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

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: Adelie, Chinstrap, and Gentoo; “island” has three levels: Biscoe, Dream and Torgersen; “sex” is a binary variable with two levels: male and female. The data set has 344 rows and 8 columns. The mean flipper length is 200.9152mm after excluding the missing value NAs.

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

```
## [1] "species"           "island"            "bill_length_mm"
## [4] "bill_depth_mm"     "flipper_length_mm" "body_mass_g"
## [7] "sex"               "year"
```

```
levels(penguins$species)
```

```
## [1] "Adelie"    "Chinstrap" "Gentoo"
```

```
levels(penguins$island)
```

```
## [1] "Biscoe"    "Dream"     "Torgersen"
```

```
nrow(penguins)
```

```
## [1] 344
```

```
ncol(penguins)
```

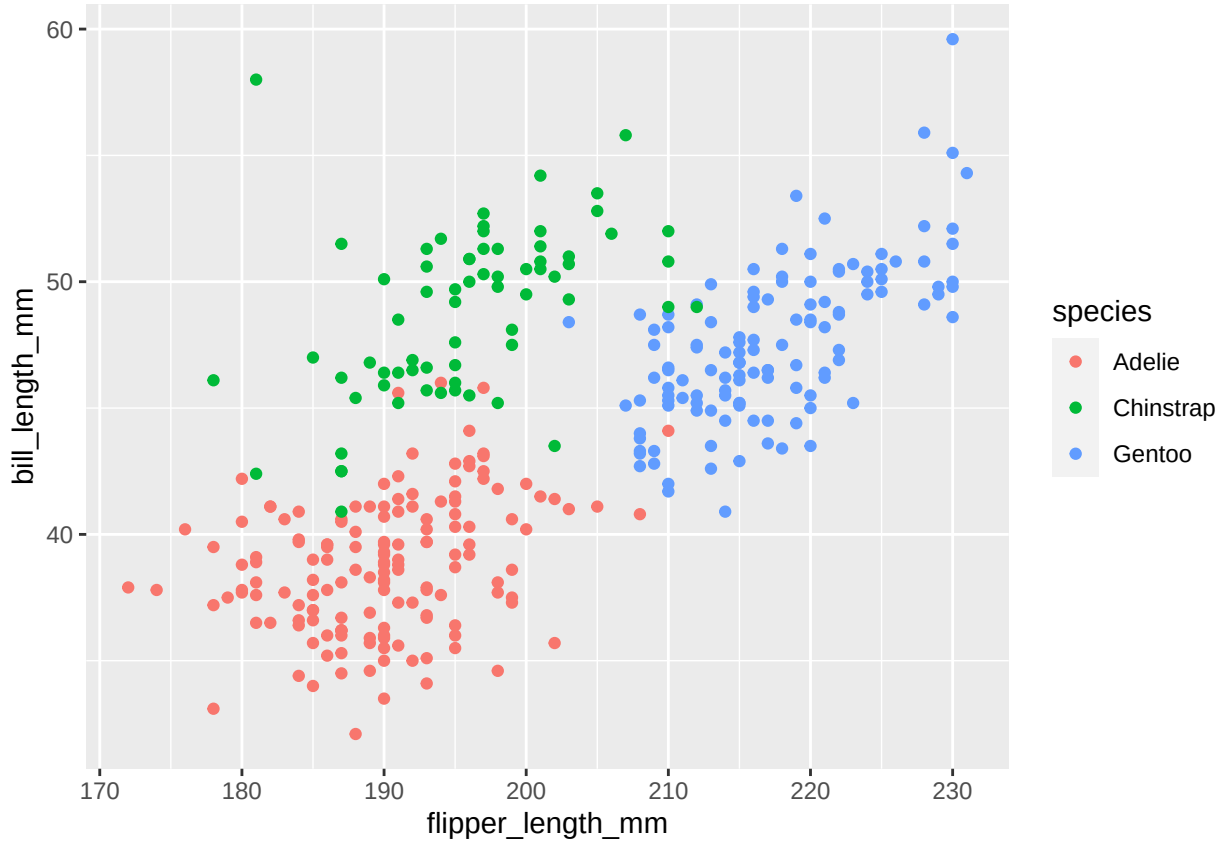
```
## [1] 8
```

```
mean(penguins$flipper_length_mm, na.rm = TRUE)
```

```
## [1] 200.9152
```

```
ggplot(penguins, aes(x = flipper_length_mm, y = bill_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

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

```
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] 0.6813646
```

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

```
## [1] 0.9
```

```
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
```