Lab Exercise 1

Priyanka Verma

16/01/23

Table of contents


```
#install.packages("tidyverse")
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
#library(ggplot2)
```

Read Data

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

```
Warning: 494 parsing failures.
row    col               expected actual
108 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
109 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Male   no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Total  no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
... ...... ..................... ...... ...............................................
See problems(...) for more details.
```

```r
head(dm)
```

```
# A tibble: 6 x 5
   Year Age    Female    Male   Total
  <dbl> <chr>   <dbl>   <dbl>   <dbl>
1  1921 0     0.0978  0.129   0.114
2  1921 1     0.0129  0.0144  0.0137
3  1921 2     0.00521 0.00737 0.00631
4  1921 3     0.00471 0.00457 0.00464
5  1921 4     0.00461 0.00433 0.00447
6  1921 5     0.00372 0.00361 0.00367
```

Q1

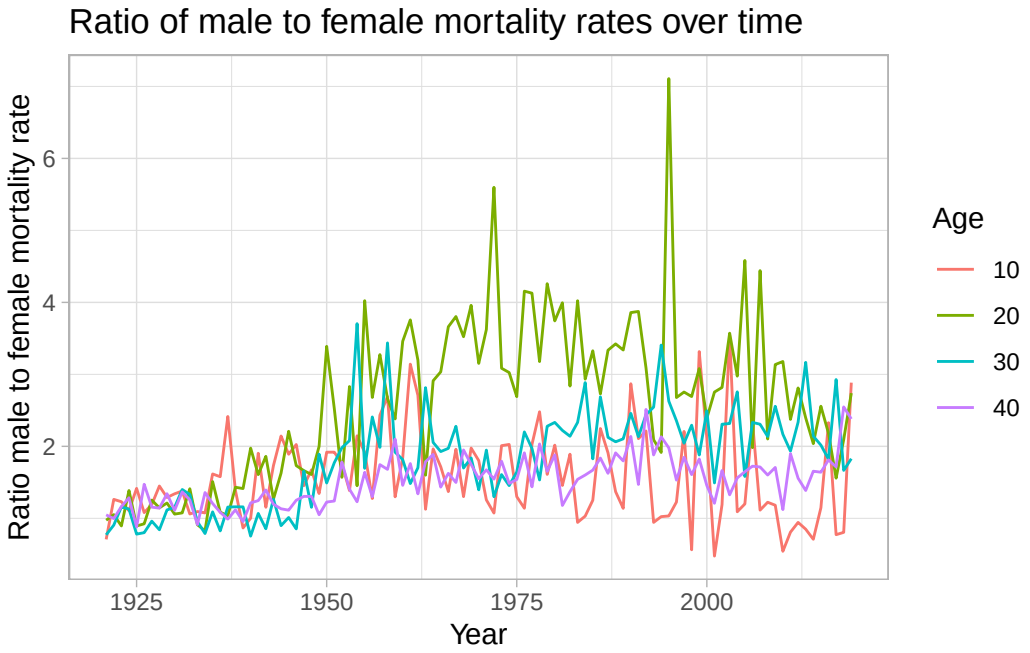
Plot the ratio of male to female mortality rates over time for ages 10,20,30 and 40 (different color for each age) and change the theme

```r
d_1 <- dm |>
  filter(Age==30 | Age==40 | Age==10 | Age==20 ) |>
  mutate(mf_ratio = Male/Female) |>
  select(Year, Age, mf_ratio)
d_1
```

```
# A tibble: 396 x 3
    Year Age   mf_ratio
   <dbl> <chr>    <dbl>
 1  1921 10       0.708
 2  1921 20       0.974
 3  1921 30       0.771
 4  1921 40       1.05
```

```
 5  1922 10        1.26
 6  1922 20        1.05
 7  1922 30        0.906
 8  1922 40        0.996
 9  1923 10        1.23
10  1923 20        0.894
# ... with 386 more rows
```

```
d_1 |>
  ggplot(aes(x=Year, y=mf_ratio, color= Age)) +
  geom_line()+
  ylab("Ratio male to female mortality rate") +
  ggtitle("Ratio of male to female mortality rates over time")+
  theme_light() #change of theme
```



Q2

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

```
d_2 <- dm |>
  select(Year, Age, Female) |>
```

```
    group_by(Year) |>
    mutate(maxF = max(Female, na.rm = TRUE)) |>
    group_by(Year) |>
    filter(Female == maxF) |>
    select(Year, Age, Female) # prints the Age along with the value of the highest Female mo
  d_2
```

```
# A tibble: 102 x 3
# Groups:   Year [99]
    Year Age   Female
   <dbl> <chr>  <dbl>
 1  1921 106    6
 2  1922 98     0.603
 3  1923 104    0.524
 4  1924 107    6
 5  1925 98     0.514
 6  1926 106    4.16
 7  1927 106    6
 8  1928 104    2.13
 9  1929 104    1.32
10  1930 105    6
# ... with 92 more rows
```

