# Lab 1

Qiaoyu (Terence) Liang

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

```
dm %>%
  mutate(mf_ratio = Male / Female) %>%
  filter(Age==10|Age==20|Age==30|Age==40) %>%
  ggplot(aes(x = Year, y = mf_ratio, color = factor(Age))) +
  geom_line() +
  theme_bw() +
  scale_color_brewer(palette = "Set1", name = "Age") +
  labs(title="Ratio of Male to Female Mortality Rates",
       subtitle="Over Time For Ages 10, 20, 30 and 40",
       y = "ratio")
```

Ratio of Male to Female Mortality Rates
Over Time For Ages 10, 20, 30 and 40
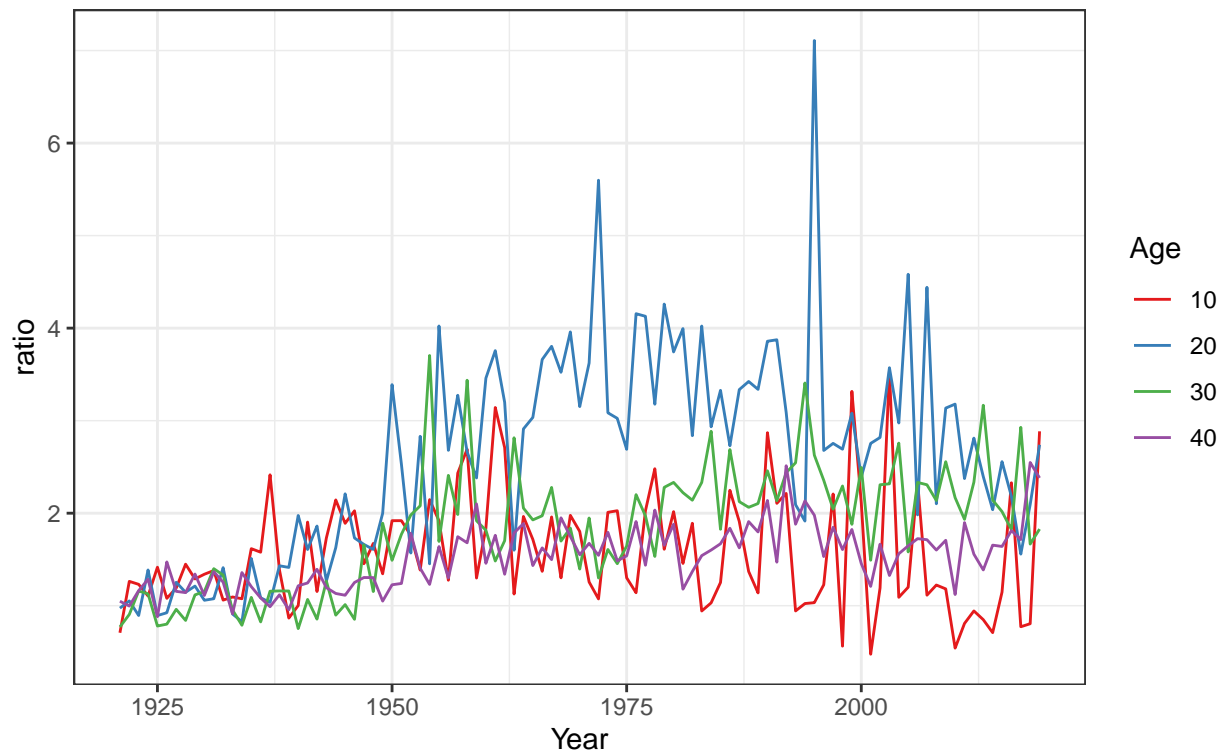
2. Find the age that has the highest female mortality rate each year

```
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.

```r
dms <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.txt",
                  skip = 2, col_types = "ddddd")
dms  %>%
  group_by(Age) %>%
  summarise(across(c(Male, Female, Total), sd, na.rm = T))
```

```
## # A tibble: 111 x 4
##      Age    Male   Female    Total
##    <dbl>   <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.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.

