1.2.5 Exercises

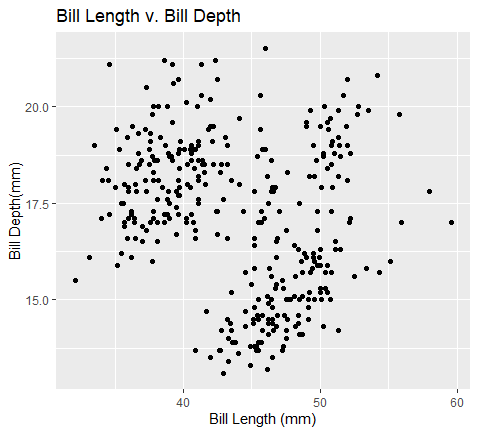
1. How many rows are in penguins? How many columns?

There are 344 rows and 8 columns.

1. What does the bill\_depth\_mm variable in the penguins data frame describe? Read the help for [?penguins](https://allisonhorst.github.io/palmerpenguins/reference/penguins.html) to find out.

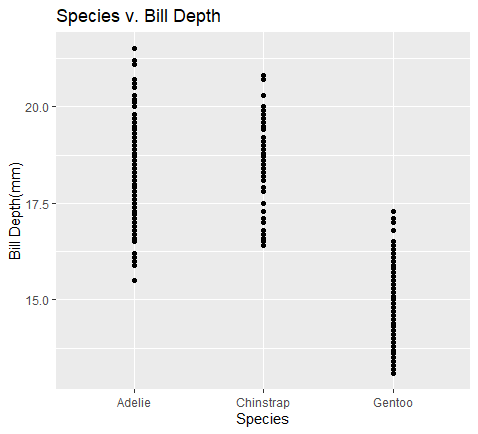
The bill\_depth\_mm variable describes a number denoting bill depth (millimeters). The bill\_depth\_mm is a double variable type.

1. 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.

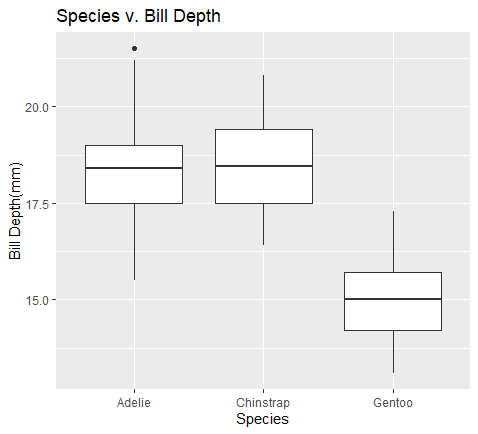


There is no relationship between these two variables.

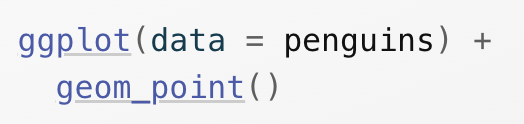
1. What happens if you make a scatterplot of species vs. bill\_depth\_mm? What might be a better choice of geom?

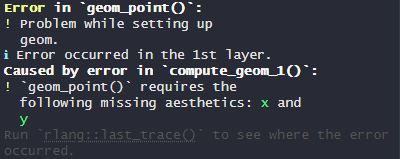


It is a very bad representation of the data. A better goem to use would be goem\_boxplot, which presents the data as a boxplot.



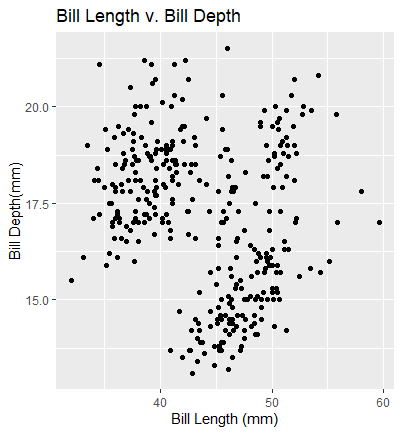
1. Why does the following give an error and how would you fix it?





(I don’t know why the error picture is blurry, but it is).

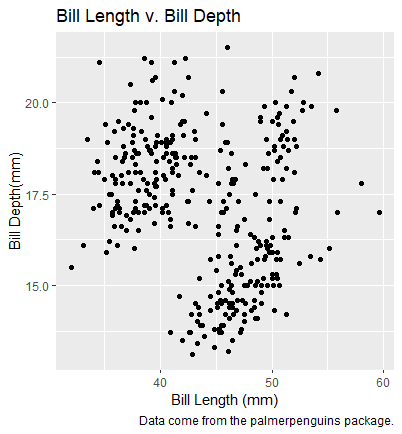
The following gives an error because it is asking to make a scatter plot without assigning values to the x-axis or y-axis. You could fix it by assigning values to the x-axis and y-axis.

