

P8105 Homework 1

Jia Ji (jj3205)

2022-09-24

Problem 1

```
## -- 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()
```

Loading libraries and penguins dataset.

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

```
## tibble [344 x 8] (S3: tbl_df/tbl/data.frame)
## $ species      : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ island       : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ bill_length_mm : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
## $ bill_depth_mm  : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
## $ flipper_length_mm: int [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
## $ body_mass_g    : int [1:344] 3750 3800 3250 NA 3450 3650 3625 4675 3475 4250 ...
## $ sex           : Factor w/ 2 levels "female","male": 2 1 1 NA 1 2 1 2 NA NA ...
## $ year          : int [1:344] 2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

```
nrow(penguins)
```

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

```
ncol(penguins)
```

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

```
mean(na.omit(penguins$flipper_length_mm))
```

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

Dataset penguins includes 344 rows and 8 columns, which are species, islands, bill length, bill depth, flipper length, body mass, sex, and year. The mean flipper length is 200.9152 (mm).

```
length_plot = ggplot(penguins, aes(x = bill_length_mm, y = flipper_length_mm)) +
  geom_point(aes(color = species))
ggsave("scatterplot.pdf", length_plot)
```

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

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

Export and save the first scatterplot to the project directory.

Problem 2

```
df = tibble(
  vec_numeric = rnorm(10),
  vec_logical = (vec_numeric > 0),
  vec_char = c("hi", "This", "is", "a", "character", "vector", "of", "length", "ten", "."),
  vec_factor = factor(c("alpha", "beta", "omega", "beta", "beta", "beta", "omega", "alpha",
                        "alpha", "beta"))
)

mean(pull(df,1))

## [1] -0.3118061
mean(pull(df,2))

## [1] 0.5
mean(pull(df,3))

## Warning in mean.default(pull(df, 3)): argument is not numeric or logical:
## returning NA
## [1] NA
mean(pull(df,4))

## Warning in mean.default(pull(df, 4)): argument is not numeric or logical:
## returning NA
## [1] NA

The mean of character vector and factor vector cannot work.

## [1] 0 0 0 1 1 1 1 0 1 0

## Warning: NAs introduced by coercion
## [1] NA NA NA NA NA NA NA NA NA NA

## [1] 1 2 3 2 2 2 3 1 1 2
```

We can take the mean of numeric variables and logical variables. Only logical variables and factor variables can be converted to numeric variables. Therefore, we can take the mean of logical variables rather than character variables unless it concludes number.