Lab_1

Yaqi Shi, 1003813180

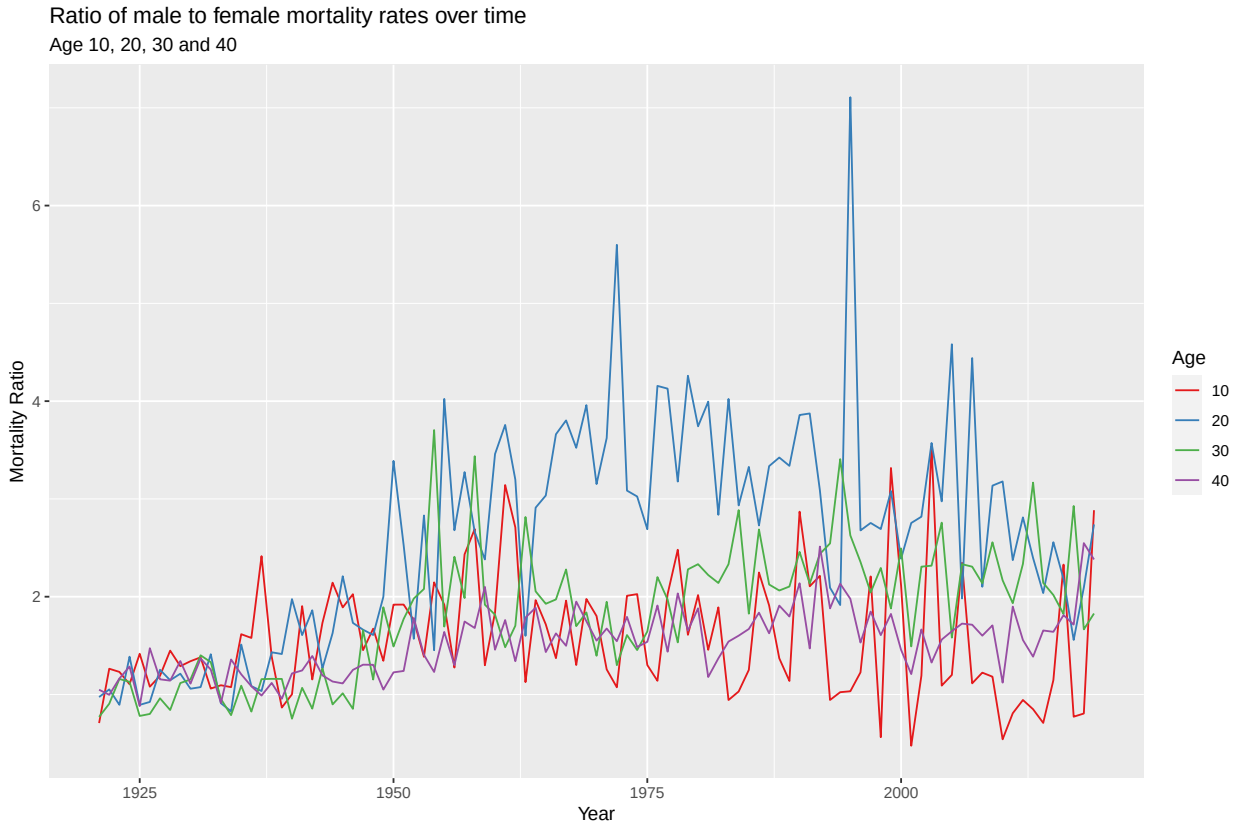
2023-01-12

Lab Exercises

Q1

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

dm1 <- dm %>%
  filter(Age==10|Age==20|Age==30|Age==40) %>%
  mutate(mf_ratio = Male/Female)

dm1 %>%
  ggplot(aes(x = Year, y = mf_ratio, color = Age)) +
  geom_line() +
  scale_color_brewer(palette = "Set1") +
  labs(title = "Ratio of male to female mortality rates over time",
       subtitle = "Age 10, 20, 30 and 40",
       y = "Mortality Ratio")
```



Q2

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

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

```
## # 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
```

Q3

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

