Dplyr Homework

The data for this homework comes from an Open Data Science website called Kaggle. Kaggle has many open source datasets for you to use and most of them come with code uploaded by other users showing different ways to explore the data. It is a great way to learn about data-wrangling and analysis and if you are interested then set up your own account and get going.
For this task we will make use of their **Starbucks Calorie** dataset. You can find out more information about each dataset and what each column represents , but we have put the version you need in your data folder.

1 Question 1

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

```
[1] 242  18
 [1] "Beverage_category"       "Beverage"                "Beverage_prep"
 [4] "Calories"                "Total Fat (g)"           "Trans Fat (g)"
 [7] "Saturated Fat (g)"       "Sodium (mg)"             "Total Carbohydrates (g)"
[10] "Cholesterol (mg)"        "Dietary Fibre (g)"       "Sugars (g)"
[13] "Protein (g)"             "Vitamin A (% DV)"        "Vitamin C (% DV)"
[16] "Calcium (% DV)"          "Iron (% DV)"             "Caffeine (mg)"
```

Character and Doubles

2 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.

3 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 the calories from 135 to see which drink have more or less than 10 cups of sugar).

4 Question 4

Summarise the mean number of calories in each beverage_category. Which 3

drinks have the most calories? Which 3 drinks have the least? Write a small summary of your findings.

Beverage_category <chr>	mean_calories <dbl>			
Smoothies	282.2222			
FrappuccinoÂ ® Blended Coffee	276.9444			
Signature Espresso Drinks	250.0000			

Beverage_category <chr>	mean_calories <dbl>			
Classic Espresso Drinks	140.1724			
Shaken Iced Beverages	114.4444			
Coffee	4.2500			

5 Question 5

Let’s look at this a different way. What is the average number of calories in each Beverage_prep type?

Beverage_prep <chr>	mean_bev <dbl>			
Whole Milk	283.75000			
Venti Nonfat Milk	260.00000			
2% Milk	218.00000			
Grande Nonfat Milk	209.61538			
Soymilk	207.27273			

Tall Nonfat Milk	147.82609			
Venti	118.57143			
Short Nonfat Milk	99.16667			
Grande	85.71429			
Tall	63.42857			

Next
12
Previous

6 Question 6

Which Beverage_prep type contains more than the average calories of all drinks?
Hint: to answer this, you'll have to first figure out what the average calories across all drinks are, and then use that as a filter for the grouped Beverage_prep data.

Beverage_prep <chr>	**mean_bev** <dbl>			
Whole Milk	283.7500			
Venti Nonfat Milk	260.0000			
2% Milk	218.0000			
Grande Nonfat Milk	209.6154			

7 Question 7

Which is the best type of **coffee** drink to get if you're worried about consuming too many calories?

Beverage_category <chr>	**Beverage** <chr>	**Beverage_prep** <chr>	**Calories** <dbl>	**calorie_diff** <dbl>
Coffee	Brewed Coffee	Short	3	-132