population_pyramid_D

TeamD

3/9/2022

For this exercise, we will need to load the following packages:

```
library(tidyverse)
library(readxl)
tinytex::install_tinytex()
```

(1) Open XLSX file

```
sg <- read_excel("outputFile.xlsx", sheet = "T2", skip = 10, na = "na") |>
  slice(27:50, 52:75)
```

```
Missing values appear as 'NA' values in the spreadsheet.
```

(2) Remove columns for all other years and change name of column with age group

```
sg <- sg |>
  select(age = "Data Series", as.character(seq(1960, 2020, 10)))
```

(3) Change strings

```
sg <- mutate(
  sg,
  age = str_replace(age, "(.*) Years & Over$", ">\\1") |>
    str_remove(" Years") |>
    (\(x) {
      factor(x, levels = unique(x))
    })()
)
```

(4) Append a column to `sg` that indicates `sex`

```
sg <- sg |>
  mutate(sex = as.factor(rep(c("male", "female"), each = nrow(sg)/2)))
```

(5) Change age brackets

```
# Automated function
repl_col <- function(d) {
  d <- as.data.frame(d)
  x <- 0
  for (i in 2:ncol(d)) { # For each column, the code will look through each row, thus the double for-lo
    for (k in 1:nrow(d)) {
      if (is.na(d[k, i])) { # If the cell is NA, the code will look for a non-NA cell who's age group f
        for (j in k:nrow(d)) {
          if (grepl(str_remove(d[j, 1], ">"), str_remove(d[k, 1], "-"), fixed = TRUE)) {
            d[k, i] <- d[j, i]
            x <- j
            break # if the corresponding age group is found, the NA cell will be replaced by the value
          }
        }
        k <- x # since the value is found, the code will start checking for NA values after that non-NA
      }
    }
  }
  return(as_tibble(d))
}

sg_auto <- repl_col(sg)

# Hard code
# Male
sg[18, "1990"] <- sg[23, "1990"]
sg[18, "1980"] <- sg[23, "1980"]
sg[17, "1970"] <- sg[22, "1970"]
sg[15, "1960"] <- sg[20, "1960"]

# Female
sg[42, "1990"] <- sg[47, "1990"]
sg[42, "1980"] <- sg[47, "1980"]
sg[41, "1970"] <- sg[46, "1970"]
sg[39, "1960"] <- sg[44, "1960"]

all.equal(sg_auto, sg)
```

```
## [1] TRUE
```

(6) Remove rows that do not correspond to given age groups

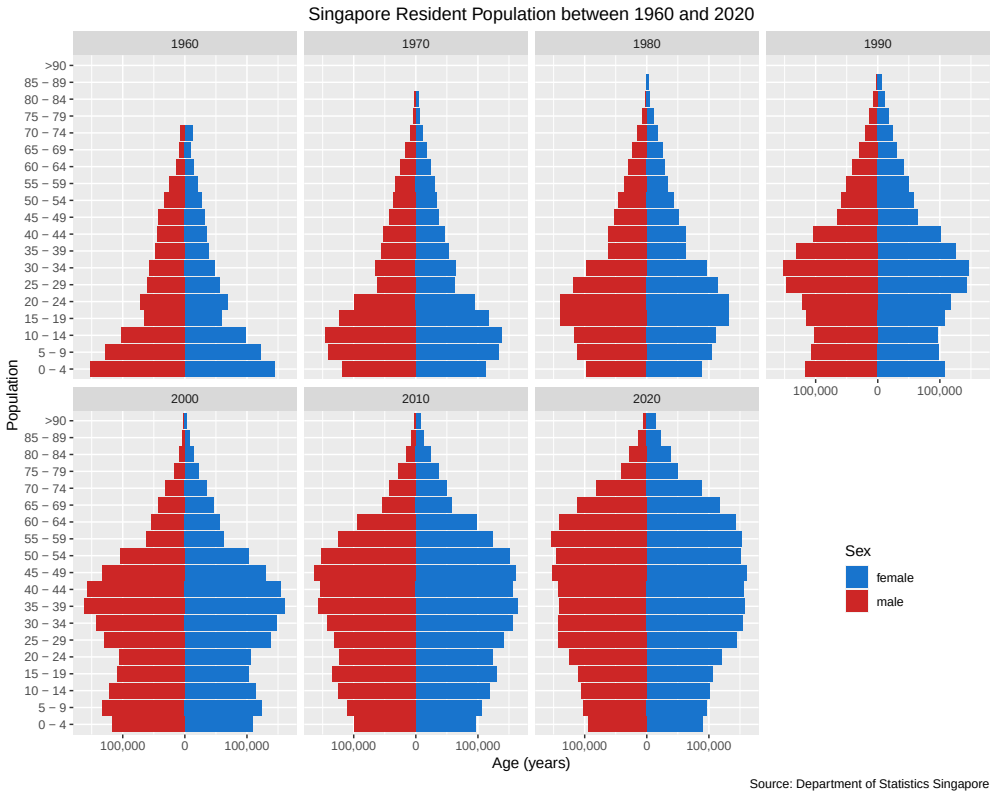
```
sg <- sg |>
  filter(!(age %in% c(">65", ">70", ">75", ">80", ">85")))
```

(7) Pivot the data

```
sg <- sg |>
  pivot_longer(
    as.character(seq(1960, 2020, 10)),
    names_to = "year",
    values_to = "pop"
  )
```

(8) Make Faceted Population Pyramid

```
sg |>
  filter(!is.na(pop)) |>
  mutate(pop2 = if_else(sex == "male", pop * -1, pop)) |>
  ggplot(aes(x = pop2, y = age)) +
  geom_col(aes(fill = sex), na.rm = TRUE) +
  scale_fill_manual(values = c("dodgerblue3", "firebrick3")) +
  facet_wrap(~year, ncol = 4, nrow = 2) +
  scale_x_continuous(
    breaks = c(-100000, 0, 100000),
    labels = c("100,000", "0", "100,000")
  ) +
  labs(
    x = "Age (years)",
    y = "Population",
    title = "Singapore Resident Population between 1960 and 2020",
    caption = "Source: Department of Statistics Singapore",
    fill = "Sex"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = c(0.88, 0.23)
  )
```



(9) Comment on the plot. What does the plot tell us about the development of Singapore's resident population?

The plot tells us that the proportion of the adults in Singapore's resident population is increasing over time, indicative of an ageing population. The proportion of children is decreasing in the population, possibly due to declining reproduction rates.