STA2201 Lab 1

Timothy Jordan Regis

2023-01-14

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v stringr 1.4.0
## v tidyr   1.2.1     v forcats 0.5.2
## v readr   2.1.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
```

```
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 x 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
```

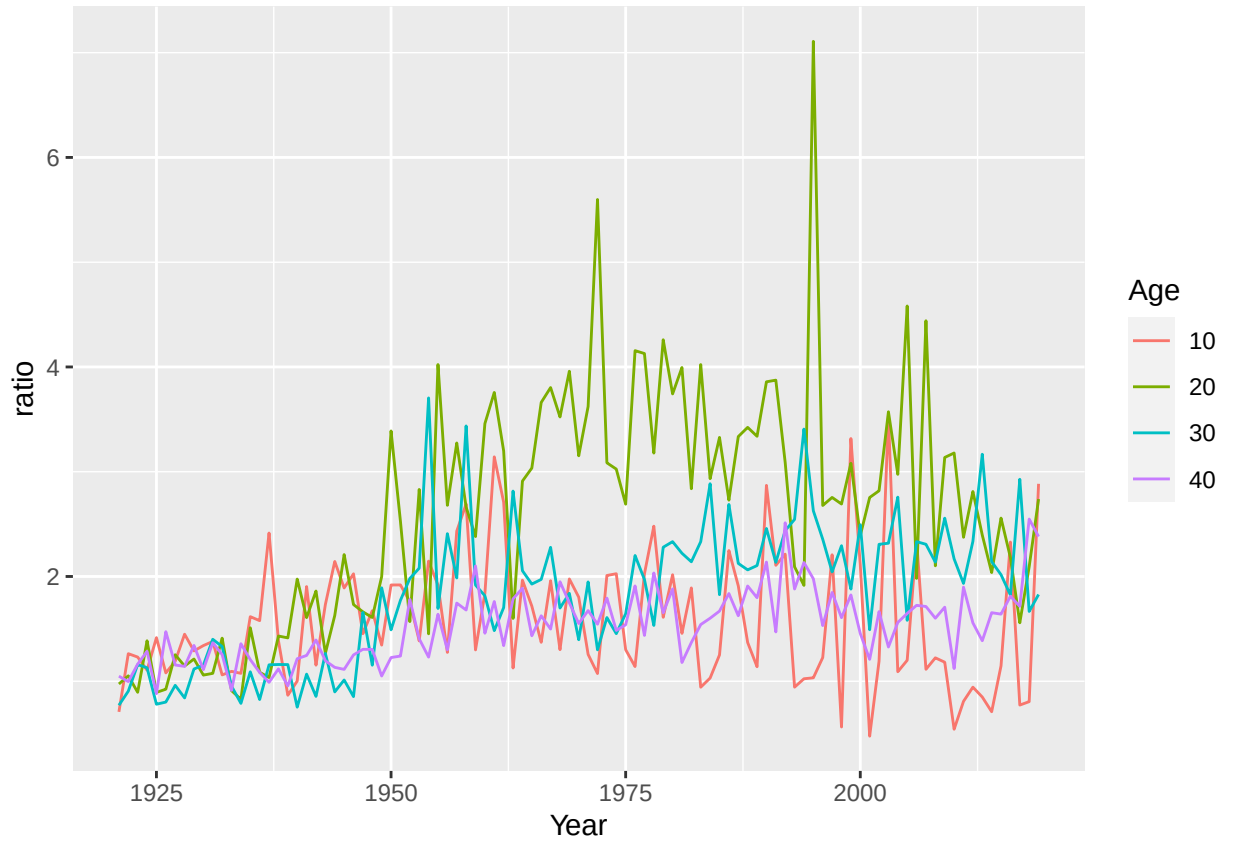
```
dm <- data.frame(dm)

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

```
## # A tibble: 6 x 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.
```

```
pop <- data.frame(pop)
```

Q1

```
mort_rates = dm %>%
  filter(
  Age == 10 | Age == 20 | Age == 30 | Age == 40
) %>%
  mutate(
  ratio = Male/Female
)
mort_rates %>%
  ggplot(aes(y = ratio, x = Year, color = Age)) +
  geom_line()
```




```
#plot(mort_rates$ratio, mort_rates$Year, type = "line", color = Age)
```

Q2

```
max(dm$Female)
```

```
## [1] NA
```

```
# Highest Absolute
dm %>% group_by(Year) %>% filter(Female == max(Female)) %>% select(Year, Age, Female)
```

```
## # A tibble: 49 x 3
## # Groups:   Year [48]
##     Year Age   Female
##    <dbl> <chr>  <dbl>
##  1  1948 99     0.677
##  2  1949 102    0.573
##  3  1950 102    0.629
##  4  1951 110+   2.65
##  5  1953 106    1.16
##  6  1954 110+   4.08
##  7  1955 107    1.02
##  8  1956 110+   4.08
##  9  1957 107    2.08
## 10  1958 110+   4.08
## # ... with 39 more rows
```

