

lab1

Yovna Junglee

2023-01-12

Lab 1

1.

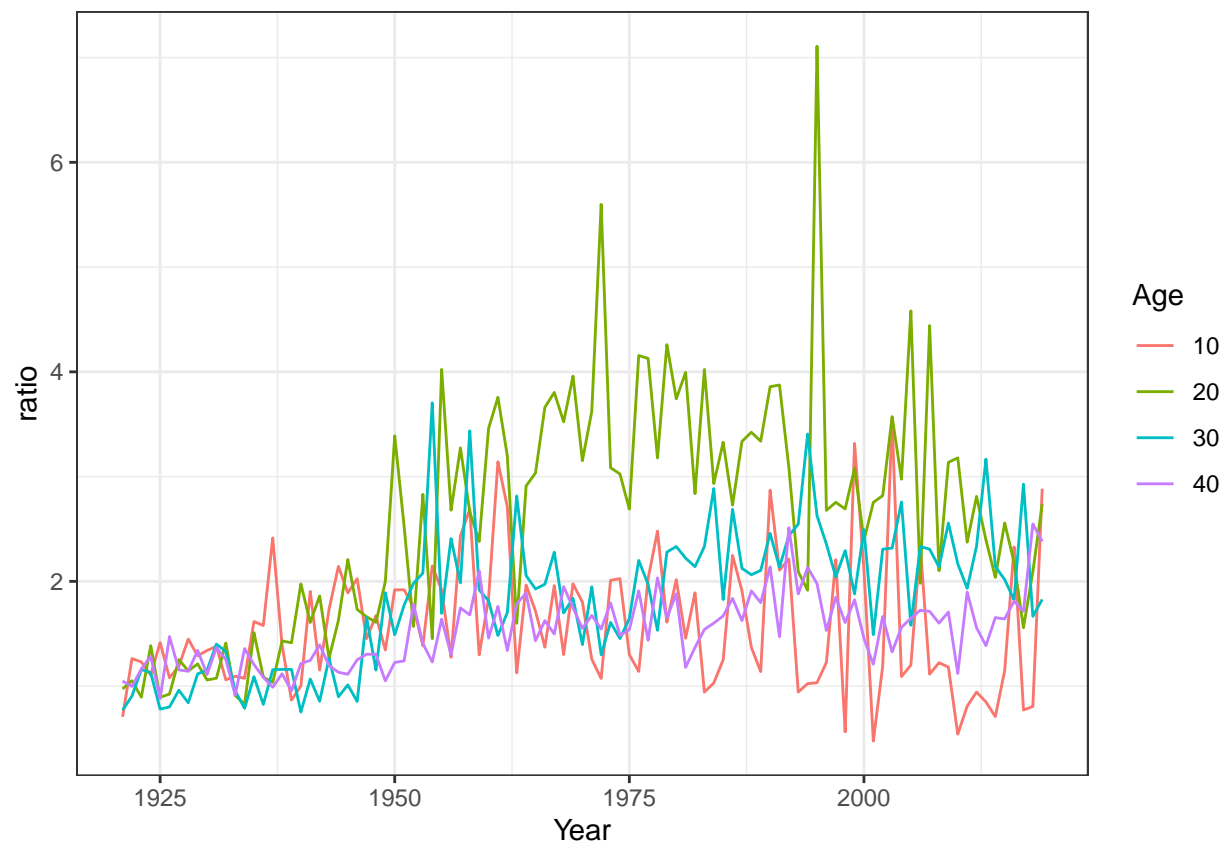
Plot the ratio of male to female mortality rates over time for ages 10,20,30 and 40 (different color for each age) and change the theme

```
library(tidyverse)

# Load data set
dm <- read_table("https://www.prnh.umontreal.ca/BDLC/data/ont/Mx_1x1.txt",
                 skip = 2, col_types = "dfddd")
#head(dm)

ratio_fm <- dm %>% filter(Age %in% c(10,20,30,40)) %>% mutate(ratio = Male/Female)

ratio_fm %>% ggplot(aes(x = Year, y = ratio, colour = Age)) + geom_line() + theme_bw()
```



2.

Find the age that has the highest female mortality rate each year

```
dm %>% group_by(Year) %>% filter(Female == max(Female, na.rm = T)) %>% select(Year, Age)
```

```
## # A tibble: 102 x 2
## # Groups:   Year [99]
##   Year Age
##   <dbl> <fct>
## 1 1921 106
## 2 1922 98
## 3 1923 104
## 4 1924 107
## 5 1925 98
## 6 1926 106
## 7 1927 106
## 8 1928 104
## 9 1929 104
## 10 1930 105
## # ... with 92 more rows
```

3.

Use the `summarize(across())` syntax to calculate the standard deviation of mortality rates by age for the Male, Female and Total populations.

```
dm %>% group_by(Age) %>% summarise(across(c(Male, Female, Total), sd, na.rm = T))
```

```
## # A tibble: 111 x 4
##   Age      Male      Female      Total
##   <fct>    <dbl>    <dbl>    <dbl>
## 1 0      0.0330    0.0256    0.0294
## 2 1      0.00396   0.00352   0.00374
## 3 2      0.00175   0.00154   0.00164
## 4 3      0.00127   0.00113   0.00120
## 5 4      0.000987  0.000925  0.000947
## 6 5      0.000820  0.000748  0.000776
## 7 6      0.000849  0.000631  0.000731
## 8 7      0.000749  0.000590  0.000664
## 9 8      0.000693  0.000496  0.000590
## 10 9      0.000604  0.000473  0.000530
## # ... with 101 more rows
```

4.

