

practice__exercise

Ekta Chaudhary

28/06/2020

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.2
```

```
## v ggplot2 3.2.1    v purrr  0.3.2
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflicts
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)
library(dplyr)
```

```
practice_data = read_excel("./data/Practice_exercise.xlsx", sheet = "Data") %>%
  janitor::clean_names() %>%
  select(observation_number, quarter, employee_id, sex = sex_male_1, race, age, hospital_visit = hospital_visit)
  mutate(
    age_cat = case_when(
      age < 30 ~ 1,
      age <= 45 ~ 2,
      age > 45 ~ 3
    )
  )
```

```
sapply(practice_data, function(x) sum(is.na(x)))
```

```
## observation_number      quarter      employee_id
##                0                0                0
##                sex          race          age
##                71          2123                0
##      hospital_visit      salary      health_score
##                0                0                0
##                age_cat
##                0
```

```
practice_data %>%
  select(everything()) %>% # replace to your needs
  summarise_all(funs(sum(is.na(.))))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
```

```
##
## # Simple named list:
## list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`:
## tibble::lst(mean, median)
##
## # Using lambdas
## list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.

## # A tibble: 1 x 10
##   observation_num~ quarter employee_id sex race age hospital_visit
##             <int>   <int>       <int> <int> <int> <int>      <int>
## 1             0       0           0    71  2123    0          0
## # ... with 3 more variables: salary <int>, health_score <int>,
## #   age_cat <int>
```

```
sapply(practice_data, function(x) min(x))
```

```
## observation_number      quarter      employee_id
##      1.000000e+00      1.000000e+00      1.000000e+00
##           sex           race           age
##           NA           NA           7.000000e+00
##   hospital_visit      salary      health_score
##      0.000000e+00      2.835070e+04      6.265991e-01
##           age_cat
##      1.000000e+00
```

```
sapply(practice_data, function(x) max(x))
```

```
## observation_number      quarter      employee_id
##      19103.00           12.00           2000.00
##           sex           race           age
##           NA           NA           172.00
##   hospital_visit      salary      health_score
##           1.00      68826.34           10.00
##           age_cat
##           3.00
```