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

dm %>%
  rename(Male_mortality = Male,
         Female_mortality = Female) %>%
  select(-Total) %>%
  left_join(ds %>%
              rename(M_pop = Male,
                     F_pop = Female)) %>%
  drop_na() %>%
  group_by(Year) %>%
  summarise(Female_mortality = sum(Female_mortality*F_pop, na.rm=TRUE)/sum(F_pop, na.rm=TRUE),
            Male_mortality = sum(Male_mortality*M_pop, na.rm=TRUE)/sum(M_pop, na.rm=TRUE)) %>%
  pivot_longer(Female_mortality:Male_mortality, names_to="Sex", values_to="Mortality")%>%
  ggplot(aes(x = Year, y = Mortality, color = Sex)) +
    geom_line() +
    labs(y = "Population Weighted Average Mortality Rate",
         title = "Population Weighted Average Mortality Rate",
         subtitle = "For Males and Females Over Years")
```

Population Weighted Average Mortality Rate

For Males and Females Over Years
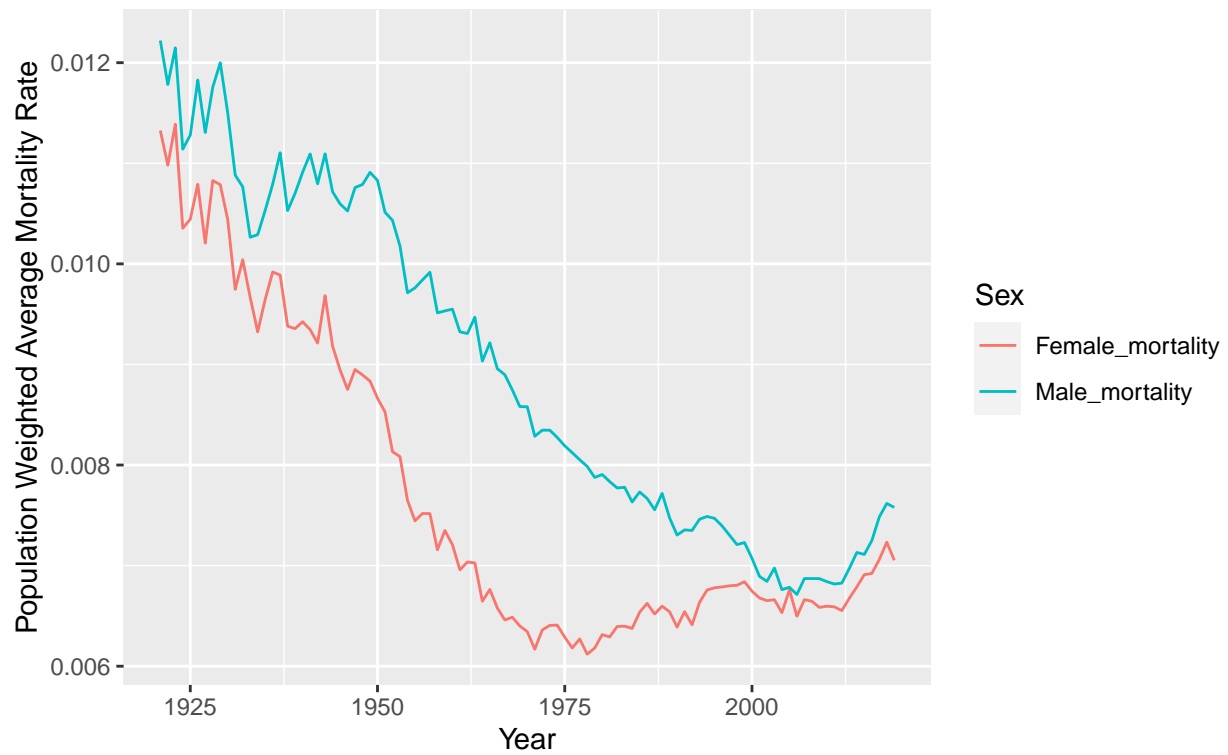
Comment: From the plot above, we find that the population weighted average mortality rate generally decreases over years and we can see that the mortality rates in the male population tend to be higher than the mortality rates in the female population over time.