The Canadian HMD also provides population sizes over time (<https://www.prhd.umontreal.ca/BDLC/data/ont/Population.txt>). Use these to calculate the population weighted average mortality rate separately for males and females, for every year. Make a nice line plot showing the result (with meaningful labels/titles) and briefly comment on what you see (1 sentence). Hint: `left_join` will probably be useful here.

```
dsize <- read_table("https://www.prhd.umontreal.ca/BDLC/data/ont/Population.txt",
                    skip = 2, col_types = "dfddd")

colnames(dsize)[3:5] <- c("PopF", "PopM", "PopTotal")
```

```

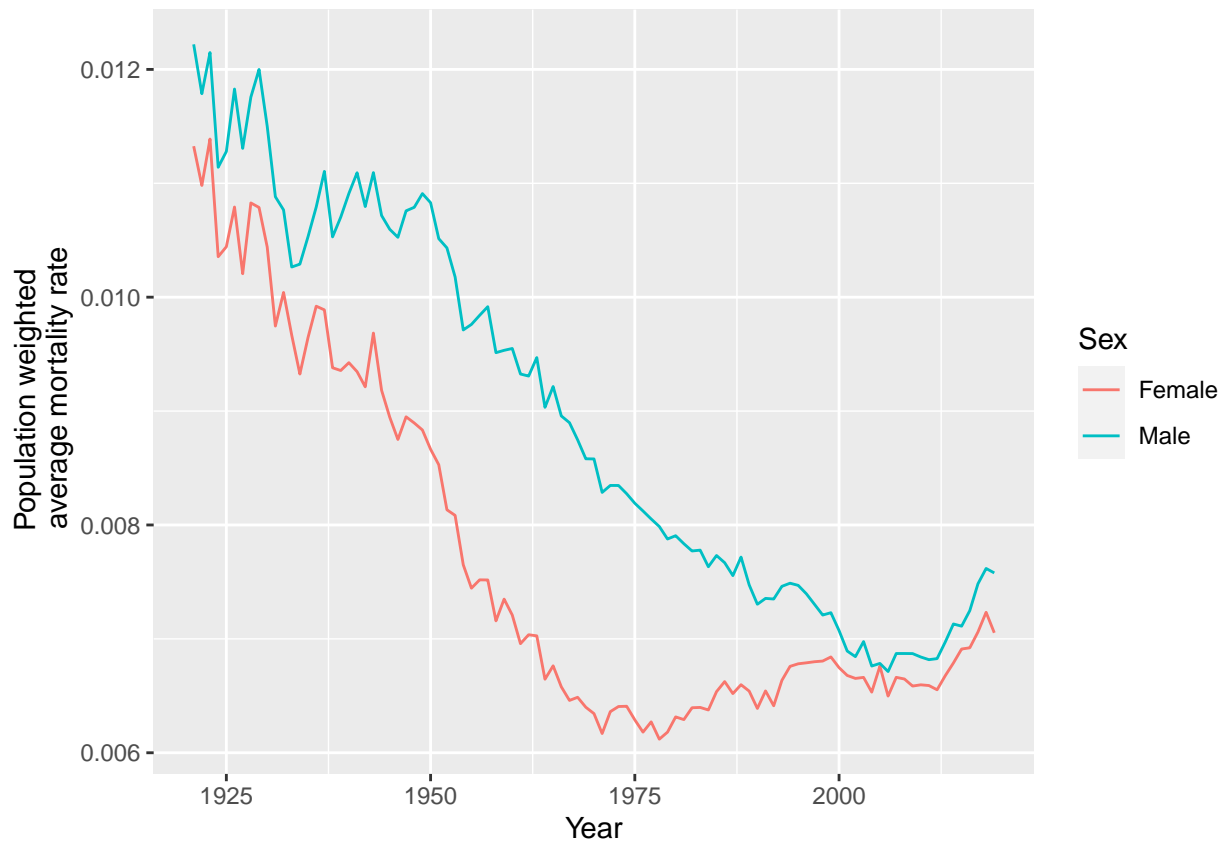
dtot <- left_join(dm %>% select(Male, Female, Year, Age), dsize %>% filter(Year!=2020),
  by = c("Year", "Age")) %>% mutate(DeathsF = (Female*PopF),
  DeathsM = (Male*PopM)) %>%

  group_by(Year) %>%
  summarise(across(c(DeathsF, DeathsM, PopF, PopM), sum, na.rm = T))

dw <- dtot %>% mutate(Female = DeathsF/PopF, Male = DeathsM/PopM)

dw %>% pivot_longer(c(Female, Male), names_to = "Sex", values_to = "Weighted") %>%
  ggplot(aes(x = Year, y = Weighted, colour = Sex)) + geom_line() +
  labs(y = "Population weighted \n average mortality rate")

```



We note that the population weighted average mortality rate has decreased drastically since 1921 but the mortality rates in the male population remain higher than the female population every year.