Q3

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

```
  d_3 <- dm  |>
    group_by(Age) |>
    summarize(across(Female:Total, sd))
  d_3
```

```
# A tibble: 111 x 4
   Age      Female      Male     Total
   <chr>     <dbl>     <dbl>     <dbl>
 1 0      0.0256    0.0330    0.0294
 2 1      0.00352   0.00396   0.00374
 3 10     0.000474  0.000561  0.000509
 4 100    0.0928    0.138     0.0729
```

```
 5 101    0.125     0.158     0.0995
 6 102    0.143     0.214     0.114
 7 103    0.252     0.371     0.208
 8 104    0.449    NA         0.363
 9 105   NA        NA        NA
10 106   NA        NA        NA
# ... with 101 more rows
```

Q4

The Canadian HMD also provides population sizes over time (https://www.prdh.umontreal.ca/BDLC/data/ont/
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, c
head(dp)
```

```
# A tibble: 6 x 5
   Year Age   Female   Male  Total
  <dbl> <chr>  <dbl>  <dbl>  <dbl>
1  1921 0     30157. 31530. 61687.
2  1921 1     30391. 31319. 61711.
3  1921 2     30962. 31785. 62747.
4  1921 3     31306. 32031. 63336.
5  1921 4     31364. 32046. 63409.
6  1921 5     31175. 31847. 63021.
```

```
df_join <- merge(x = dm, y = dp, by = c("Year", "Age"), all.x = TRUE)
colnames(df_join) <- c("Year","Age","Female_m", "Male_m", "Total_m","Female_p","Male_p","T

df_join <- df_join |>
  select(-c("Total_m", "Total_p")) # remove the total mortality and total population colum
```

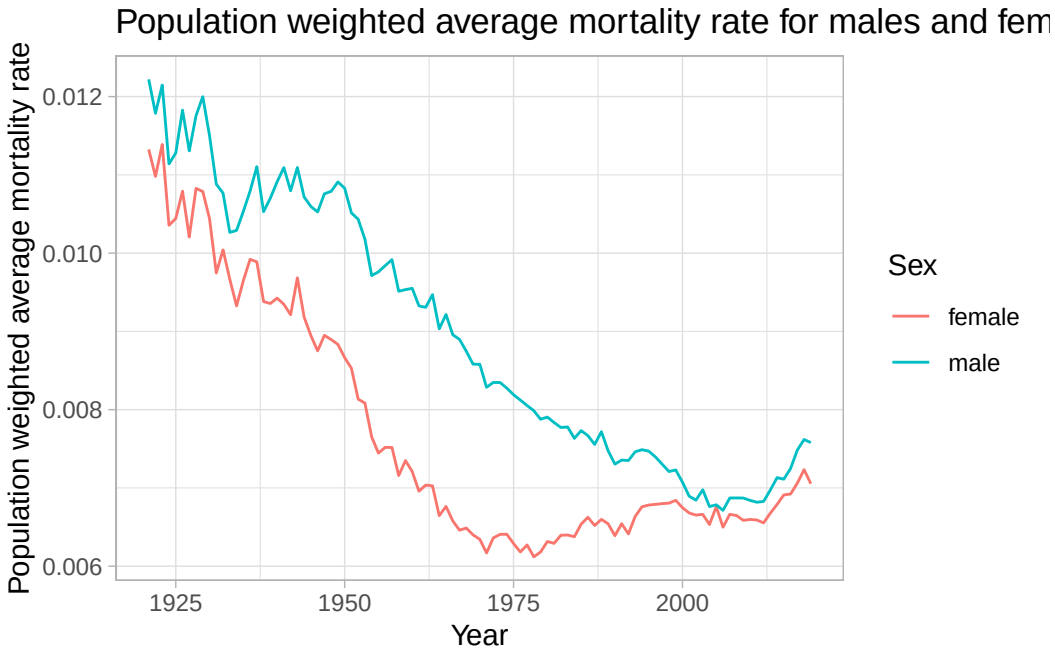
```
df_join <- df_join |>
  group_by(Year) |>
  mutate(total_f = sum(Female_p, na.rm = TRUE)) |>
  mutate(total_m = sum(Male_p, na.rm = TRUE))
```

```r
df_join <- df_join |>
  mutate(w_f= Female_m * Female_p/total_f) |>
  mutate(w_m = Male_m * Male_p/total_m)
```

```r
df_avg <- df_join |>
  group_by(Year) |>
  summarise(female = sum(w_f, na.rm = TRUE), male = sum(w_m, na.rm = TRUE))
```

```r
d_plot <- df_avg |>
  pivot_longer(female:male, names_to = "Sex", values_to = "weighted_avg")
```

```r
p1 <- d_plot |>
  ggplot(aes(x=Year, y=weighted_avg, color= Sex)) +
  geom_line()+
  ggtitle("Population weighted average mortality rate for males and females")+
  ylab("Population weighted average mortality rate") +
  theme_light() #change of theme
p1
```



Interpretation: The population weighted average mortality rate decreases and then increases

for both females and males.