1. What does the na.rm argument do in [geom\_point()](https://ggplot2.tidyverse.org/reference/geom_point.html)? What is the default value of the argument? Create a scatterplot where you successfully use this argument set to TRUE.

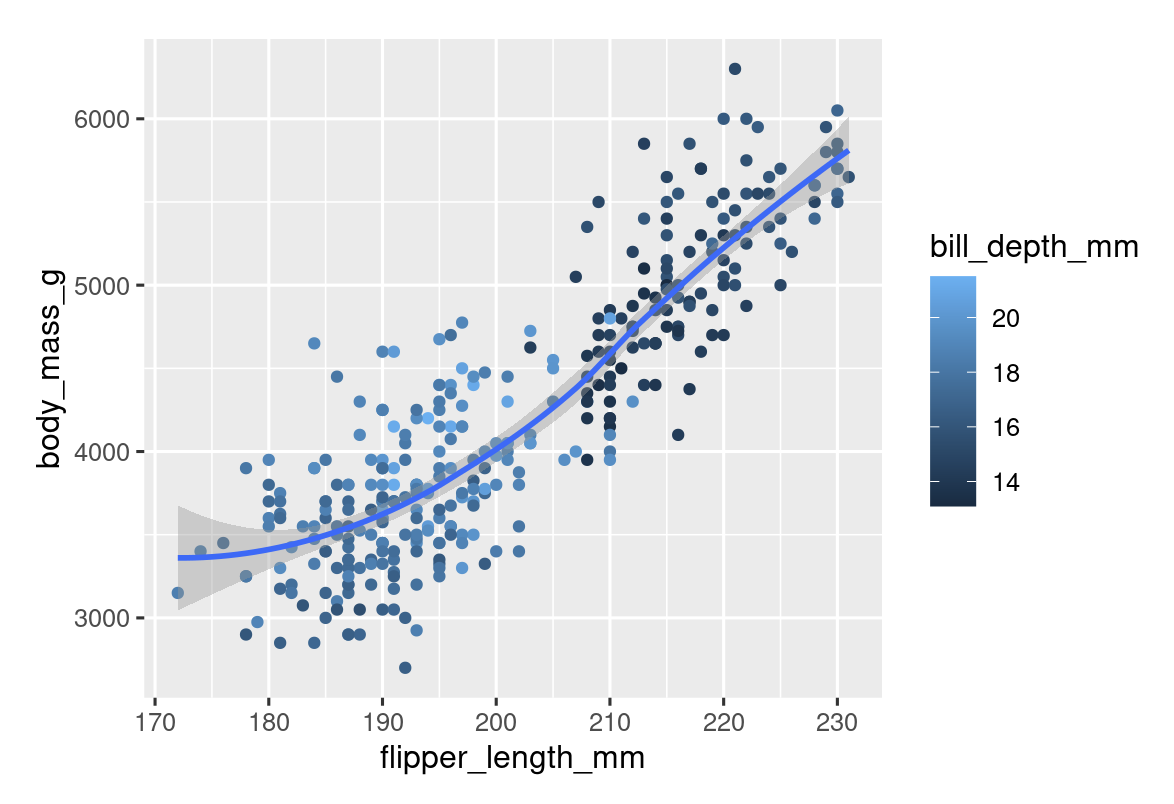
The na.rm argument works with missing values. If FALSE, the missing values are removed with a warning, and if TRUE, the missing values are removed silently. The default value is FALSE.



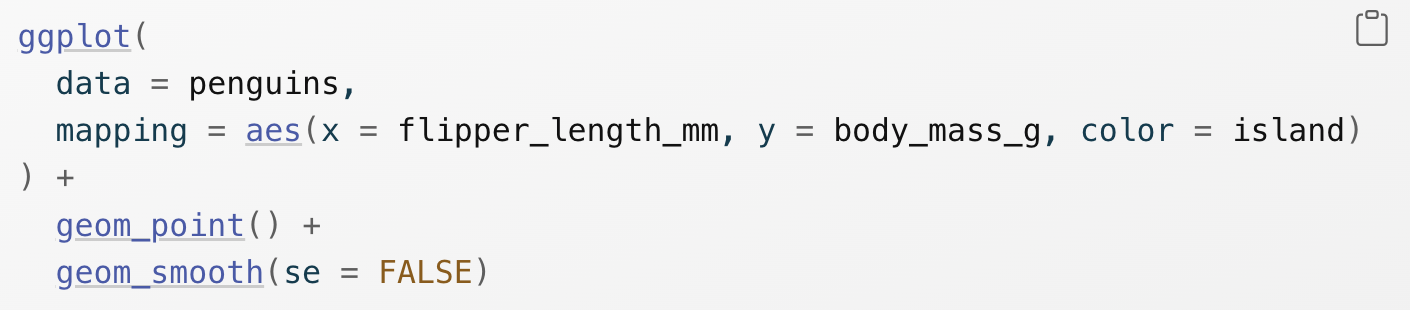
1. Add the following caption to the plot you made in the previous exercise: “Data come from the palmerpenguins package.” Hint: Take a look at the documentation for [labs()](https://ggplot2.tidyverse.org/reference/labs.html).

