p8105_hw1_jl5788

Jie Liu

9/26/2021

Problem 1

```
data_df = tibble(
  norm_samp = rnorm(10),
  norm_samp_pos = norm_samp >0,
  character_vector = c("a","b","c","d","e","f","g","h","i","j"),
  type = factor(c("positive","zero","negative","positive","zero","negative","positive","zero","negative
)
```

Data Frame

```
mean(pull(data_df,norm_samp))
```

1. Caculate the mean of the first variable of dataframe Success! The result is -0.3926475.

```
mean(pull(data_df,norm_samp_pos))
```

2. Caculate the mean of the second variable of dataframe Success! The result is 0.3.

```
mean(pull(data_df,character_vector))
```

3. Caculate the mean of the third variable of dataframe Failure! The result is NA.

```
mean(pull(data_df,type))
```

4. Caculate the mean of the forth variable of dataframe Failure! The result is NA.

```
as.numeric(3>2)
```

When using as.numeric function to the logical, character, and factor variables, what will happen? Success! The result is 1. Output is 1 because logical variable True equal to 1. On the contrary, if logical variable is False, as.numeric function results to 0.

```
as.numeric("character")
```

The result is NA and in the Console,it writes warning message. Because "character" (character variable) cannot equal to any numeric variable like "1"(character variable) equal to 1 (numeric variable), so it just shows NA indicating it has been forced to transform character variables to numeric variables NA.

```
as.numeric(factor("good","bad"))
```

The result is NA. The reason is the same as shows above.

Does this help explain what happens when you try to take the mean? These three code chunks above and results they showed help explain what happens when try to take the mean. Because when we use as.numeric() function to transform character vector and factor vector to numeric vector, the results are both NA and we cannot caculate the average of NA, which is meaningless, so Console reports an error when we try it.

Problem 2

```
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
```

Short description of the penguins dataset

1. The names of the data in this dataset are species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex, year.
2. There are 344 rows and 8 columns in this dataset.
3. There are 3 speices of penguins, which are Adelie, Gentoo, Chinstrap.
4. They live in 3 islands, which are Torgersen, Biscoe, Dream.
5. The mean of flipper length is 200.9152047.

```
pen_plot = ggplot(penguins, aes(x = bill_length_mm, y = flipper_length_mm, color = species)) +geom_poin
ggsave("pen_plot.pdf")
```

Scatterplot of flipper_length_mm (y) vs bill_length_mm (x)

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

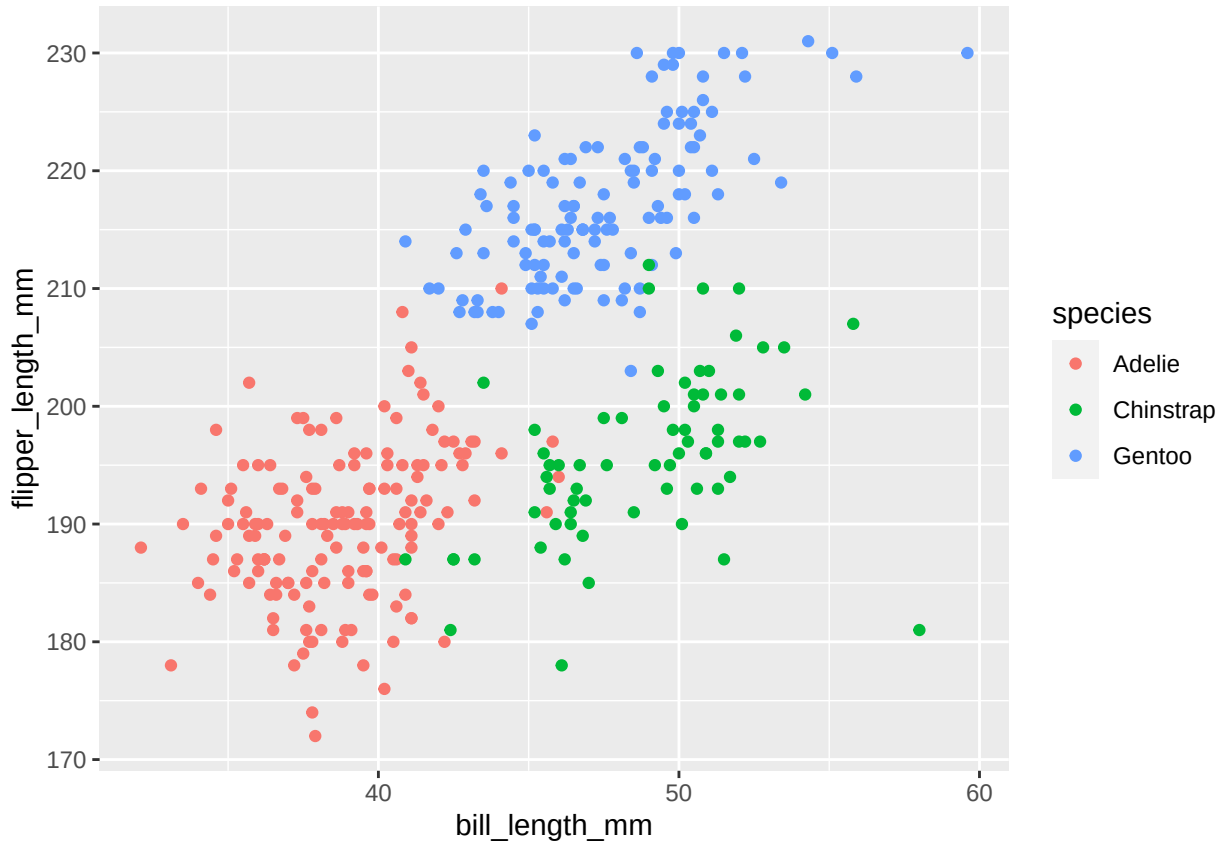

Figure 1: scatterplot