Homework 1

Samiha Reza

2024-09-15

# Problem 1

Here is the code chunk for ***data description***.

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

## species island bill\_length\_mm bill\_depth\_mm   
## Adelie :152 Biscoe :168 Min. :32.10 Min. :13.10   
## Chinstrap: 68 Dream :124 1st Qu.:39.23 1st Qu.:15.60   
## Gentoo :124 Torgersen: 52 Median :44.45 Median :17.30   
## Mean :43.92 Mean :17.15   
## 3rd Qu.:48.50 3rd Qu.:18.70   
## Max. :59.60 Max. :21.50   
## NA's :2 NA's :2   
## flipper\_length\_mm body\_mass\_g sex year   
## Min. :172.0 Min. :2700 female:165 Min. :2007   
## 1st Qu.:190.0 1st Qu.:3550 male :168 1st Qu.:2007   
## Median :197.0 Median :4050 NA's : 11 Median :2008   
## Mean :200.9 Mean :4202 Mean :2008   
## 3rd Qu.:213.0 3rd Qu.:4750 3rd Qu.:2009   
## Max. :231.0 Max. :6300 Max. :2009   
## NA's :2 NA's :2

nrow(penguins)

## [1] 344

ncol(penguins)

## [1] 8

The “penguins” data includes 8 variables: species, island, bill length (mm), bill depth (mm), flipper length (mm), body mass (g), sex, and year. There are three species: Adelie (152 penguins), Chinstrap (68 penguins), and Gentoo (124 penguins). There are three islands: Biscoe (168 pengiuins), Dream (124 penguins), and Torgersen (52 penguins). There are 8 columns, and 344 rows, so 344 data samples.The mean flipper length in mm is 200.9.

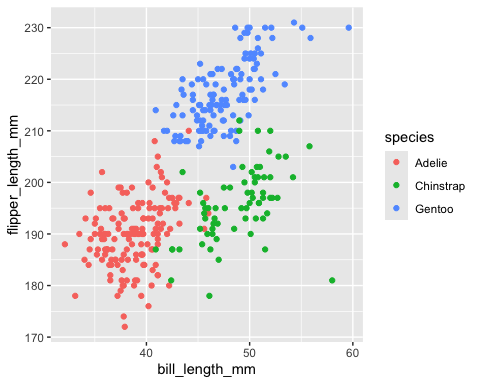
Here is the code chunk for the ***scatterplot***

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

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

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggsave("penguins\_scatter.pdf")

## Saving 5 x 4 in image

## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

# Problem 2

Since we have already loaded tidyverse, we do not need to do it again. Here we have a code chunk to create the ***data frame***.

question\_2\_df = tibble(  
 random\_sample = rnorm(10),  
 logic\_vector = random\_sample > 0,  
 character\_vector = rep("character", 10),  
 factor\_vector = factor(rep(c("Level1", "Level2", "Level3"), length.out =10)),  
)  
  
print(question\_2\_df)

## # A tibble: 10 × 4  
## random\_sample logic\_vector character\_vector factor\_vector  
## <dbl> <lgl> <chr> <fct>   
## 1 -0.0817 FALSE character Level1   
## 2 1.86 TRUE character Level2   
## 3 -0.602 FALSE character Level3   
## 4 -0.807 FALSE character Level1   
## 5 1.06 TRUE character Level2   
## 6 1.31 TRUE character Level3   
## 7 -2.12 FALSE character Level1   
## 8 -2.06 FALSE character Level2   
## 9 1.94 TRUE character Level3   
## 10 0.549 TRUE character Level1

Here we pull variables from the dataframe and try to find the means of each.

random\_sample <- question\_2\_df$random\_sample  
logic\_vector <- question\_2\_df$logic\_vector  
character\_vector <- question\_2\_df$character\_vector  
factor\_vector <- question\_2\_df$factor\_vector  
  
mean(random\_sample)

## [1] 0.1046619

mean(logic\_vector)

## [1] 0.5

mean(character\_vector)

## Warning in mean.default(character\_vector): argument is not numeric or logical:  
## returning NA

## [1] NA

mean(factor\_vector)

## Warning in mean.default(factor\_vector): argument is not numeric or logical:  
## returning NA

## [1] NA

Only random\_sample and logic\_vector has a mean value. Character and Factor vector do not because “argument is not numeric or logical.”

We now check if we can convert the variables to numeric ones.

numeric\_logical = as.numeric(logic\_vector)  
# numeric\_character = as.numeric(character\_vector)  
# numeric\_factor = as.numeric(factor)

Character and factor vector cannot be converted to numeric, which is why it isn’t possible to find the mean.