```
dm %>%
  mutate(Age = as.numeric(Age)) %>%
  group_by(Age) %>%
  summarize(across(c("Male","Female", "Total"), sd, na.rm = TRUE))
```

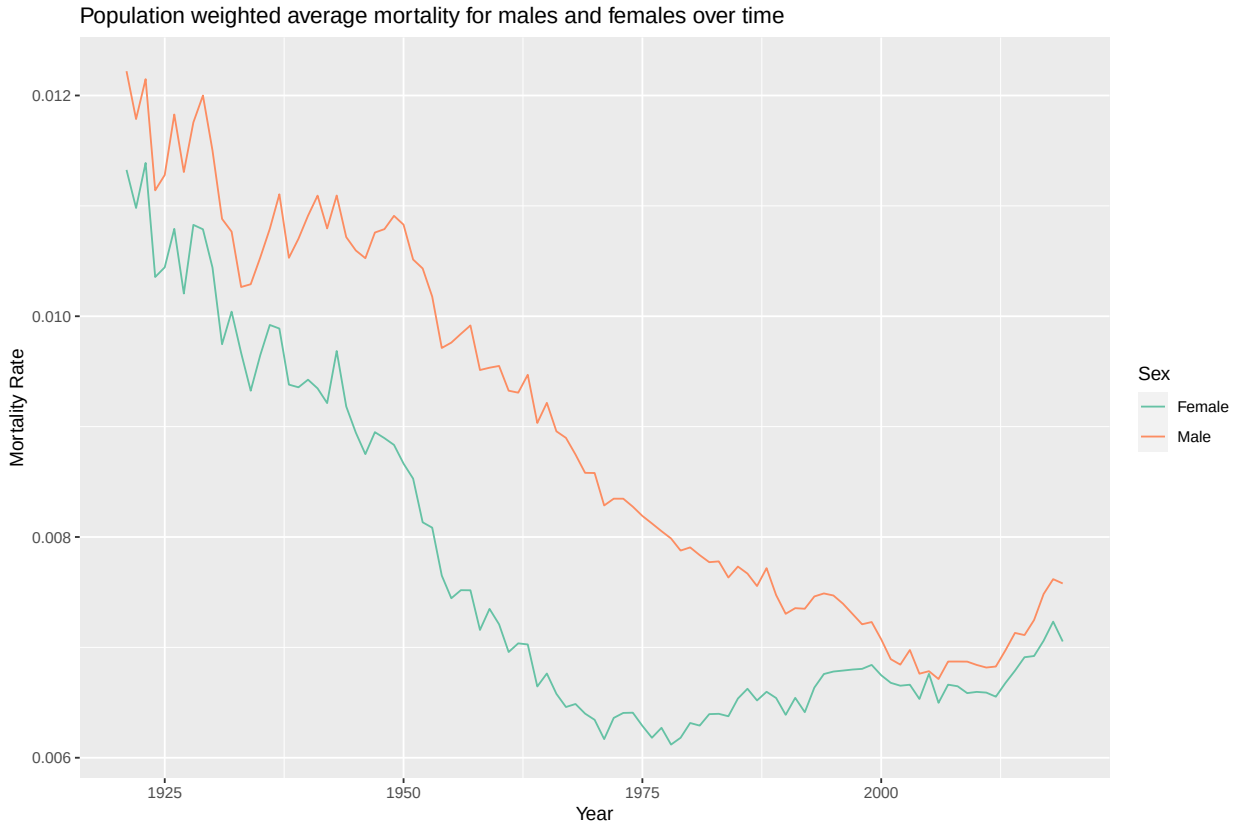
```
## # A tibble: 111 x 4
##      Age     Male   Female    Total
##    <dbl>    <dbl>    <dbl>    <dbl>
##  1     0 0.0330   0.0256   0.0294
##  2     1 0.00396  0.00352  0.00374
##  3     2 0.00175  0.00154  0.00164
##  4     3 0.00127  0.00113  0.00120
##  5     4 0.000987 0.000925 0.000947
##  6     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
```

Q4

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.

```
# Handle population data
hmd <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt", skip = 2, col_types = "c
hmd <- rename(hmd, Female_Pop  = Female, Male_Pop = Male)
hmd <- hmd %>% select(-Total)

# Handle mortality data
dm <- rename(dm, Female_Mort  = Female, Male_Mort = Male)
dm <- dm %>% select(-Total)

# Combine datasets and calculate weighted mortality rate
comb <- dm %>% left_join(hmd)
comb <- comb %>% mutate(Female_death = Female_Mort * Female_Pop, Male_death = Male_Mort * Male_Pop)

comb <- comb %>%
  group_by(Year) %>%
  summarize(sum_Female_Pop = sum(Female_Pop, na.rm = TRUE),
            sum_Female_Death = sum(Female_death, na.rm = TRUE),
            sum_Male_Pop = sum(Male_Pop, na.rm = TRUE),
            sum_Male_Death = sum(Male_death, na.rm = TRUE))

comb <- comb %>%
```

```
  group_by(Year) %>%
  mutate(Male = sum_Male_Death/sum_Male_Pop,
         Female = sum_Female_Death/sum_Female_Pop) %>%
  select(Year,Male,Female) %>%
  pivot_longer(Male:Female, names_to = "Sex", values_to = "Mortality")

# Produce plots
comb %>%
  ggplot(aes(x = Year, y = Mortality, color = Sex)) +
  geom_line() +
  scale_color_brewer(palette = "Set2") +
  labs(title = "Population weighted average mortality for males and females over time",
       y = "Mortality Rate")
```



Comment: The mortality rate for male were in general higher than the mortality rate for female across all those years and both mortality rates have dropped over time.