

# PHW251 Problem Set 4

Clara Voong

9/29/2024

For this problem set you will tidy up a dataset of 500 individuals. We also want to calculate each individual's BMI and appropriately categorize them.

Load your data (500\_Person\_Gender\_Height\_Weight.csv):

## Question 1

Clean the column headers to be all lower case, have no spaces, and rename “Location information” to location.

```
data <- data %>%  
  rename(  
    location = `Location information`,  
    gender = GENDER  
  )
```

```
colnames(data)
```

```
## [1] "location" "gender"   "height"   "weight"
```

## Question 2

Create a new variable that calculates BMI for each individual.

You will need to navigate the different system of measurements (metric vs imperial). Only the United States is using imperial.

- BMI calculation and conversions:
  - metric:  $BMI = weight(kg) / [height(m)]^2$
  - imperial:  $BMI = 703 * weight(lbs) / [height(in)]^2$
  - 1 foot = 12 inches
  - 1 cm = 0.01 meter

Although there's many ways you can accomplish this task, we want you to use an `if_else()` to calculate BMI with the appropriate formula based on each person's location.

```
# head(data)

data <- data %>%
  mutate(
    bmi = if_else(
      location %in% c("United Kingdom", "Taiwan"),
      weight / ((height * 0.01)^2),
      703 * weight / ((height * 12)^2)
    )
  )
head(data)
```

```
## # A tibble: 6 x 5
##   location      gender height weight  bmi
##   <chr>         <chr>   <dbl> <dbl> <dbl>
## 1 New York      Male     5.71  212.  31.7
## 2 United Kingdom Male    189    87  24.4
## 3 New York      Female   6.07  243.  32.1
## 4 Taiwan        Female  195   104  27.4
## 5 Taiwan        Male    149    61  27.5
## 6 Taiwan        Male    189   104  29.1
```

### Question 3

Create a new variable that categorizes BMI with `case_when()`:

- Underweight: BMI below 18.5
- Normal: 18.5-24.9
- Overweight: 25.0-29.9
- Obese: 30.0 and Above

```
data <- data %>% mutate(  
  bmi_cat =  
    case_when(  
      bmi < 18.5 ~ "Underweight",  
      bmi <= 24.9 ~ "Normal",  
      bmi <= 29.9 ~ "Overweight",  
      TRUE ~ "Obese"  
    )  
)
```

```
head(data)
```

```
## # A tibble: 6 x 6  
##   location      gender height weight   bmi bmi_cat  
##   <chr>         <chr>   <dbl> <dbl> <dbl> <chr>  
## 1 New York      Male    5.71  212.  31.7 Obese  
## 2 United Kingdom Male    189    87  24.4 Normal  
## 3 New York      Female  6.07  243.  32.1 Obese  
## 4 Taiwan        Female  195   104  27.4 Overweight  
## 5 Taiwan        Male   149    61  27.5 Overweight  
## 6 Taiwan        Male   189   104  29.1 Overweight
```

Could we have used `if_else()`?

Yes, but we would have to specify the lower bound as well for each line in contrast with `case_when()`, where the lines are run sequentially and the first match in the function would be the output.

#### Question 4

Arrange your data first by location and then by descending order of BMI.

```
data <- data %>%  
  arrange(location, desc(bmi))  
  
head(data)  
  
## # A tibble: 6 x 6  
##   location gender height weight   bmi bmi_cat  
##   <chr>    <chr>   <dbl>  <dbl> <dbl> <chr>  
## 1 Colorado Female    4.66   351.   78.8 Obese  
## 2 Colorado Female    4.59   322.   74.6 Obese  
## 3 Colorado Male      4.72   320.   70.1 Obese  
## 4 Colorado Female    4.95   348.   69.4 Obese  
## 5 Colorado Female    4.66   302.   67.9 Obese  
## 6 Colorado Male      4.95   340.   67.7 Obese
```

#### Question 5

Use a dplyr method to remove the height, weight, and BMI columns from your data.

```
# colnames(data)  
  
data <- data %>% select(location, gender, bmi_cat)  
  
head(data)  
  
## # A tibble: 6 x 3  
##   location gender bmi_cat  
##   <chr>    <chr> <chr>  
## 1 Colorado Female Obese  
## 2 Colorado Female Obese  
## 3 Colorado Male   Obese  
## 4 Colorado Female Obese  
## 5 Colorado Female Obese  
## 6 Colorado Male   Obese
```

## Optional Challenge

Perform all the actions in this problem set with one dplyr call.

```
data <- read_csv("~/phw251_fall124_cv/week 5/500_Person_Gender_Height_Weight.csv") %>%
  rename(
    location = `Location information`,
    gender = GENDER
  ) %>%
  mutate(
    bmi = if_else(
      location %in% c("United Kingdom", "Taiwan"),
      weight/((height*0.01)^2),
      703 * weight/((height*12)^2)
    )
  ) %>%
  mutate(
    bmi_cat =
      case_when(
        bmi < 18.5 ~ "Underweight",
        bmi <= 24.9 ~ "Normal",
        bmi <= 29.9 ~ "Overweight",
        TRUE ~ "Obese"
      )
  ) %>%
  arrange(location, desc(bmi)) %>%
  select(location, gender, bmi_cat)
```

```
## Rows: 500 Columns: 4
## -- Column specification -----
## Delimiter: ","
## chr (2): Location information, GENDER
## dbl (2): height, weight
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
head(data)
```

```
## # A tibble: 6 x 3
##   location gender bmi_cat
##   <chr>    <chr> <chr>
## 1 Colorado Female Obese
## 2 Colorado Female Obese
## 3 Colorado Male   Obese
## 4 Colorado Female Obese
## 5 Colorado Female Obese
## 6 Colorado Male   Obese
```