Tasks: Day 3

Solveig Bjørkholt

1. Make an R Markdown file. Title the document: "Food shortage?"
2. Make a chunk where you `library` in the *tidyverse* package.
3. Download the dataset `global_food_prices.csv` from Canvas. Place it in a folder on your computer and find the path to the file. With this path, read the dataset into R using the function `read_csv`. Assign the dataset to an object called `foodprices`.

The data is gathered from this site, which again has gotten it from the World Food Programme and Humanitarian Data Exchange.

The dataset contains Global Food Prices data from the World Food Programme covering foods such as maize, rice, beans, fish, and sugar for 76 countries and some 1500 markets. The data goes back as far as 1992 for a few countries, although many countries started reporting from 2003 or thereafter. It includes these main variables: country, locality, market, goods purchased, price & currency used, quantity exchanged, and month/year of purchase.

All the names of the variables are given below:

- adm0_id: country id
- adm0_name: country name
- adm1_id: locality id
- adm1_name: locality name
- mkt_id: market id
- mkt_name: market name
- cm_id: commodity purchase id
- cm_name: commodity purchased
- cur_id: currency id
- cur_name: name of currency
- pt_id: market type id
- pt_name: market type (Retail/Wholesale/Producer/Farm Gate)
- um_id: measurement id
- um_name: unit of goods measurement
- mp_month: month recorded
- mpyear: year recorded
- mpprice: price paid
- mp_commoditysource: Source supplying price information

4. How many variables are there in this dataset? How many rows? Comment briefly on the size of the dataset in your R Markdown report. Why is it so big?
5. To do some preliminary tests on the data, you decide to subset only a few observations and variables. First, use `filter` to get all observations where the year equals 2015 and 2020. Write it into a new object that you call `foodprices_subset`. How many observations and variables does this dataset have?

```
_______ <- foodprices %>%
  ____(mp_year %in% c("____", "____"))
```

6. Second, use `select` to fetch the variables `adm0_name`, `mp_year`, `adm1_name`, `mkt_name`, `cm_name`, `cur_name`, and `mp_price`. Overwrite the old object by adding an arrow in and calling the object the same, `foodprices_subset`. How many observations and variables does the dataset have now?

```
foodprices_subset <- __________ %>%
  _____(___, ___, ___, ___, ___, ___, ___)
```

7. Give the variables some names that work better for you to remember what they mean. You are free to choose which names you want, take a look over in the document to see what the different variables contain of information. Below is an example of some new names for the variables:

- adm0_name - country
- mp_year - year
- adm1_name - locality
- mkt_name - market
- cm_name - commodity
- cur_name - currency
- mp_price - price

Remember that when you use `rename`, the new name comes before the old name.

```
foodprices_subset <- foodprices_subset %>%
  _____(___ = adm0_name,
        ____ = mp_year,
        ____ = adm1_name,
        ____ = mkt_name,
        ____ = ____,
        ____ = ____,
        ____ = ____)
```

8. What is the extent of missing values (`NA`) in our dataset? Use `is.na` and `table` as in the syntax below to figure it out.

```
foodprices_subset %>%
  ____() %>%
  ____()
```

9. Use some maths in R to figure out what the percentage of missing values is. Recall that percentages are calculated by the number of observations that are missing, divided by all the observations, and multiplied by one hundred. Write the number in your report.

```
___/(___+___)*100
```

10. Use `group_by` and `summarise` to figure out what the sum of the prices for food was in each country for each year. Recall the the function used to find the sum is `sum`, and that to avoid trouble because of missing values, you have to add `na.rm = TRUE`.

```
foodprices_subset %>%
  ____(country, year) %>%
  ____(food = ___(price, ____))
```

11. What was the total price for food for Kenya? Use the code you wrote above and add a row using `filter` to figure it out.

```
foodprices_subset %>%
  ____(country, year) %>%
  ____(food = ___(price, ____)) %>%
  ____(country == "____")
```

12. What's the average price for food for all countries in 2015 and 2020 respectively? Remember that the function `mean` gives you the average.

```
foodprices_subset %>%
  group_by(____) %>%
  _____(foodprice = ____(price, ____ = TRUE))
```

13. What does the code below do? Comment each line to explain what the different lines do. Remember that comments are made by setting a hashtag in the code chunk and writing your comment after that; # comment

Which country traded most spices in our dataset?

```
foodprices_subset %>%
  group_by(country, year) %>%
  count(commodity, name = "commodity_number") %>%
  mutate(spices_commodity = ifelse(commodity %in% c("Salt - Retail", "Sugar - Retail"),
                                   "spice",
                                   "other")) %>%
  filter(spices_commodity == "spice") %>%
  ungroup() %>%
  arrange(desc(commodity_number))
```