Question 1.

Load in the dataset and have a look at it. What are the dimensions, variable types, variable names, etc.?

- 242x18

- Characters & Doubles

- [1] "Beverage\_category" "Beverage" "Beverage\_prep" "Calories" "Total Fat (g)" "Trans Fat (g)"

[7] "Saturated Fat (g)" "Sodium (mg)" "Total Carbohydrates (g)" "Cholesterol (mg)" "Dietary Fibre (g)" "Sugars (g)"

[13] "Protein (g)" "Vitamin A (% DV)" "Vitamin C (% DV)" "Calcium (% DV)" "Iron (% DV)" "Caffeine (mg)"

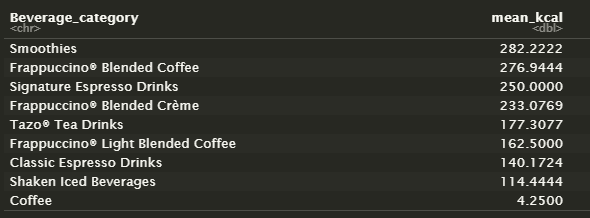
Question 2. Let’s first investigate the calories of different drinks. Select the variables Beverage\_category, Beverage, Beverage prep and Calories from your data. Since we are interested in the calorie content, check if there are any NA values in the data, and drop them if there are.

- Done, None N/A

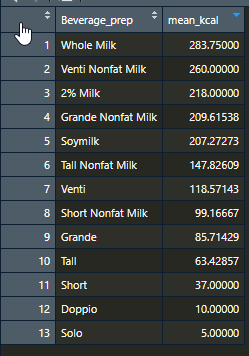
Question 3. Create a new variable (column) called calorie\_diff, which stores the difference between 135 calories (135 calories = 10 cubes of sugar!) and the calories in each drink. (hint: you’ll want to subtract 135 from the calories to see which drink have more or less than 10 cups of sugar).

- Done

Question 4. Summarise the mean number of calories in each beverage\_category. Which 3 category of drinks have the most calories? Which 3 drinks have the least? Write a small summary of your findings.



Question 5. Let’s look at this a different way. What is the average number of calories in each Beverage\_prep type?



Question 6. Which Beverage\_prep type contains more than the average calories of all drinks?

* Whole Milk
* Venti Nonfat Milk

Question 7. Which is the best type of coffee drink to get if you’re worried about consuming too many calories?

Brewed Coffee