Week 5 Tutorial Worksheet

AY22/23 Semester 2

Instructions

1. Answer all questions in a single `R` Markdown file. Make sure your code generates the required objects.
2. Use headers to divide your code into appropriate, titled sections and subsections.
3. Remember to use the usual relative path settings. Make sure your `R` Markdown file can knit to HTML.

Question 1. Retrenchment by industry

In this question, we would like to practice how to obtain data from online APIs. We want to retrieve the data set **Retrenchment by Industry (Level 1)** from the following website:

https://data.gov.sg/dataset/retrenched-employees-by-industry-and-occupational-group-quarterly?view_id=c51ab6a4-6af8-4334-8061-e2dfc39a06ea&resource_id=3d180571-81d3-4834-a759-8374806b731e

1. Use the Data API link on the web page to download the full data set by its resource id. *Hint:* The first few observations of your data frame should look like the following.

```
##   _id retrench retrench_term_contract quarter retrench_permanent     industry1
## 1   1     6170                   1060 1998-Q1               5110 manufacturing
## 2   2      560                    480 1998-Q1                 90  construction
## 3   3     2100                    160 1998-Q1               1940      services
```

- While querying the full data set, think about the following:
  - How many rows of data have we retrieved?
  - How many rows should the final data set contain?

2. Remove the `_id` column and name the new data frame as `df_retrench`.

3. Continue to work with `df_retrench`. Convert `retrench`, `retrench_term_contract`, `retrench_permanent` to numeric. Convert `industry1` to factor.

4. Keep only the observations in the year of 2020. There are missing values in the data frame. Replace the missing entries with zero. Store it in a new object named `df_retrench2020`.

5. Explore the data by yourself. **Create a graph** to answer one question you find interesting about the data. Include the code you use, and summarize (in words) what you found.

Question 2 Hawkers data set

The file `hawker_ctr_raw.rds` contains information on hawker centers in Singapore, retrieved from the OneMap website. We first encountered this data set in Week 1.

Read the data into `R` and name it `hawk`. The following command will return a list of length 116. Each component of that list will contain information on a hawker center. Inspect the structure to verify this.

```
hawk = readRDS("../data/hawker_ctr_raw.rds")
hawk = hawk[[1]][-1]
```

1. Write a function that works in the following way: Given two hawker center names, it computes the point-to-point distance between them using the `XY` coordinates:

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

For example, given the two hawker center names below, the function returns:

```
compute_dist("Tanglin Halt Market", "Taman Jurong Market & Food Centre")
```

```
## [1] 9257.798
```

2. Run the `compute_dist()` function and generate a data frame that contains **all** $\binom{116}{2}$ combinations of hawker center names and the distance between them. Save the data frame as an object named `dist_df`. The first few rows of the data frame looks like:

```
head(dist_df)
```

```
##                                hawker1                    hawker2      dist
## 1 Blks 1A/ 2A/ 3A Commonwealth Drive  Blks 20/21 Marsiling Lane 16106.573
## 2 Blks 1A/ 2A/ 3A Commonwealth Drive Blks 221A/B Boon Lay Place 10722.325
## 3 Blks 1A/ 2A/ 3A Commonwealth Drive   Blks 22A/B Havelock Road  3764.473
## 4 Blks 1A/ 2A/ 3A Commonwealth Drive   Blks 79/79A Circuit Road 10141.455
## 5 Blks 1A/ 2A/ 3A Commonwealth Drive   Blks 91/92 Whampoa Drive  6780.956
## 6 Blks 1A/ 2A/ 3A Commonwealth Drive         Bukit Timah Market  5016.367
```

Requirements

1. Your code should create the following objects:

    - `df_retrench`
    - `df_retrench2020`
    - `hawk`
    - `dist_df`
    - A function called `compute_dist()`

2. Your `Rmd` file should also include

    - One graph to answer the question you explored in Question 1 part 5, followed by some discussion/summary in words.