R and Tidyverse Entry Test

If you've completed the "Whole game" section from the book *R for Data Science*, these exercises should be straightforward for you. Try solving them—we'll discuss your solutions in class.

Good Luck!

Task 1: Prepare Your Environment

First, let's set up your workspace. If you haven't done so yet, ensure you have the following software installed on your computer:

- A current version of the R programming language
- The latest version of RStudio Desktop
- Git

Once everything is installed:

1. **Open RStudio.**

2. **Open your terminal** within RStudio and navigate to your preferred project directory using the `cd` command.

3. **Clone** the Git repository containing the course materials by running:

   ```
   git clone https://github.com/winf-hsos/data-analytics-code.git
   ```

4. A new folder named `data-analytics-code` will be created. Inside this folder, locate the file named `data-analytics-code.Rproj` and open it in RStudio via **File → Open Project**. Or double-click the file from a file explorer.

5. Within your project, navigate to the `your_scripts` folder and create a new R-script file named `entry_test.R`. At the top of your script, include:

```r
install.packages("tidyverse")
install.packages("janitor")
library(tidyverse)
library(janitor)
```

These commands install and load the necessary libraries for this exercise and future classes.

Task 2: First Steps with R and the Tidyverse

Loading and Preparing the Data

1. Load the REWE product data set from the `data` folder into a tibble named `rewe` using `readRDS()`. The file is called `rewe.rds`.

2. Quickly get an overview of your data set (column names, data types, values).

3. Ensure column names are in `snake_case`. If necessary, rename them.

Data Visualization

4. **Bar Chart:** Visualize how many products exist in each category.

5. **Histogram:** Show the distribution of energy content in kcal across all products.

6. **Box Plots:** Create box plots comparing fat content (in grams) across different product categories.

7. **Stacked Bar Chart:** Visualize the proportion of organic products within each product category.

Data Transformation

8. Identify and print the ten most expensive products (show only product names and prices).

9. List all unique countries from which the products originate.

10. List unique countries again, this time including the number of products from each country, sorted by descending order.

11. Compute a new column representing the difference between total fat and saturated fat.

12. Select and display products containing more than 99 grams of fat. Display only the product name and fat content.

13. Filter your data set to retain only columns with numeric data.

14. Convert the columns `product_id` and `gtin` to the character data type.