Week 1 Lab

Lab Exercises

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

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.0
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()
```

```r
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.
```

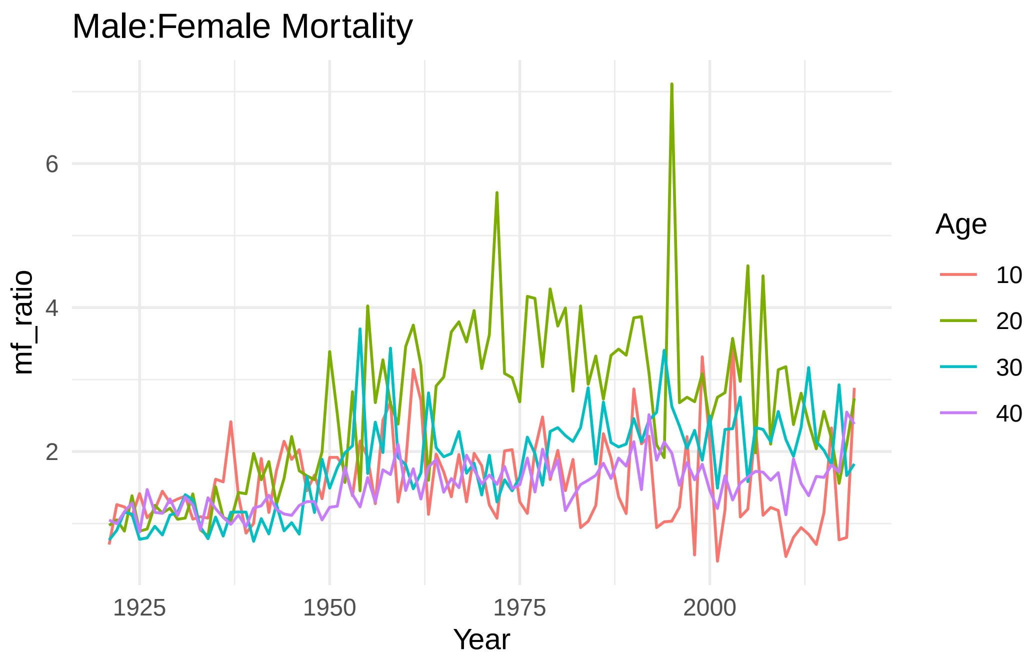
```r
head(dm)
```

```
# 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
```

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

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

dlab |>
  ggplot(aes(x = Year, y = mf_ratio, color = Age)) +
  geom_line() +
  theme_minimal()+
  labs(title="Male:Female Mortality")
```



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

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

```
# A tibble: 102 x 2
# Groups:   Year [99]
    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
# ... 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) |>
  summarize(across(c("Male","Female","Total"),sd,na.rm=TRUE))
```

```
# A tibble: 111 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    1.01     0.449    0.363
```

```
 9 105   1.29     1.27     1.27
10 106   1.13     1.21     1.20
# ... with 101 more rows
```

4. The Canadian HMD also provides population sizes over time (https://www.prdh.umontreal.ca/BDLC/data 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. Reformat existing dataset:

```
dm4 <- dm |>
select(Year:Male)
```

Get new dataset:

```
df <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt", skip = 2, c
```

