

Problem Set 4

NAME HERE

9/17/2020

- Due date: Monday, September 28

Today you will tidy up a dataset of 500 individuals that provides their height and weight. Our goal is to:

Load your data:

Question 1

Clean the column headers to have all lower case and no spaces and rename location_information to location.

```
bmi_1 <- rename_with(bmi, ~ tolower(gsub(" ","_", .x, fixed=TRUE)))  
bmi_1 <- rename(bmi_1, location = location_information)
```

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()`.

```
bmi_2 <- bmi_1 %>%  
  mutate(bmi = if_else(  
    location %in% c("New York", "Colorado", "Hawaii"),  
    (703 * weight) / (height * 12)^2,  
    (weight / (height / 100)^2)))  
head(bmi_2$bmi)
```

```
## [1] 31.69571 24.35542 32.13784 27.35043 27.47624 29.11453
```

Question 3

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

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

```
bmi_3 <- bmi_2 %>%  
  mutate(bmi_cat = case_when(  
    bmi > 30 ~ "Obese",  
    bmi > 25 ~ "Overweight",  
    bmi > 18.4 ~ "Normal",  
    TRUE ~ "Underweight"))
```

Could we have used `if_else()`?

Yup, we could have! It's a matter of preference and how you approach the problem.

Question 4

Arrange your data by location and descending order of bmi.

```
bmi_4 <- bmi_3 %>%  
  arrange(location, desc(bmi))
```

Question 5

Filter your data to remove the height, weight, and bmi column.

```
bmi_5 <- bmi_4 %>% select(-c(height, weight, bmi))
```

Challenge

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

```
bmi_all <- bmi %>%
  rename_with(~ tolower(gsub(" ","_", .x, fixed=TRUE))) %>%
  rename(location = location_information) %>%
  mutate(bmi = if_else(location %in% c("New York", "Colorado", "Hawaii"),
    (703 * weight)/(height * 12)^2,
    weight/(height/100)^2),
    bmi_cat = case_when(bmi > 30 ~ "Obese",
      bmi > 25 ~ "Overweight",
      bmi > 18.4 ~ "Normal",
      TRUE ~ "Underweight")) %>%
  arrange(location, desc(bmi)) %>%
  select(-c(height, weight, bmi))

head(bmi_all)
```

```
## # A tibble: 6 x 7
##   location gender    x5    x6 condition data bmi_cat
##   <chr>    <chr> <dbl> <dbl> <lgl>    <lgl> <chr>
## 1 Colorado Female  78.8     1 NA      NA     Obese
## 2 Colorado Female  74.6     1 NA      NA     Obese
## 3 Colorado Male    70.1     1 NA      NA     Obese
## 4 Colorado Female  69.4     0 NA      NA     Obese
## 5 Colorado Female  67.9     0 NA      NA     Obese
## 6 Colorado Male    67.7     0 NA      NA     Obese
```