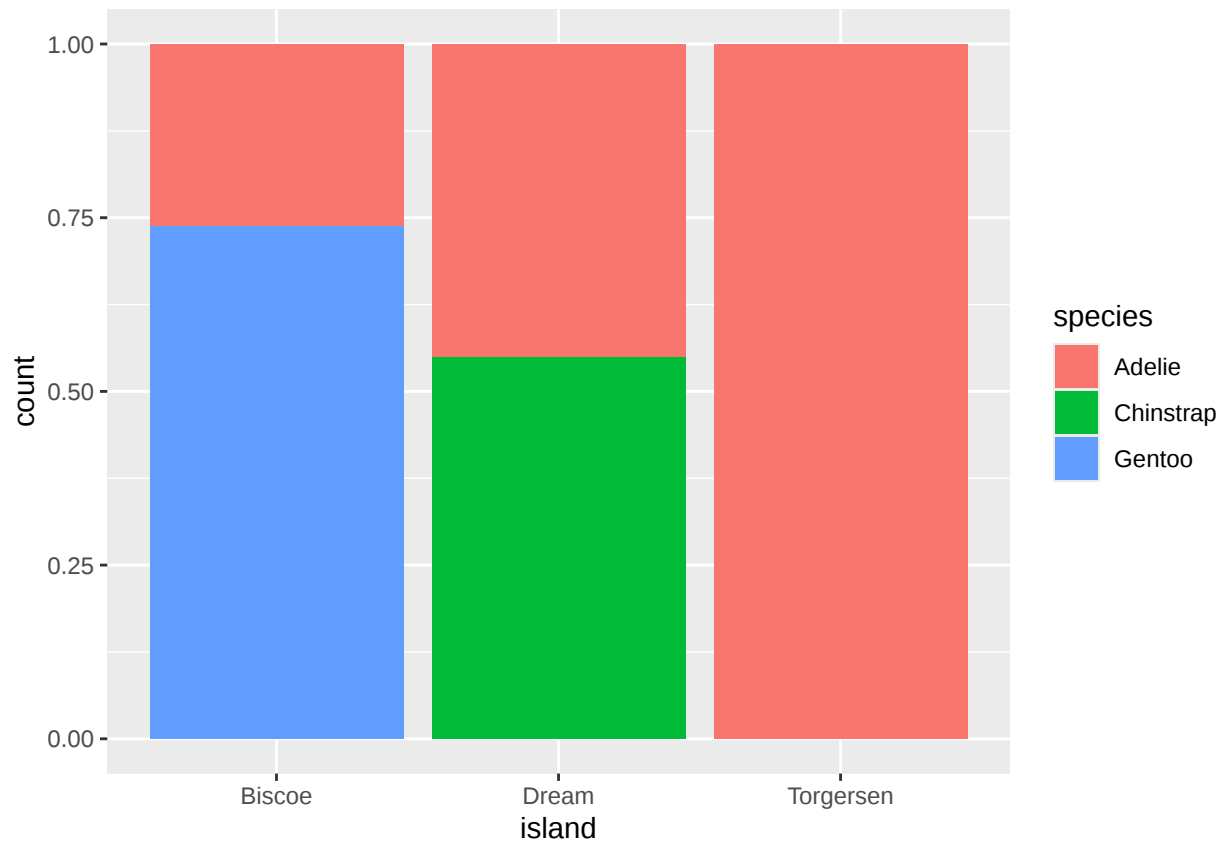
16. Make a scatterplot of `hwy` vs. `displ` using the `mpg` data frame. Next, map a third, numerical variable to `color`, then `size`, then both `color` and `size`, then `shape`. How do these aesthetics behave differently for categorical vs. numerical variables?
17. In the scatterplot of `hwy` vs. `displ`, what happens if you map a third variable to `linewidth`?
18. What happens if you map the same variable to multiple aesthetics?
19. Make a scatterplot of `bill_depth_mm` vs. `bill_length_mm` and color the points by `species`. What does adding coloring by species reveal about the relationship between these two variables? What about faceting by `species`?
20. Why does the following yield two separate legends? How would you fix it to combine the two legends?

```
ggplot(
  data = penguins,
  mapping = aes(
    x = bill_length_mm, y = bill_depth_mm,
    color = species, shape = species
  )
) +
  geom_point() +
  labs(color = "Species")
```

