Assignment 1

Due Saturday, Aug 3rd, 11:59 p.m., China Standard Time

Your assignment should be a single Rmd document that produces nice-looking and well-organized output file (html or pdf). This should consist of a header, some code chunks (with **comments as appropriate**), and **informative text between code chunks**. Please have you code chunks run and show the output in the final file.

A numerical summary can be provided using the summary() function (which takes a tibble as input). Unique values of a variable can be seen using the unique() function which takes as input a specific variable of a tibble (e.g., unique(my\_tibble$my\_variable)).

1. Install required software and packages:
   1. RStudio
   2. R collection of packages tidyverse
   3. R package knitr
   4. R package haven
   5. R package readxl
   6. R package janitor
2. The Tooth Growth dataset contains the results of an experiment conducted on Guinea pigs to evaluate the effect of vitamin C supplements on tooth growth.
   1. Read the file tooth\_growth\_data.xlsx into R
   2. How many variables and observations are in this dataset?
   3. What is each variable's type?
   4. Are there any missing data values? If so which variables(s) and how many missing data values?
   5. What is the mean of each numerical variable?
3. The Penguin dataset contains observations made on individual penguins 3 islands in the Palmer Archipelago, Antarctica
   1. Read the file penguin\_data.csv into R
   2. Provide a (numerical) summary of the data.
   3. How many variables and observations are in this dataset?
   4. Are there any missing data values? If so which variables(s) and how many missing data values?
4. Using the Penguin dataset, create new tibbles (each with an informative new name and informative variable names) and produce a (numerical) summary of each:
   1. Male penguins from Dream island having body mass at least 4200g, excluding any rows with missing values. This tibble should not have a sex variable or an island variable (since they should all be the same).
   2. All penguins with bill length, bill width, and flipper length in cm, and with body mass in kg. There should not be redundant information in the table (e.g., if there is a variable with flipper length in cm there should not be another variable with flipper length in any other units).