

Homework 2

Due February 11, 2025

Homework is not collected in this class. When applicable, we will review homework answers during class.

Assignment Goals:

1. Learn about tables.
2. Practice data manipulations.

Part 1: Reading

Read Chapter 11 of *Better Data Visualizations* and complete [Quiz 2](#). The quiz is open-book (obviously) and due on February 11, 2025 at 5:30PM ET. Quizzes submitted after class begins will not count.

Part 2: Practice Data Manipulations

1. Download the dataset [here](#) and load it into a new notebook using the `pd.read_csv()` function. Make sure to also:
 - look over the dataset's metadata under **About Dataset**
 - import the pandas library as we did in class
 - set the environment the notebook uses to `data502`
 - name the notebook `HW.ipynb`
 - move the notebook into a new folder (in the course folder) named `Homework`
 - move the `.csv` file to the `data` folder, right click the file to get its relative path
2. Recode all the cereal manufacturers to be their full name (e.g. "G" should be replaced with "General Mills", "K" with "Kelloggs", and so forth)
3. Create a dataframe called `df_gm` that contains only cereals manufactured by General Mills.
4. Create a dataframe called `df_gm_k` that contains cereals manufactured by either General Mills or Kelloggs.
5. Create a column called `sodium_grams` equal to `sodium` divided by 1000.
6. Rename the `shelf` column as `display_shelf`.
7. How many cereals have `type == "hot"` ?

At the beginning of next week's class, students will gather in groups to review their answers to the homework. Representatives from each group will present during the full class homework discussion afterwards.