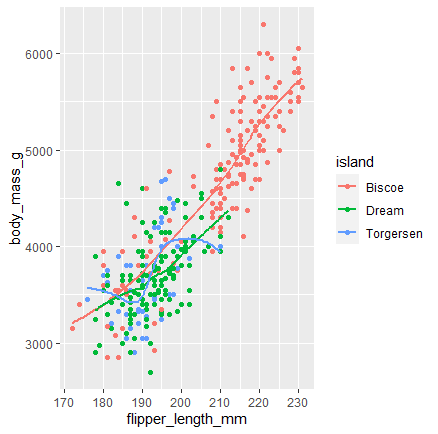


1. Recreate the following visualization. What aesthetic should bill\_depth\_mm be mapped to? And should it be mapped at the global level or at the geom level?

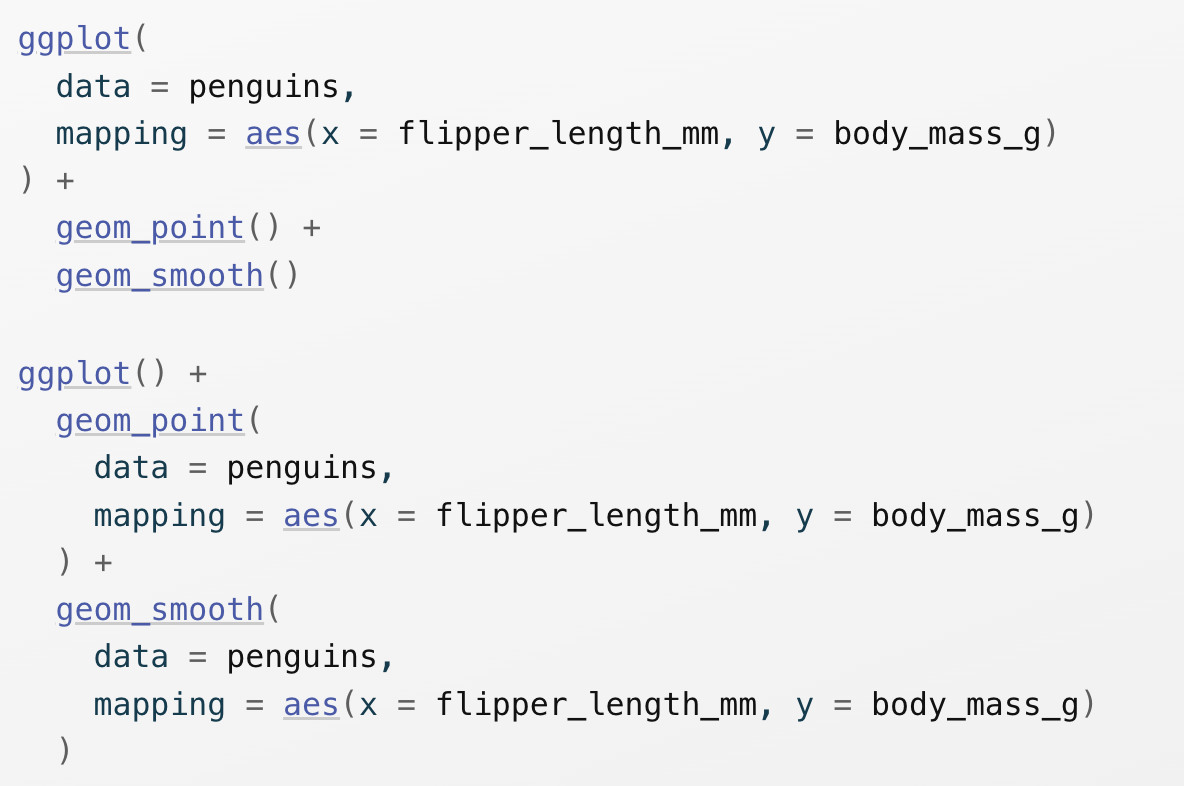
The bill\_depth\_mm should be mapped to the color aesthetic. 

