Instructions

- Clear the environment
- Open a new R Script called `day5_exercise_script` where you will do the exercise and later save in the `day5` project directory.
- Add the purpose of the file and the author [MANDATORY]
- Here are the main activities for this excercise
  1) Load the `rio`, `lubridate`,`epikit`, `janitor`, `infer` and `tidyverse` package
  2) Load the *WHR2018.csv* dataset using the `import` function.
  3) clean the column names to remove spaces
  4) Explore the distribution of all continuous variables using density, bar and boxplots

1 Importing the dataset into R and clean names

- Download the *WHR2018.csv* from your emails and save it in the `Data` folder
- Import the dataset into R using the `import` function
- Clean the all column names to remove spaces betwen the names

2 Subsetting

- Subset the imported dataset and onle keep the following variables:
  - country
  - year
  - freedom_to_make_life_choices
  - confidence_in_national_government
  - positive_affect
  - negative_affect

3 Droping records with missing data

- In all the columns selected above, remove records that have missing data.

4 Scatter plot, correlation coefficient and line of best fit

- Make a scatter plot of `negative_affect` vs. `positive_affect`
- Calculate and interpret the correlation coefficient of `negative_affect` vs. `positive_affect`
- Add a line of best fit. *HINT:* add a layer of `geom_smooth()`

5 Linear regression

- Fit a linear regression model on `negative_affect` vs. `positive_affect`
- Interpret the output

6 Chi square test

- Do a `chi-square` test of facility utilization Pre-covid and during covid
- Use the covid_tab table below
- What do you think of the association?

```
## CHi square covid
covid_tab <- matrix(
  c(3819, 7076, 215, 1261, 2390, 2345),
    ncol=2,  byrow=F ,
  dimnames = list(c("Maternal service utilization", "Ooutpatient Utilization" ,"Inpatient utilization")
                  c("Pre COVID", "During COVID")))

covid_tab
```