Lab 1: Intro to Quarto and Tidyverse

TJ Ayoub

01/11/23

Lab Exercises

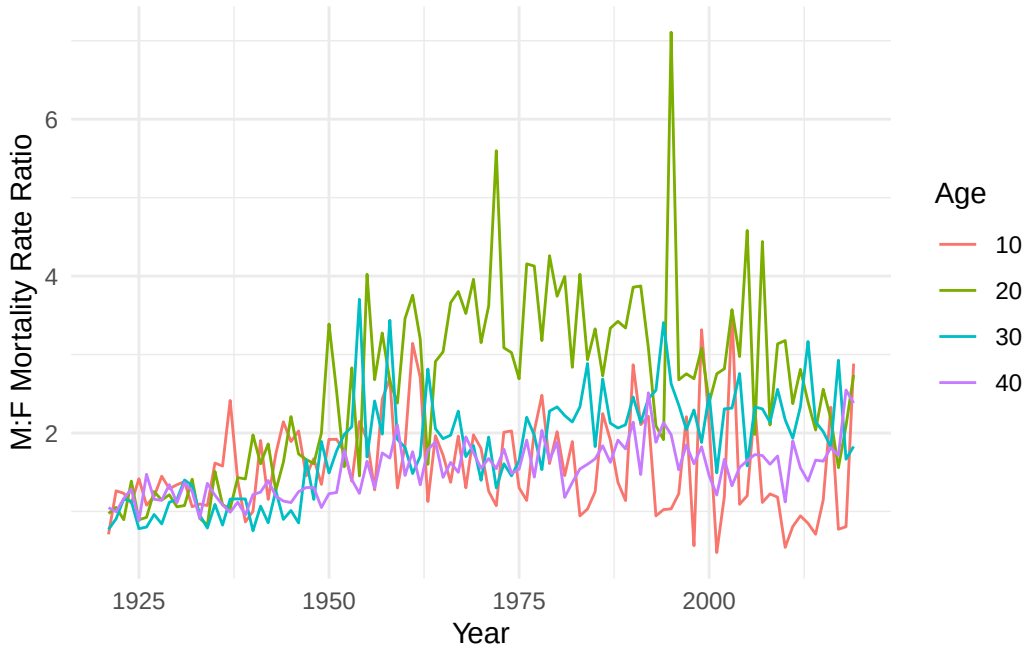
```
library(tidyverse)

dm <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.txt",
                 skip = 2, col_types = "dcddd")
```

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

```
p1 = dm |>
      filter(Age==10 | Age==20 | Age==30 | Age==40) |>
      mutate(mf_ratio = Male/Female) |>
      ggplot(aes(x=Year, y=mf_ratio, color=Age)) + geom_line()

p1 + ylab("M:F Mortality Rate Ratio") + theme_minimal()
```



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

```
two = dm |>
        group_by(Year) |>
        slice_max(Female) |>
        select(Year:Age)

# Only a snippet shown for the first 10 years
two[1:10,]
```

```
# A tibble: 10 x 2
# Groups:   Year [10]
    Year Age
   <dbl> <chr>
 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
```

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

```
three = dm |>
        group_by(Age) |>
        summarize(across(c("Male","Female","Total"), sd))

three[1:10,]
```

```
# A tibble: 10 x 4
   Age        Male    Female     Total
   <chr>     <dbl>     <dbl>     <dbl>
 1 0      0.0330    0.0256    0.0294
 2 1      0.00396   0.00352   0.00374
 3 10     0.000561  0.000474  0.000509
 4 100    0.138     0.0928    0.0729
 5 101    0.158     0.125     0.0995
 6 102    0.214     0.143     0.114
 7 103    0.371     0.252     0.208
 8 104    NA        0.449     0.363
 9 105    NA        NA        NA
10 106    NA        NA        NA
```

4. The Canadian HMD also provides population sizes over time (https://www.prdh.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.

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

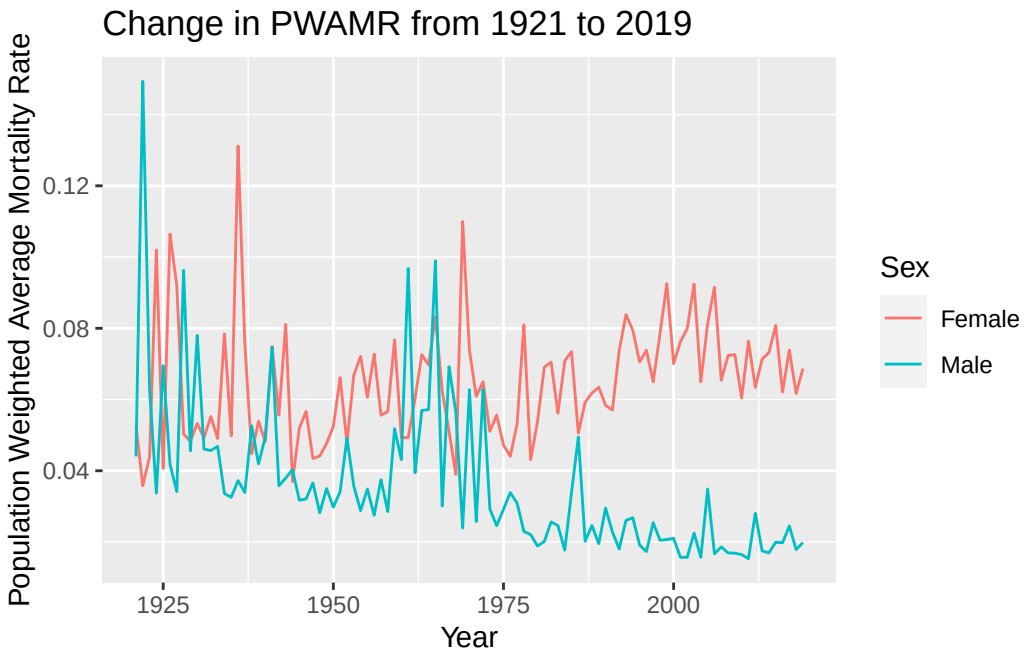
```r
dj = left_join(dm, dp, by = c("Year","Age"))

four = dj |>
        mutate(weightedMortFem = Female.x * Female.y / (Female.y + Male.y),
               weightedMortMale = Male.x * Male.y / (Female.y + Male.y)) |>
        group_by(Year) |>
        summarize(popWeightFem = mean(weightedMortFem, na.rm=TRUE),
                  popWeightMale = mean(weightedMortMale, na.rm=TRUE))

four[1:10,]
```

```
# A tibble: 10 x 3
    Year popWeightFem popWeightMale
   <dbl>        <dbl>         <dbl>
 1  1921       0.0528        0.0440
 2  1922       0.0358        0.149
 3  1923       0.0439        0.0629
 4  1924       0.102         0.0338
 5  1925       0.0407        0.0695
 6  1926       0.106         0.0418
 7  1927       0.0920        0.0342
 8  1928       0.0503        0.0963
 9  1929       0.0483        0.0456
10  1930       0.0533        0.0780
```

```r
four |>
  pivot_longer(popWeightFem:popWeightMale, names_to = "Sex", values_to = "PWAMR") |>
  ggplot(aes(x=Year, y=PWAMR, color=Sex)) + geom_line() +
    ylab("Population Weighted Average Mortality Rate") +
    ggtitle("Change in PWAMR from 1921 to 2019") +
    scale_color_hue(labels=c('Female', 'Male'))
```



We see a clear divergence in the average mortality rate when taking into account population size, between males and females.