Introduction to ggplot2

Arjun Mannem

Friday, January 26, 2024

library(palmerpenguins)  
library(tidyverse)  
library(knitr)

## The ggplot2 package

First, let’s load the Palmer Penguins data set and explore some high-level characteristics of the data set.

data(penguins, package = "palmerpenguins")  
  
glimpse(penguins)

## Rows: 344  
## Columns: 8  
## $ species <fct> Adelie, Adelie, Adelie, Adelie, Adelie, Adelie, Adel…  
## $ island <fct> Torgersen, Torgersen, Torgersen, Torgersen, Torgerse…  
## $ bill\_length\_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, …  
## $ bill\_depth\_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, …  
## $ flipper\_length\_mm <int> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186…  
## $ body\_mass\_g <int> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, …  
## $ sex <fct> male, female, female, NA, female, male, female, male…  
## $ year <int> 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007, 2007…

### Plot in layers

Start by creating a blank canvas using the penguins data frame.

ggplot(data = penguins)



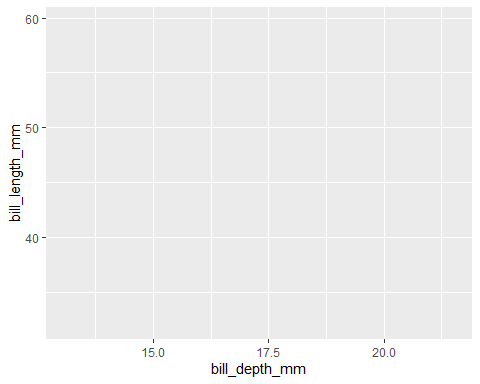
Then, in a new code chunk map bill depth to the x-axis.

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm))



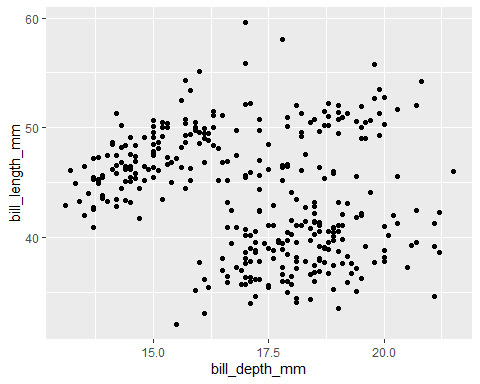
Next, in another new code chunk, add bill length to the y-axis.

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm))



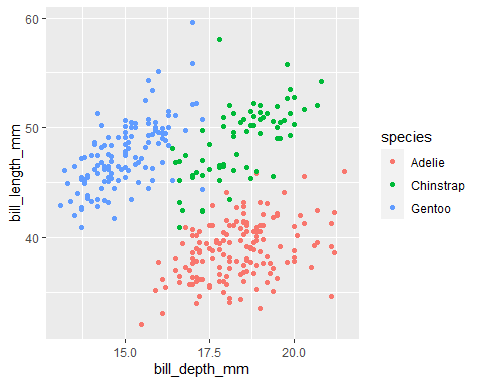
Create a scatter plot, representing each two-dimensional observation with a point by adding a geom\_point() layer. Additional layers are added to a ggplot using the + operator.

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm)) +   
 geom\_point()



Let’s color the points in our scatter plot based on the species of the penguin being Adelie, Chinstrap, or Gentoo, reproducing the plot below.

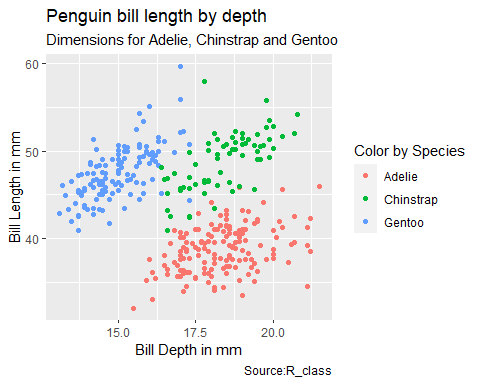
ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 color = species)) +  
 geom\_point()



Next, add another layer to our ggplot to customize the title and subtitle using the labs() function to reproduce the plot below.

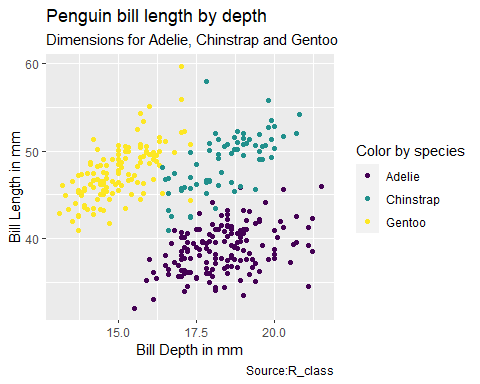
We can also customize the legend title by using the color option and a caption for the overall plot using the caption option in the labs() function. Customize the legend title to be “Species” instead of “species”, and modify the plot caption to match our overall desired plot.

