03 Practice clean data

Richard Dien Giau Bui

Our data today

- **Starbuck nutrition** data

- In the `data/raw/starbucks-nutrition.xlsx` file

- Please open the Excel file first to see the data

Clean steps [1]

Clean steps [2]

6. Any column has the wrong data type? If you can find them, let's convert them into the correct data format.

7. The whip column should have only two values: 1 if we added whip cream and 0 otherwise. But we have some other values like 2 or 3. Please change all these typos to 1.

8. Some `size` is not quite popular, let's delete them. Only keep these size: `short`, `tall`, `grande`, `venti`

Clean steps [3]

9. Create a new column `milk_type` as follow:

- If milk = 0 then `milk_type` = "none"
- If milk = 1 then `milk_type` = "nonfat"
- If milk = 2 then `milk_type` = "2%"
- If milk = 3 then `milk_type` = "soy"
- If milk = 4 then `milk_type` = "coconut"
- If milk = 5 then `milk_type` = "whole"

10. Save your final data into the `data/process` folder as the name: `starbucks_clean.rds`

Let's go

- Life needs luck! (Richard, 2024)