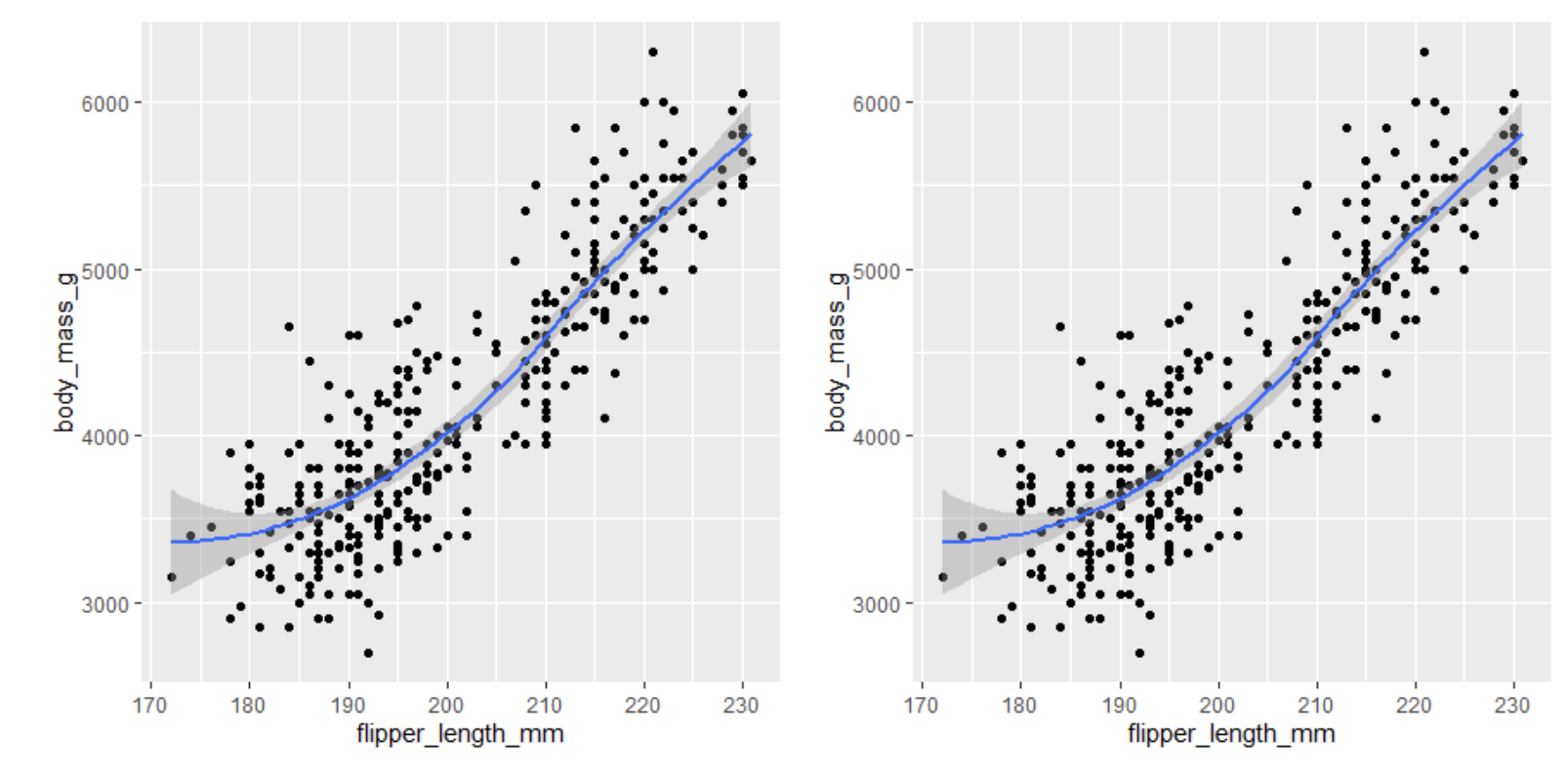
1. Run this code in your head and predict what the output will look like. Then, run the code in R and check your predictions.

The output will be a scatterplot of flipper\_length\_mm versus body\_mass\_g with the color aesthetic of islands. The graph also includes a line of best fit for each island but no statistical errors.

