Week 12 Tutorial Worksheet

AY23/24 Semester 2

No submission required

Demographic structure in Singapore

In this question, we will use a data set on population distribution in Singapore from the 2015 General Household Survey. You can download the raw data directly from SingStat via: https://www.singstat.gov.sg/-/media/files/publications/ghs/ghs2015/excel/t1-9

The file contains multiple spreadsheets. We will start by working with the one named **T7 (Total)**. For questions (1)-(3), we would like to extract information on total population in each planning area by age groups. For questions (4) and (5), we will use information in two additional spreadsheets, **T7 (Male)** and **T7 (Female)**.

1. Read the data into `R`. Keep only data at planning-area level, convert the population sizes in each planning area into numeric, and remove unwanted rows/columns. After that, save the resulting data frame as `qn1_1`.

   The structure of `qn1_1` should resemble the following:

```
str(qn1_1)
```

```
## tibble [55 x 20] (S3: tbl_df/tbl/data.frame)
##  $ Planning Area: chr [1:55] "Ang Mo Kio" "Bedok" "Bishan" "Boon Lay" ...
##  $ Total        : num [1:55] 174770 289750 90700 30 139270 ...
##  $ 0 - 4        : num [1:55] 6790 11690 3430 NA 5510 ...
##  $ 5 - 9        : num [1:55] 7660 13400 4330 NA 6890 7480 7630 4660 NA 230 ...
##  $ 10 - 14      : num [1:55] 8290 14750 4710 NA 7970 ...
##  $ 15 - 19      : num [1:55] 9320 16930 5520 NA 9240 ...
##  $ 20 - 24      : num [1:55] 10310 19450 6860 NA 10070 ...
##  $ 25 - 29      : num [1:55] 11170 19860 6460 NA 10460 ...
##  $ 30 - 34      : num [1:55] 12250 19270 5720 NA 9760 ...
##  $ 35 - 39      : num [1:55] 13070 20850 6000 NA 9940 ...
##  $ 40 - 44      : num [1:55] 13710 22520 7070 NA 10980 ...
##  $ 45 - 49      : num [1:55] 13000 21460 6800 NA 11310 ...
##  $ 50 - 54      : num [1:55] 14010 23430 7540 NA 12190 ...
##  $ 55 - 59      : num [1:55] 13800 23380 7700 NA 11770 ...
##  $ 60 - 64      : num [1:55] 12980 20590 6360 NA 9370 ...
```

1

```
##  $ 65 - 69      : num [1:55] 11050 16750 4860 NA 5990 ...
##  $ 70 - 74      : num [1:55] 6670 9310 2730 NA 3060 6480 2740 2380 NA 50 ...
##  $ 75 - 79      : num [1:55] 5140 7330 2140 NA 2230 5590 2000 1880 NA 30 ...
##  $ 80 - 84      : num [1:55] 3250 4760 1370 NA 1450 3540 1290 1210 NA 30 ...
##  $ 85 & Over    : num [1:55] 2300 4010 1090 NA 1090 3010 1010 1000 NA 40 ...
```

2. Find the top four planning areas with the largest total resident populations in 2015. Save the names of these areas as a vector called `qn1_2`.

3. The data frame `qn1_1` is currently in wide format. Convert it from wide to long and compute the population proportion by age cohort and planning area. For each planning area, the population proportion for the j-th cohort can be derived as:

$$pop\_prop_j = \frac{population_j}{\sum_j population_j}$$

Save the resulting data frame as `qn1_3`. After that, recreate, as much as you can, the graph below on the top 4 planning areas with the largest total resident populations in 2015.
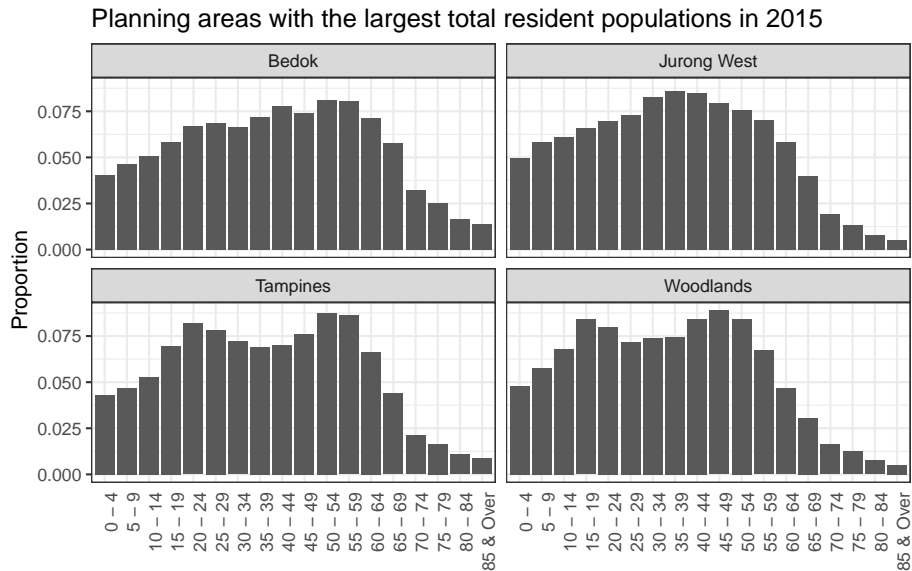
Source: Singapore Department of Statistics

4. In spreadsheets **T7 (Males)** and **T7 (Females)**, you will find information about the total population by age group, gender, and region. Read the data into R first.

Suppose we wish to show the demographic age structure for female and male residents in Singapore in 2015. Your task is to prepare the data and create a plot that effectively serves this purpose using `ggplot2` functions.

5. The file `sg_masterplan2019.rds` contains data on Singapore's planning area boundaries. During lectures, we used it to generate a map of Singapore. To work with this data, you need to install and load the `sf` package, which is designed for simple features (`sf`) object.

   Use the population data to create a map that displays the total size of population in each planning region.

Requirements

- You code should generate data frames named `qn1_1`, `qn1_2`, and `qn1_3`.

- The knitted HTML should contain three plots, one each for Question 1.3, Question 1.4, and Question 1.5.