Renaming Columns:

```
df=rename(df,pop_f=Female,pop_m=Male)
```

Combine Datasets:

```
df2 <- dm4 |>
  left_join(df)
```

```
Joining, by = c("Year", "Age")
```

```
df2
```

```
# A tibble: 10,989 x 7
    Year Age    Female    Male  pop_f  pop_m  Total
   <dbl> <chr>   <dbl>   <dbl>  <dbl>  <dbl>  <dbl>
 1  1921 0     0.0978  0.129   30157. 31530. 61687.
 2  1921 1     0.0129  0.0144  30391. 31319. 61711.
 3  1921 2     0.00521 0.00737 30962. 31785. 62747.
 4  1921 3     0.00471 0.00457 31306. 32031. 63336.
 5  1921 4     0.00461 0.00433 31364. 32046. 63409.
 6  1921 5     0.00372 0.00361 31175. 31847. 63021.
 7  1921 6     0.00265 0.00393 30808. 31466. 62274.
 8  1921 7     0.00295 0.00351 30295. 30922  61217.
```

```
 9   1921 8     0.00237 0.00285 29660. 30270. 59930.
10   1921 9     0.00198 0.00255 28923  29494. 58417.
# ... with 10,979 more rows
```

Add Total Deaths by Sex:

```
df3 <- df2 |>
  mutate(deaths_f=Female*pop_f, deaths_m=Male*pop_m)
```

Group by Year:

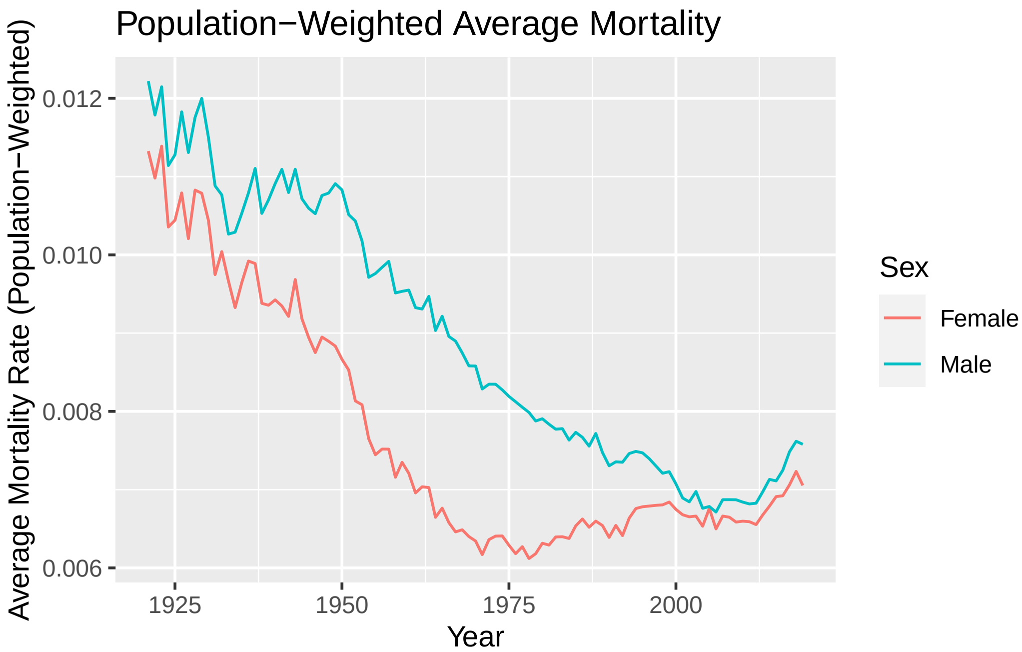
```
df4 <- df3 |>
  group_by(Year) |>
  summarise(across(pop_f:deaths_m,sum,na.rm=TRUE))
```

Calculate Population-weighted Mortality by Year:

```
df5 <- df4 |>
  mutate(Female=deaths_f/pop_f,Male=deaths_m/pop_m) |>
  pivot_longer(Female:Male,names_to="Sex",values_to="Weighted_Mortality")
```

Plot the Chart:

```
df5 |>
  ggplot(aes(x=Year,y=Weighted_Mortality,color=Sex))+
  geom_line()+
  labs(title="Population-Weighted Average Mortality",
       y="Average Mortality Rate (Population-Weighted)")
```



The population-weighted average mortality for both sexes trended downwards until the early 2000s when it started to trend upwards; the male average is significantly higher than the female average, though also since the early 2000s, the two are becoming much closer aligned.