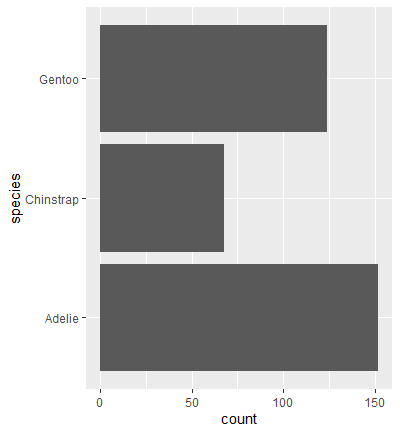


1. Will these two graphs look different? Why/why not?

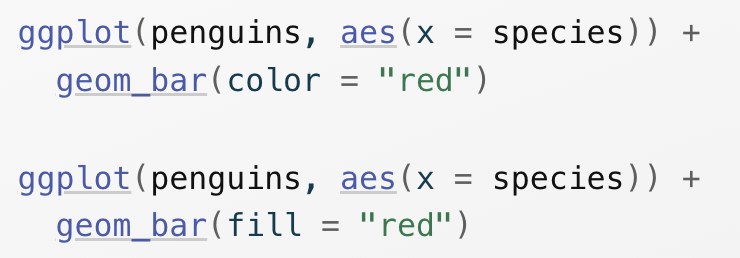
These two graphs will look the same. The first graph is done globally, and the second graph is done globally but manually.

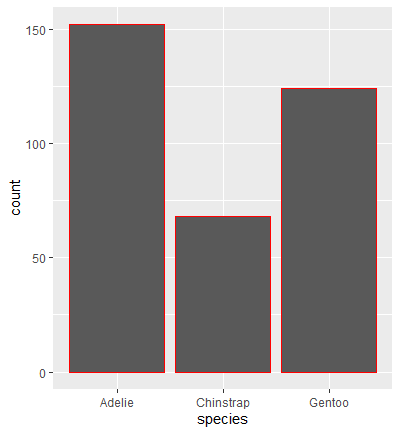


1.4.3 Exercises

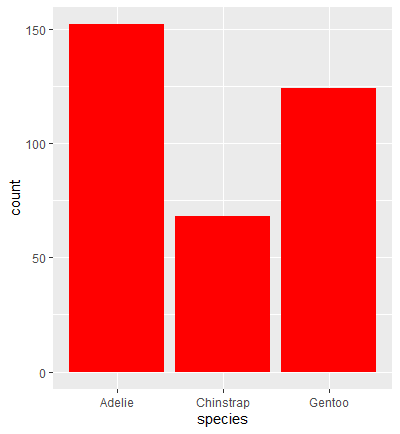
1. Make a bar plot of species of penguins, where you assign species to the y aesthetic. How is this plot different?

This plot is different because the graph is horizontal instead of vertical. Basically, the x-axis and y-axis are flipped.

1. How are the following two plots different? Which aesthetic, color or fill, is more useful for changing the color of bars?



The first graph (using color = “red”) has a red outline of the bars.

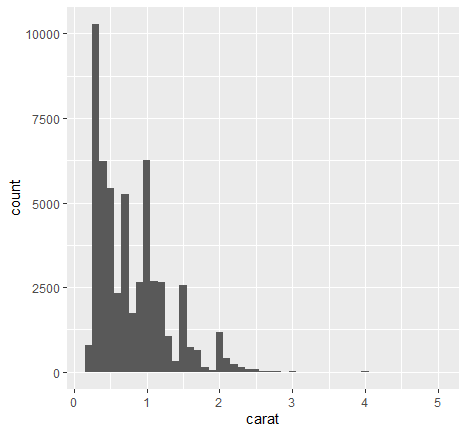


The second graph (using fill = “red”) actually fills the bars with the color red.

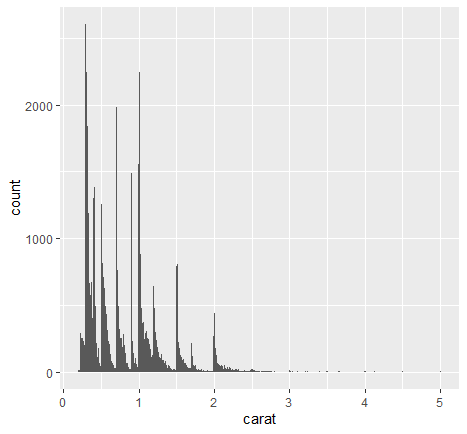
As can be seen, the fill aesthetic is more useful for changing the colors of the bars.

1. What does the bins argument in [geom\_histogram()](https://ggplot2.tidyverse.org/reference/geom_histogram.html) do?

The bins argument is the number of bins. The value defaults to 30 bins and is overridden by binwidth.

1. 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?

