

Workshop 5: Transforming, selecting, filtering, and plotting

Author's name here

2022-11-08

- 1 Intro
- 2 Get the assignment repo
- 3 Load and clean the data
 - 3.1 **Step 1:** Verify the type of your variables
 - 3.2 **Step 2:** Convert the physical activity variables
- 4 **Plot 1:** BMI distribution by sex
 - 4.1 **Extract:** Make a subset
 - 4.2 **Plot with Esquisse:** Violin Plot
- 5 **Plot 2:** Male respondents' Light Physical Activity (LPA, in minutes per week)
 - 5.1 **Extract:** Make a data subset
 - 5.2 **Plot with Esquisse:** Histogram of Light Physical Activity (LPA) of Male Respondents
- 6 **Plot 3:** Adults complying to OMS/WHO recommendations' Moderate to Vigorous Physical Activity (MVPA, minutes per week)
 - 6.1 **Extract:** Make a data subset
 - 6.2 **Plot with Esquisse:** Boxplots of Moderate to Vigorous Physical Activity per Age Group
- 7 Submission: Upload HTML

1 Intro

Welcome!

For this workshop, we will be cleaning a dataset. It is a hands-on approach to using the `select()`, `filter()`, and `mutate()`.

The assignment should be submitted individually, but you are encouraged to brainstorm with partners.

The final due date for the assignment is Tuesday, November 15th at 23:59 PM UTC+2.

2 Get the assignment repo

To get started, you should download and look through the assignment folder.

1. First download the repo to your local computer [here](#).

You should ideally work on your local computer, but if you would rather work on RStudio Cloud, you can upload the zip file to RStudio Cloud through the Files pane. Consult one of the instructors for guidance on this.

2. Unzip/Extract the downloaded folder.

If you are on macOS, you can simply double-click on a file to unzip it.

If you are on Windows and are not sure how to “unzip” a file, see [this image](#). You need to right-click on the file and then select “extract all”.

3. Once done, click on the RStudio Project file in the unzipped folder to open the project in RStudio.
4. In RStudio, navigate to the Files tab and open the “rmd” folder. The instructions for your exercise are outlined there (these are the same instructions you see here).
5. Open the “data” folder and observe its components. You will work with the “obesity.csv” file. (You can also open the “00_info_about_the_dataset” file to learn more about this dataset.)

3 Load and clean the data

Now that you understand the structure of the repo, you can load in and clean your dataset.

In the code section below, **load in the needed packages**.

```
## [1] "WRITE_YOUR_CODE_HERE"
```

Now, **read the dataset into R**. The data frame you import should have 142 rows and 9 columns. Remember to use the `here()` function to allow your Rmd to use project-relative paths.

```
## [1] "WRITE_YOUR_CODE_HERE"
```

3.1 Step 1: Verify the type of your variables

Before jumping into wrangling or plotting, you should take the time to know what data types you are working with. You can look at these types with `summary()` or `typeof()`.

Some of your variables are numeric but they should be factors (i.e. categories), some are characters but should be factors, and some are characters but should be numeric. Having them in the correct type will be essential for the next manipulations and for plotting !

Use `mutate()` to convert your variables into the right type.

3.2 Step 2: Convert the physical activity variables

Currently, the variables of physical activity are in seconds per day. There are 3 types of physical activity variables: sedentary physical activity (`sedentary_ap_s_day`), light physical activity (`light_ap_s_day`), and moderate to vigorous physical activity (`mvpa_s_day`).

Please convert these numerical variables in **seconds/day** to **minutes/week**. As a kind reminder, 60 seconds = 1 minute and 7 days = 1 week.

(Hint: use `mutate()` to create new variables that are in minutes per week. If you feel more comfortable changing the variables in-place, that's also acceptable.)

Why do we perform this conversion? The [WHO \(known as OMS in French\) recommendations](#) are in minutes per week, so we want to align with these measures.

You will use this `obesity_data_cleaned` for all the following subset making and plotting.

4 Plot 1: BMI distribution by sex

4.1 Extract: Make a subset

1. Make a subset with only the variables of interest for your plot. **This is good practice to make a subset with the variables you need for plotting.**
2. Print this subset in an elegant manner for your HTML (hint: use `reactable`).

```
## [1] "WRITE_YOUR_CODE_HERE"
```

4.2 Plot with Esquisse: Violin Plot

Using `esquisse` and the subset you just made above, plot BMI distributions by sex, as a violin plot.

Violin plots are interesting because you can compare the density curves' peaks, valleys, and tails to see where the groups are similar or different.

```
## [1] "PASTE_THE_ESQUISSE_CODE_HERE"
```

5 Plot 2: Male respondents' Light Physical Activity (LPA, in minutes per week)

5.1 Extract: Make a data subset

1. To make this subset, you will only male respondents.
2. Then keep only the variables useful for the plot.
3. Print this subset in an elegant manner for your HTML (hint: use `reactable`).

```
## [1] "WRITE_YOUR_CODE_HERE"
```

5.2 Plot with Esquisse: Histogram of Light Physical Activity (LPA) of Male Respondents

Using `esquisse` and the subset you just made above, plot LPA distribution (minutes/week) for male respondents, as a histogram.

```
## [1] "PASTE_THE_ESQUISSE_CODE_HERE"
```

6 Plot 3: Adults complying to OMS/WHO recommendations' Moderate to Vigorous Physical Activity (MVPA, minutes per week)

6.1 Extract: Make a data subset

1. To make this subset, you will only keep individuals in the dataset who have complied to OMS/WHO recommendations

(Hint 1: `oms_recommendation` should be equal to `Yes`. Side-note: OMS is Organisation Mondiale de la Santé, French for WHO.)

(Hint 2: The variable `status` is encoded in French as well. "Adulte" means "Adult" and "Enfant" means "Child".)

2. Then keep only the variables useful for the plot.
3. Print this subset in an elegant manner for your HTML (hint: use `reactable`).

```
## [1] "WRITE_YOUR_CODE_HERE"
```

6.2 Plot with Esquisse: Boxplots of Moderate to Vigorous Physical Activity per Age Group

Using `esquisse` and the subset you just made above, plot MVPA distributions (minutes/week) by age groups, as boxplots.

```
## [1] "PASTE_THE_ESQUISSE_CODE_HERE"
```

7 Submission: Upload HTML

Once you have finished the tasks above, you should **knit this Rmd into an HTML** and **upload it** on the assignment page.