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 color = species)) +  
 geom\_point() +  
 labs(x = "Bill Depth in mm",  
 y = "Bill Length in mm",  
 color = "Color by Species",  
 title = "Penguin bill length by depth",  
 subtitle = "Dimensions for Adelie, Chinstrap and Gentoo",  
 caption = "Source:R\_class"  
 )



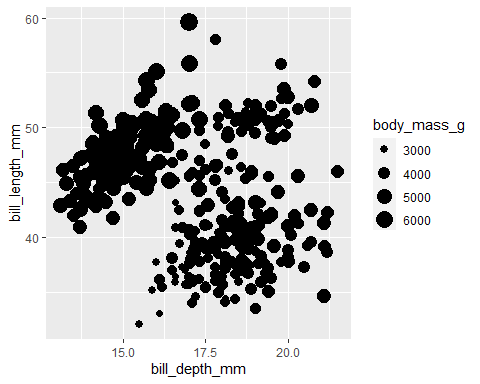
Lastly, use a discrete color scale that is inclusive of viewers with common forms of color blindness by adding a scale\_color\_viridis\_d() layer to the plot, creating the final plot below.

ggplot(data = penguins,  
 mapping = aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 color = species)) +  
 geom\_point() +  
 labs(x = "Bill Depth in mm",  
 y = "Bill Length in mm",  
 color = "Color by species",  
 title = "Penguin bill length by depth",  
 subtitle = "Dimensions for Adelie, Chinstrap and Gentoo",  
 caption = "Source:R\_class"  
 ) +  
 scale\_color\_viridis\_d()



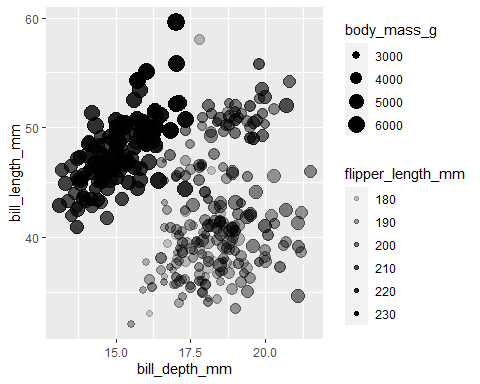
### Mapping vs. Setting

ggplot(penguins,  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 size = body\_mass\_g)) +   
 geom\_point()

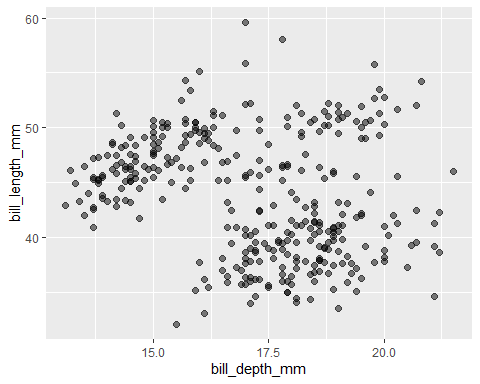


Reproduce the plot below by also mapping flipper\_length\_mm to the alpha aesthetic.

ggplot(penguins,  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 size = body\_mass\_g,  
 alpha = flipper\_length\_mm )) +   
 geom\_point()



ggplot(penguins,  
 aes(x = bill\_depth\_mm,  
 y = bill\_length\_mm,  
 alpha = flipper\_length\_mm)) +   
 geom\_point(size = 2, alpha = 0.50)



Is there any missing data? What is the plot doing with the missing values? Hint: consider using the skim() function from the skimr package to assess missingness.

library(skimr)  
  
skim(penguins)

Data summary

|  |  |
| --- | --- |
| Name | penguins |
| Number of rows | 344 |
| Number of columns | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 3 |
| numeric | 5 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

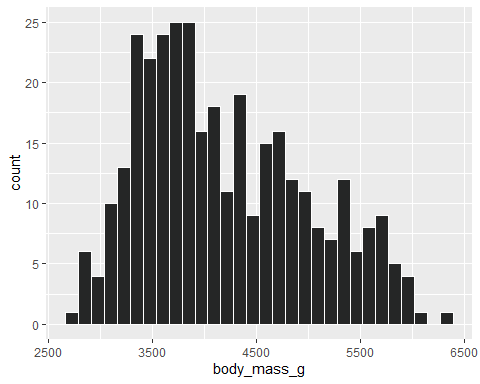
| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| species | 0 | 1.00 | FALSE | 3 | Ade: 152, Gen: 124, Chi: 68 |
| island | 0 | 1.00 | FALSE | 3 | Bis: 168, Dre: 124, Tor: 52 |
| sex | 11 | 0.97 | FALSE | 2 | mal: 168, fem: 165 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| bill\_length\_mm | 2 | 0.99 | 43.92 | 5.46 | 32.1 | 39.23 | 44.45 | 48.5 | 59.6 | ▃▇▇▆▁ |
| bill\_depth\_mm | 2 | 0.99 | 17.15 | 1.97 | 13.1 | 15.60 | 17.30 | 18.7 | 21.5 | ▅▅▇▇▂ |
| flipper\_length\_mm | 2 | 0.99 | 200.92 | 14.06 | 172.0 | 190.00 | 197.00 | 213.0 | 231.0 | ▂▇▃▅▂ |
| body\_mass\_g | 2 | 0.99 | 4201.75 | 801.95 | 2700.0 | 3550.00 | 4050.00 | 4750.0 | 6300.0 | ▃▇▆▃▂ |
| year | 0 | 1.00 | 2008.03 | 0.82 | 2007.0 | 2007.00 | 2008.00 | 2009.0 | 2009.0 | ▇▁▇▁▇ |

Reproduce the histogram of the penguin body masses displayed below using ggplot2.

ggplot(penguins,  
 aes(x = body\_mass\_g)) +   
 geom\_histogram(color = "white", fill = "#252626")



Reproduce the box plot of the penguin body masses displayed below using ggplot2. Hint: to suppress the y-axis text and ticks, add a theme(axis.text.y = element\_blank(), axis.ticks.y = element\_blank()) layer to the plot.

ggplot(penguins,  
 aes(x = body\_mass\_g,  
 color = species)) +   
 geom\_boxplot() +  
 labs(x = "Bodymass (g)",  
 color = "Color by Species",  
 title = "Penguin bodymasses",  
 caption = "Source:Palmer Station LTER/palmerpenguins package"  
 ) +  
 theme(axis.text.y = element\_blank(), axis.ticks.y = element\_blank())