This graph has a binwidth of 0.1



This graph has a binwidth of 0.01

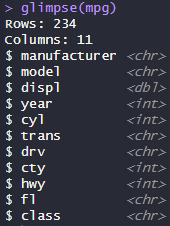
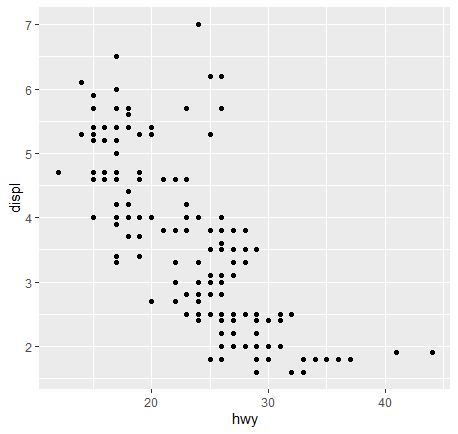
I believe that this binwidth is the most interesting for this graph because it shows more bins and, therefore, better representation of the data.

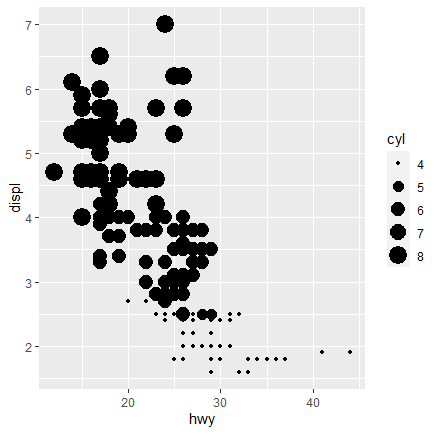
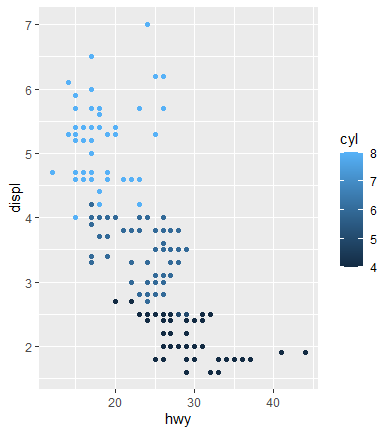
1.5.5 Exercises

1. 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](https://ggplot2.tidyverse.org/reference/mpg.html) to read the documentation for the dataset.) How can you see this information when you run mpg?

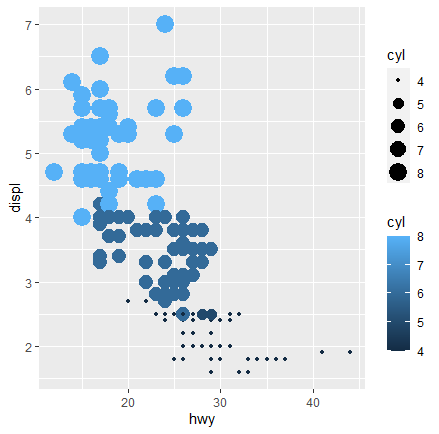
The categorical variables are manufacturer, model, trans, drv, fl, and class. The numerical variables are displ, year, cyl, cty, and hwy. When you run mpg, the variable types are listed for each column at the top. However, it cuts off. That’s why I used glimpse to get all of the variable types.

 This is from mpg.

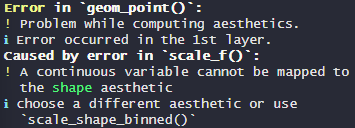
 This is from glimpse(mpg).

1. 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





The shape = cyl will not work because cyl is a continuous variable and cannot be mapped to the shape aesthetic.



For categorical variables, each category is assigned with a specific color. For numerical variables, the color is normally different shades of the same color, which can be harder to determine differences between the numerical values.

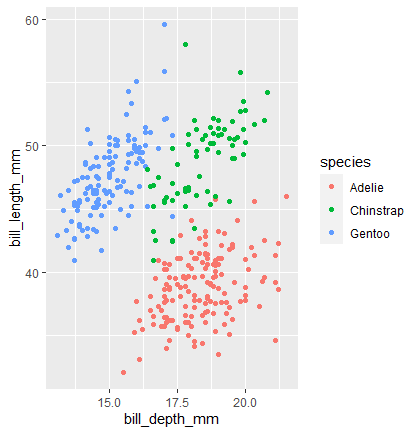
1. In the scatterplot of hwy vs. displ, what happens if you map a third variable to linewidth?

The linewidth does not work for a scatterplot. If linewidth is set to a number, it just ignores it. If linewidth is set to a variable, it says that the object is not found.