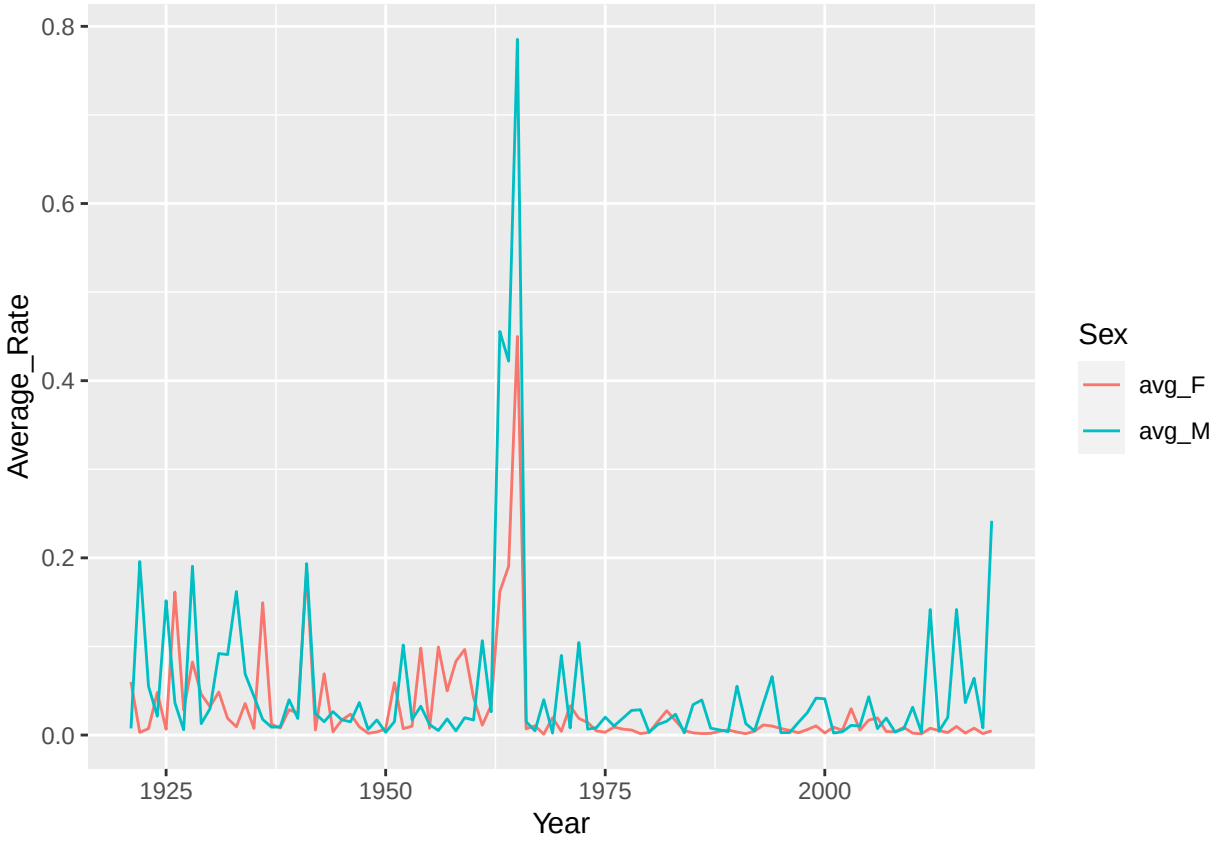
Q3

```
dm %>%
  group_by(Age) %>%
  summarise(std_M = sd(Male, na.rm = TRUE),
            std_F = sd(Female, na.rm = TRUE),
            std_T = sd(Total, na.rm = TRUE))
```

```
## # A tibble: 111 x 4
##    Age      std_M    std_F    std_T
##    <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
```

Q4

```
dm %>% left_join(y = pop, by = c("Year", "Age")) %>% mutate(
  pw_F = Female.x/Female.y,
  pw_M = Male.x/Male.y
) %>% filter(!is.nan(pw_M) & !is.nan(pw_F) &
               !is.infinite(pw_M) & !is.infinite(pw_F)) %>%
  group_by(Year) %>%
  summarise(avg_F = mean(pw_F, na.rm = TRUE), avg_M = mean(pw_M, na.rm = TRUE)) %>%
  pivot_longer(cols = c(2:3), names_to = "Sex", values_to = "Average_Rate") %>%
  ggplot(aes(x = Year, y = Average_Rate, color = Sex)) + geom_line()
```



From the looks of this graph, men tend to see a much higher variability in their death rates, often attaining greater levels than Women.