

# STA2201 Homework One(Lab One)

shaohan chang

2023-01-13

## Lab Exercises

Explanation: Download the packages of tidyverse.

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

Explanation: download the table from web-link.

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

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

```
head(dm)
```

```
## # A tibble: 6 × 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
```

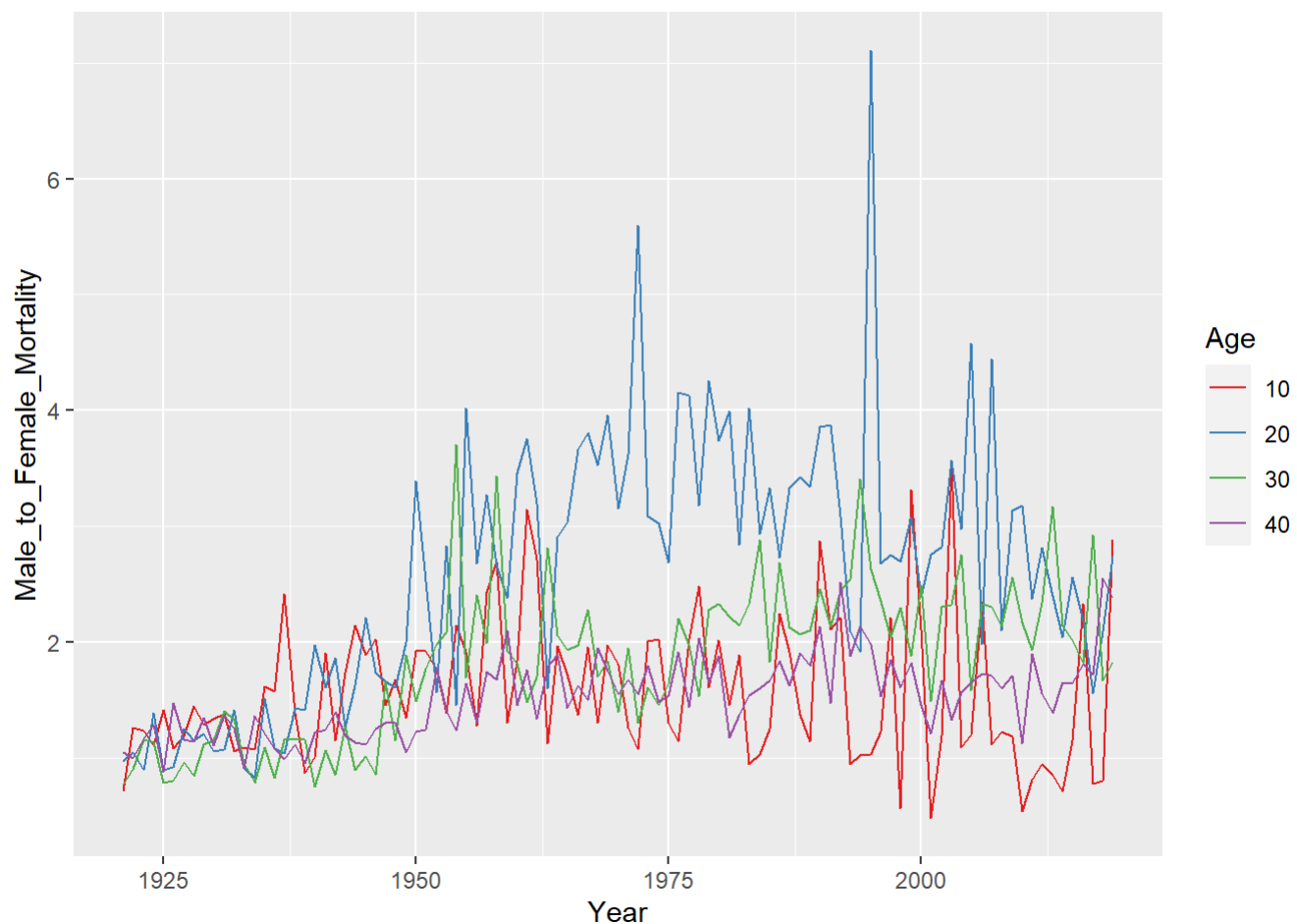
```
dm
```

```
## # A tibble: 10,989 × 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
## 7  1921 6     0.00265 0.00393 0.00330
## 8  1921 7     0.00295 0.00351 0.00323
## 9  1921 8     0.00237 0.00285 0.00262
## 10 1921 9     0.00198 0.00255 0.00227
## # ... with 10,979 more rows
```

#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

```
ratio <- dm |>
  filter(Age== 10 | Age==20 | Age == 30 | Age == 40) |>
  mutate(Male_to_Female_Mortality = Male/Female)

ratio |>
  ggplot(aes(x = Year , y = Male_to_Female_Mortality ,color = Age)) +
  geom_line() +
  scale_color_brewer(palette = 'Set1')
```



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

Explanation: Here is the list of the age that has the highest female mortality rate each year, however, the data result is not surprise, all of the result in highest female mortality rate each year, age are higher than 100, which is in my result of expectation. some of the mortality rate of the age is the same at the same year.

```
clean_data <- dm %>%
  select(-Male) %>%
  select(-Total)

group_by_year <- clean_data |>
  group_by(Year)|>
  filter(Female == max(Female, na.rm=TRUE)) |>
  select(Year, Age, Female)
group_by_year
```

```
## # A tibble: 102 × 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
```

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

Explanation:Used the method of `summarize(across())` to calculate the standard deviation of mortality rates by age for the Male, Female and Total populations.

```
rate_of_Male <- dm |>
  group_by(Age)|>
  summarize(across(c(Male,Female,Total),sd,na.rm = TRUE))
rate_of_Male
```

```
## # A tibble: 111 × 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.prhd.umontreal.ca/BDLC/data/ont/Population.txt>

(<https://www.prhd.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.

Explanation:read the table from web-link

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

```
## # A tibble: 6 × 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.
```

Explanation: AS showed in the figure, In overall, the mortality rate of population weighted average for females and males from 1921 to 2020 showed an increasing trending.

```
weighted_mean_table_female <- pp |>
  group_by (Year) |>
  summarize(weighted_average_female = weighted.mean(Female,Total,na.rm = TRUE))

weighted_mean_table_male <- pp |>
  group_by (Year) |>
  summarize(weighted_average_male = weighted.mean(Male,Total,na.rm = TRUE))

fix_up = left_join(weighted_mean_table_female,weighted_mean_table_male, by='Year')

fix_up <- fix_up |>
  select(Year,weighted_average_female,weighted_average_male) |>
  pivot_longer(weighted_average_female:weighted_average_male, names_to = "Sex", values_to = "Mortality")

fix_up |>
  ggplot(aes(x = Year, y = Mortality, color = Sex, linetype = Sex)) +
  geom_line() +
  scale_color_brewer(palette = "Set1") +
  ggtitle("Mortality Rate of Population Weighted Average for Females and Males from 1921 to 2020") +
  theme(plot.title = element_text(size = 8, color = "black", face = "bold"))
```