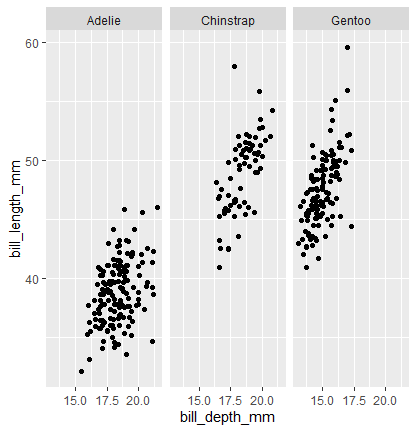
1. What happens if you map the same variable to multiple aesthetics?

If you map the same variable to multiple aesthetics, each one creates its legend, and both are used. However, it is repeated data to use color and size for the same variable because both show the differences individually, so there is no need to use both aesthetics.

1. 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?



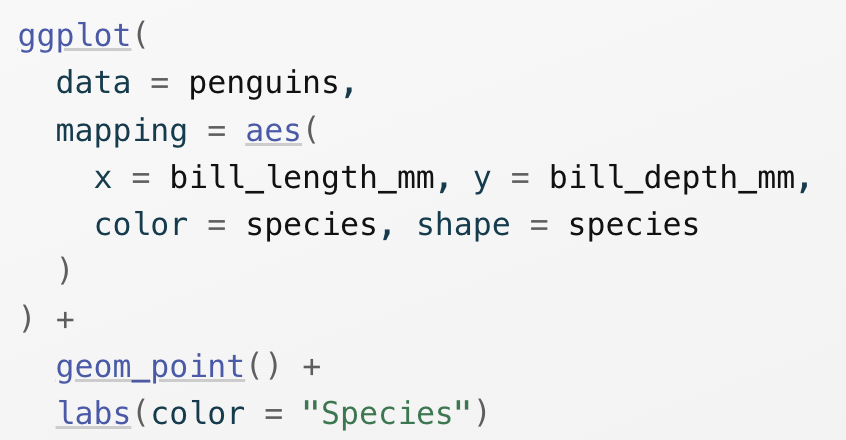
Adding color by species reveals that each species has a general range for their bill depths and lengths.



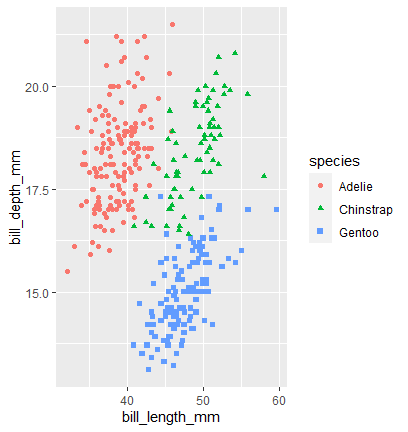


Faceting by species makes it a little easier to see the different ranges for each species.

1. Why does the following yield two separate legends? How would you fix it to combine the two legends?



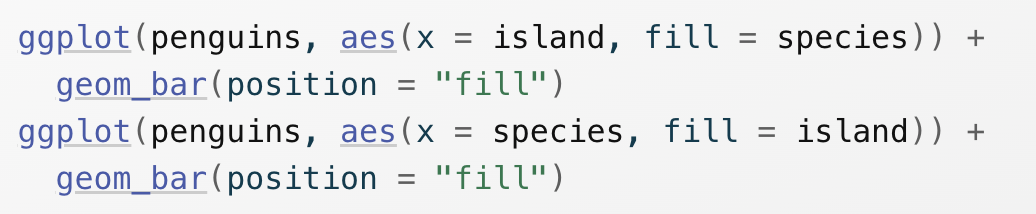
The code yields two separate legends because species is set to two aesthetics, color and shape, so it needs a legend for each aesthetic.

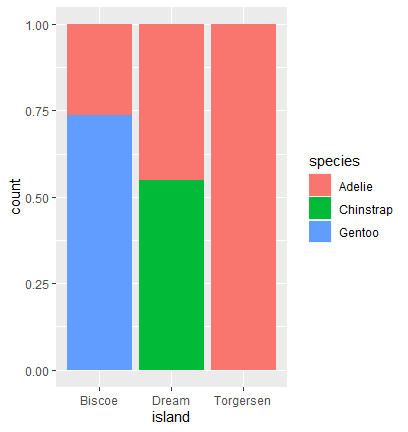




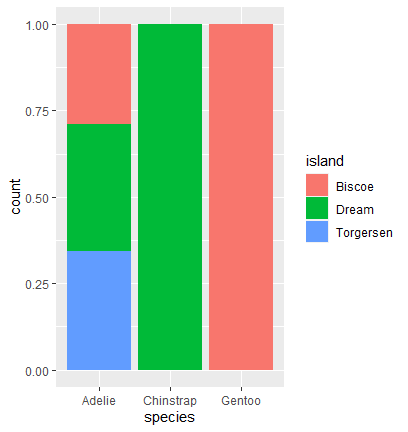
By putting the color and shape aesthetic into geom\_point, it combines the two legends into one because both aesthetics are using the same variable.

