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.prdh.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.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.
txt'
## 109 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.
txt'
## 110 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.
txt'
## 110 Male   no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.
txt'
## 110 Total  no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.
txt'
## ... ...... ..................... ......
.....................................................
## See problems(...) for more details.
```

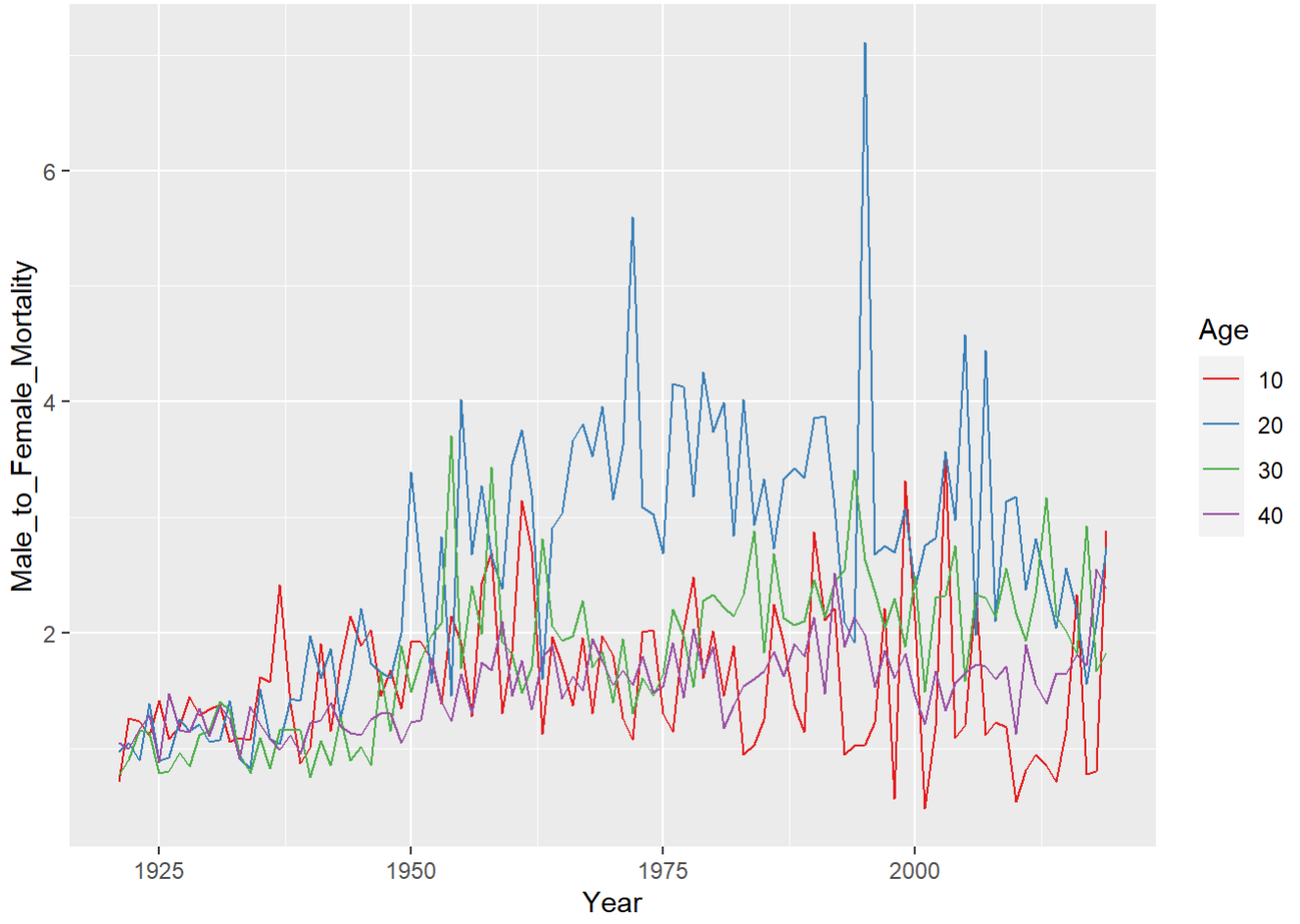
```
head(dm)
```

```
## # 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
```

#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 felmale mortality rate each year, however, the data result is not suprise, all of the result in highest felmale mortality rate each year, age are higher than 100, which is in my result of expectation.

```
group_by_year <- dm |>
  group_by(Year)|>
  summarise(highest_female_mortality_rate=Age[which.max(Female)])


group_by_year
```

```
## # A tibble: 99 × 2
##     Year highest_female_mortality_rate
##    <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 89 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(Total:Male|Female,sd,na.rm = TRUE))
rate_of_Male
```

```
## # A tibble: 111 × 4
##    Age      Total     Male   Female
##    <chr>    <dbl>    <dbl>    <dbl>
##  1 0     0.0294   0.0330   0.0256
##  2 1     0.00374  0.00396  0.00352
##  3 10    0.000509 0.000561 0.000474
##  4 100   0.0729   0.138    0.0928
##  5 101   0.0995   0.158    0.125
##  6 102   0.114    0.214    0.143
##  7 103   0.208    0.371    0.252
##  8 104   0.363    1.01     0.449
##  9 105   1.27     1.29     1.27
## 10 106   1.20     1.13     1.21
## # … with 101 more rows
```

#4. The Canadian HMD also provides population sizes over time (https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt (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.

Explanation:read the table from web-link

```
pp <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt", skip = 2, col_typ
es = "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 = "Mor
tality")

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 202
0") +
  theme(plot.title = element_text(size = 8, color = "black", face = "bold"))
```