LAB Exercise-Jan8th

Luis Correia - 1006508566

9 January 2020

Contents

1 Setup Tidyverse **1**

2 Lab Exercises **1**

1 Setup Tidyverse

Read in some packages that we'll be using:

```
library(tidyverse)
```

Read in mortality rates for Canada, as seen in class.

```
path <- "C:/Users/LuisAlvaro/Documents/GitHub/applied-stats/data/CAN_Mx_1x1.txt"
dm <- read_table(path, skip = 2)
head(dm)
```

```
## # A tibble: 6 x 5
##    Year Age   Female   Male     Total
##   <dbl> <chr> <chr>    <chr>    <chr>
## 1  1921 0     0.105821 0.138250 0.122259
## 2  1921 1     0.015593 0.017806 0.016710
## 3  1921 2     0.007409 0.008521 0.007970
## 4  1921 3     0.005442 0.006111 0.005779
## 5  1921 4     0.004563 0.004745 0.004655
## 6  1921 5     0.003433 0.003828 0.003633
```

Repeat clean up as seen in class to make the ages and mortality rates numbers not characters.

```
dm <-
  dm %>%
  mutate(Age = as.numeric(Age),
         Female = as.numeric(Female),
         Male = as.numeric(Male),
         Total = as.numeric(Male))
```

2 Lab Exercises

1. Plot the ratio of male to female mortality rates over time and change the theme (e.g. `theme_bw()`)

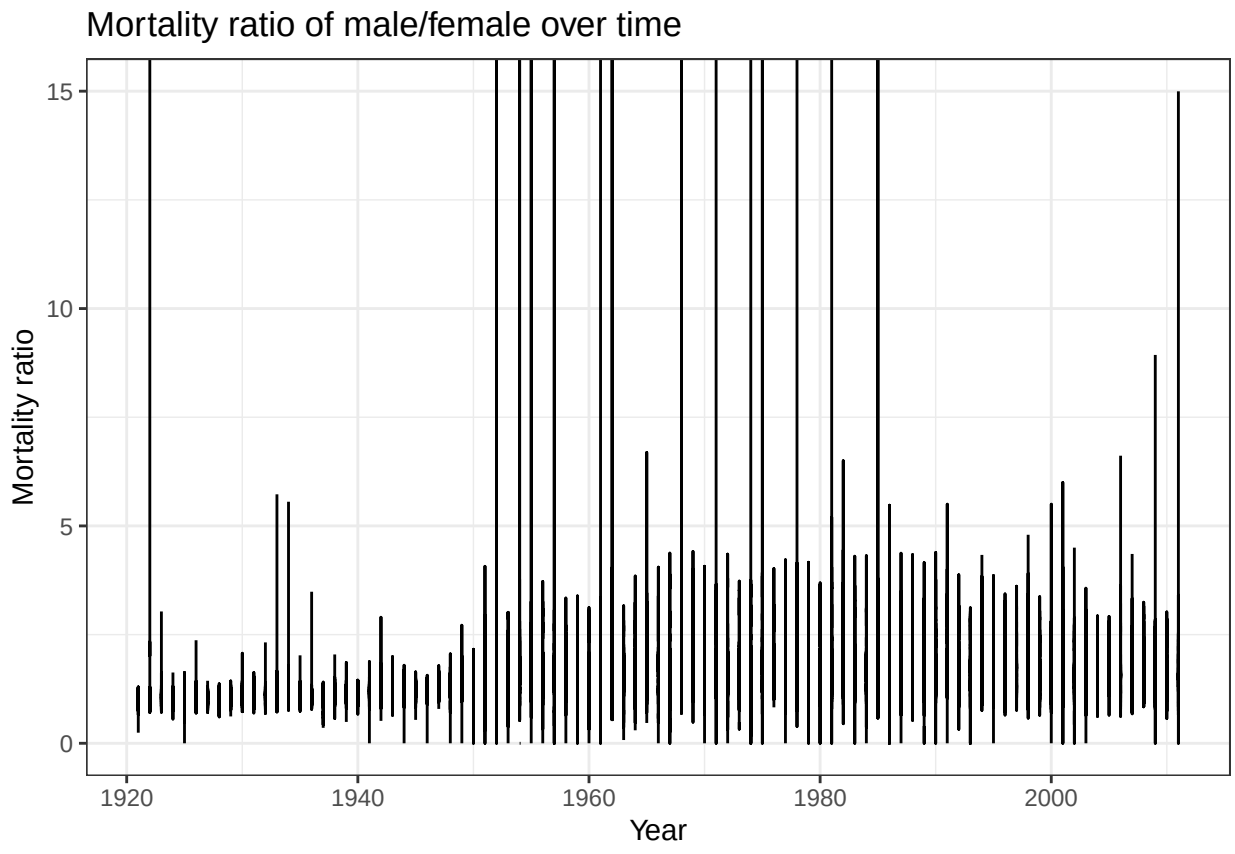
First calculate the male/female ratio

```
dm_ratio <- dm %>%
  mutate(m_f_ratio = Male/Female)
head(dm_ratio)
```

```
## # A tibble: 6 x 6
##    Year   Age  Female    Male   Total m_f_ratio
##   <dbl> <dbl>   <dbl>   <dbl>   <dbl>     <dbl>
## 1  1921     0 0.106   0.138   0.138        1.31
## 2  1921     1 0.0156  0.0178  0.0178       1.14
## 3  1921     2 0.00741 0.00852 0.00852      1.15
## 4  1921     3 0.00544 0.00611 0.00611      1.12
## 5  1921     4 0.00456 0.00474 0.00474      1.04
## 6  1921     5 0.00343 0.00383 0.00383      1.12
```

... an then Plot graphics

```
ggplot(dm_ratio, aes(Year, m_f_ratio, group = Year)) +
  geom_line()+
  ylab("Mortality ratio")+
  theme_bw()+
  ggtitle("Mortality ratio of male/female over time")
```



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

```
MaxAgeF <-
  dm %>%
  group_by(Year) %>%
  mutate(max_FemMort = (Female == max(Female, na.rm = TRUE))) %>%
```

```
  filter(max_FemMort) %>%
  summarise(Age)
head(MaxAgeF)
```

```
## # A tibble: 6 x 2
##    Year   Age
##   <dbl> <dbl>
## 1  1921   107
## 2  1922   106
## 3  1923   107
## 4  1924   107
## 5  1925   106
## 6  1926   108
```

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

```
SDev <-
  dm %>%
  group_by(Age) %>%
  summarise_at(c("Female","Male","Total"),sd,na.rm = TRUE)
head(SDev)
```

```
## # A tibble: 6 x 4
##     Age   Female    Male   Total
##   <dbl>    <dbl>   <dbl>   <dbl>
## 1     0 0.0307   0.0402  0.0402
## 2     1 0.00490  0.00562 0.00562
## 3     2 0.00220  0.00247 0.00247
## 4     3 0.00153  0.00168 0.00168
## 5     4 0.00121  0.00128 0.00128
## 6     5 0.000935 0.00104 0.00104
```