1. 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?





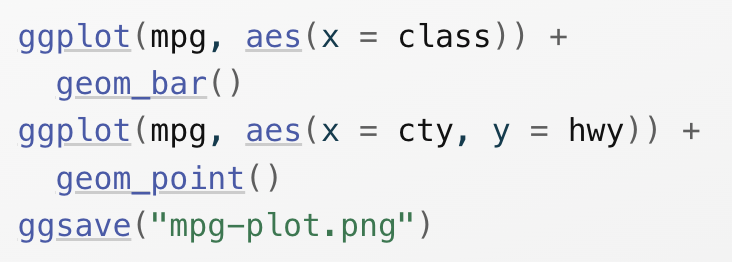
The first stacked bar plot answers the question, “What is the percentage of each species on each island?”



The second stacked bar plot answers the question, “What is the percentage of each island for each species?”

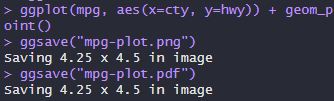
1.6.1 Exercises

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



The second plot is saved as mpg-plot.png because it is the plot most recently created to disk.

1. 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()](https://ggplot2.tidyverse.org/reference/ggsave.html)?



You can find out the types of image files that would work in ggsave() by using ?ggsave() to view its documentation.

2.5 Exercises

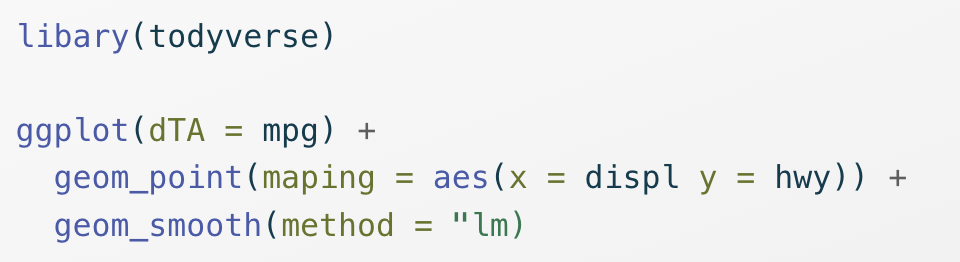
1. Why does this code not work?

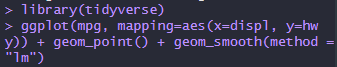


Look carefully! (This may seem like an exercise in pointlessness, but training your brain to notice even the tiniest difference will pay off when programming.)

This code does not work because variable is spelt wrong on the second line.

1. Tweak each of the following R commands so that they run correctly:

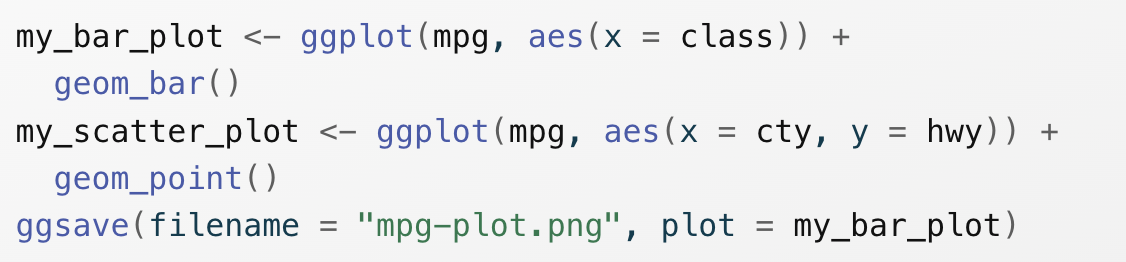




1. Press Option + Shift + K / Alt + Shift + K. What happens? How can you get to the same place using the menus?

The Keyboard Shortcut Quick Reference menu comes up after that command. You can get to the same place by clicking Tools then Keyboard Shortcuts Help.

1. Let’s revisit an exercise from the [Section 1.6](https://r4ds.hadley.nz/data-visualize#sec-ggsave). Run the following lines of code. Which of the two plots is saved as mpg-plot.png? Why?



The first plot is saved as mpg-plot.png because the code refers to my\_bar\_plot, which is the first one.