p8105_hw1_qz2493.Rmd

Qingyue Zhuo

2022-09-20

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

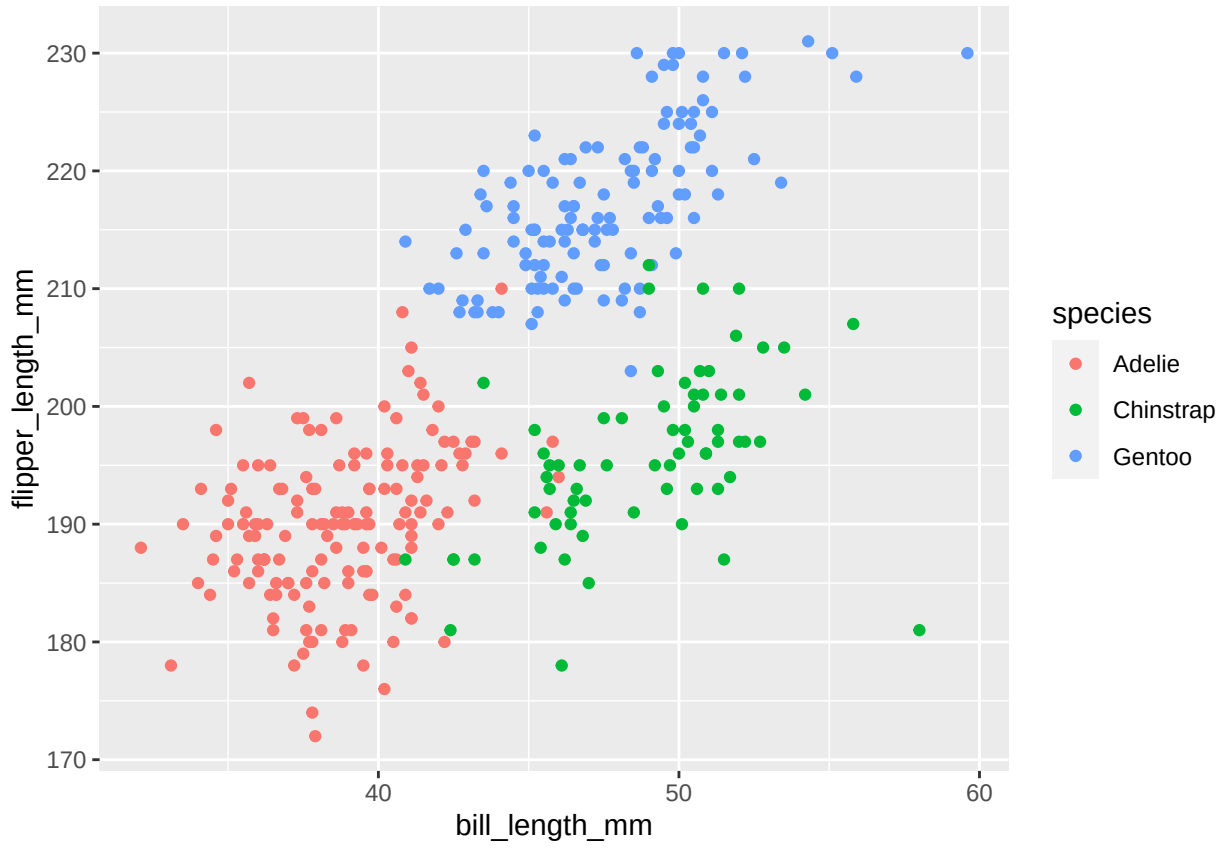
Problem 1

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

The variables of the data set are species, island, bill_length_mm, bill_depth_mm, flipper_length_mm, body_mass_g, sex, year. "Species" has three levels:Biscoe, Dream, Torgersen ; "island" has three levels: Biscoe, Dream, Torgersen; "sex" is a binary variable with two levels: female, male. The data set has 344 and 8 columns. The mean flipper length is 200.92 mm after excluding the missing value NAs.

```
ggplot(penguins, aes(x = bill_length_mm, y = flipper_length_mm, col = species)) + geom_point()
```

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




```
ggsave("Penguins_scatter_plot.pdf")
```

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

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

Problem 2

```
hw2_df = tibble(
  vec_numeric = rnorm(n = 10),
  vec_logical = vec_numeric > 0,
  vec_char = c('a','b','c','d','e','f','g','h','i','j'),
  vec_factor = factor(c("tall", "grande", "venti", "tall", "grande", "venti", "tall", "grande", "venti"
)

mean(pull(hw2_df,vec_numeric))
```

```
## [1] 0.2500993
```

```
mean(pull(hw2_df,vec_logical))
```

```
## [1] 0.7
```

```
mean(pull(hw2_df,vec_char))
```

```
## Warning in mean.default(pull(hw2_df, vec_char)): argument is not numeric or
## logical: returning NA
```

```
## [1] NA
```

```
mean(pull(hw2_df,vec_factor))
```

```
## Warning in mean.default(pull(hw2_df, vec_factor)): argument is not numeric or
## logical: returning NA
```

```
## [1] NA
```

The mean of numeric and logical vectors work, while the mean of character and factor vectors do not work.

```
as.numeric(pull(hw2_df,vec_logical))
as.numeric(pull(hw2_df,vec_char))
as.numeric(pull(hw2_df,vec_factor))
```

After applying the as.numeric function, the logical variable has two values 0 and 1; factor variable has three values 1, 2 and 3; while character variable turns into "NA"s. Logical variable only has two outcomes and can be counted as 0 and 1 in calculating mean. Factor variable is categorical variable, although it can be converted into numeric values 1, 2 and 3 but it means group 1, group 2 and group 3 therefore it makes no sense in calculating the mean. Character variable cannot be even be turned into numeric, therefore it doesn't have a mean.