Week 6 Tutorial Worksheet

AY22/23 Semester 2

Question 1. International visitor arrivals in Singapore

The file `tourist_arrivals_lengths_of_stay_2022.xlsx` was downloaded from the Singapore Department of Statistics. Read the data into `R`. Clean up the data, and then create a bar plot that shows the distribution of tourist arrival length in December 2022.

Question 2. YRBSS questionnaires

In this question, we use a subset of data from the Youth Risk Behavior Surveillance System (YRBSS). For more information about the data, read the documentation here. We will use the data to practice our data manipulation skills using `tidyverse` syntax.

```
library(tidyverse)
yrbss = read.csv("../data/yrbss_data.csv", stringsAsFactors = TRUE)
```

1. Use the code above to read data into `R`. Then conduct the following tasks:
    - Remove rows with missing values (if any).
    - Remove duplicated rows (if any).
    - Rename the columns `record` as `id`, and `stweight` as `weight_kg`.
    - Store the resulting data frame as an object named `yrbss`.
2. Subset female youth with BMI lower than 15 and then extract the following columns: `id`, `age`, `race4`. Store your result in an object named `qn2_2`.
3. What do the following commands do? Briefly describe (in words) what they do in an `Rmd` text section entitled "Question 2.3".

```
yrbss %>% select(-(grade:race7))
yrbss %>% select(starts_with("race"))
yrbss %>% select(id, sex, everything())
yrbss %>% select_if(is.numeric)
```

4. Create a column `height_m` in `yrbss`, computed based on the BMI formula:

$$BMI = \frac{weight(kg)}{height^2(m)}$$

5. Briefly describe (in words) what they do in an `Rmd` text section entitled "Question 2.5".

```
yrbss %>% mutate(grade_num = as.numeric(str_remove(grade, "th")))
yrbss %>% mutate_if(is.factor, tolower)
yrbss %>% mutate_if(is.numeric, round, digits = 0)
yrbss %>% mutate_at(vars(age:grade), toupper)
```

6. Use `yrbss` to compute the mean and standard deviation of BMI, named `bmi_mean` and `bmi_sd` respectively, for individuals in different age groups. Save your output as a data frame named `qn2_6`.

Requirements

- Clear out your environment and then knit the file again. Make sure your `Rmd` can knit to HTML.
- The code in your `Rmd` file should create three data frames: `yrbss`, `qn2_2`, `qn2_6`.
- The knitted HTML file should contain
  - a bar plot for Question 1.
  - two markdown text sections entitled "Question 2.3" and "Question 2.5" (instead of comments inside code chunks).

Important: Approach your tutor at the end of this tutorial if you are unsure about any of the requirements above.