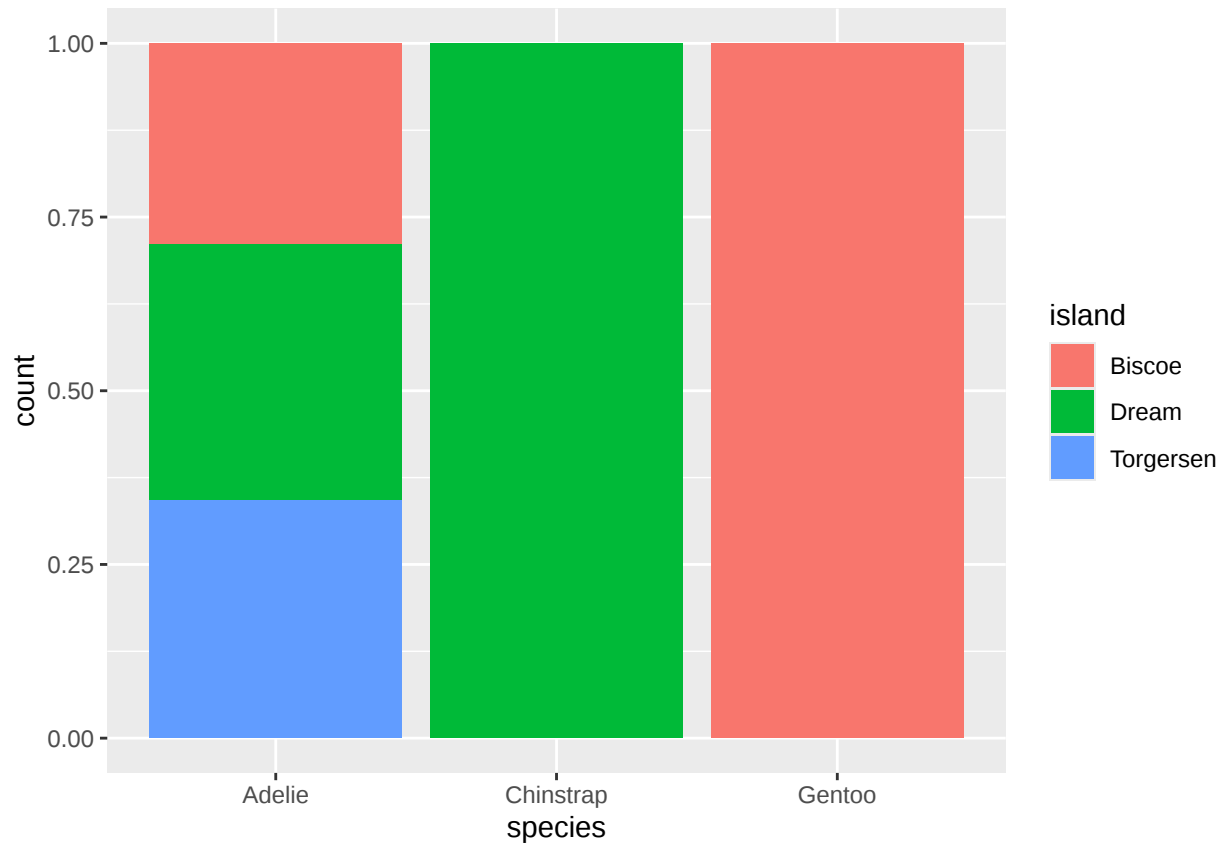


21. Create the two following stacked bar plots. Which question can you answer with the first one? Which question can you answer with the second one?

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



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



22. Run the following lines of code. Which of the two plots is saved as `mpg-plot.png`? Why?

```
ggplot(mpg, aes(x = class)) +  
  geom_bar()  
ggplot(mpg, aes(x = cty, y = hwy)) +  
  geom_point()  
ggsave("mpg-plot.png")
```

23. What do you need to change in the code above to save the plot as a PDF instead of a PNG? How could you find out what types of image files would work in `ggsave()`?