STA2201 Lab 1

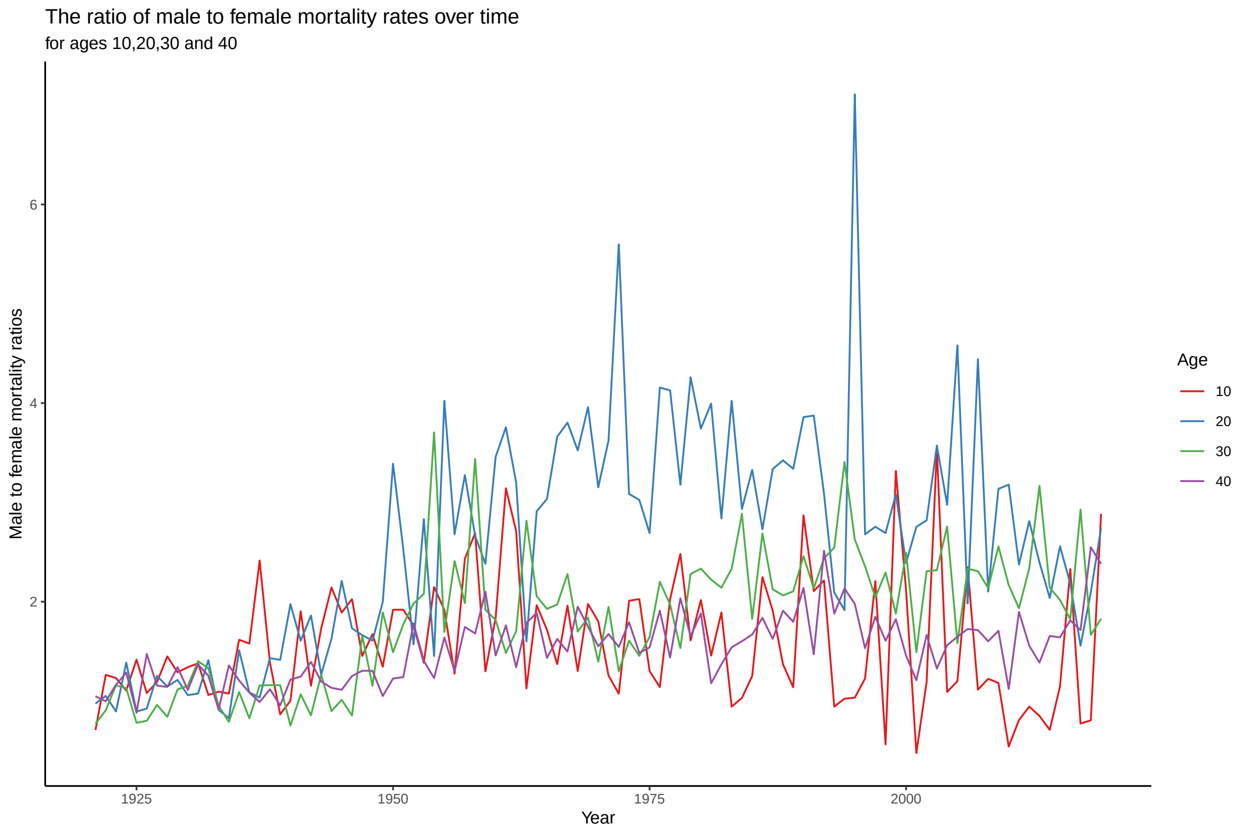
Quynh Vu

2023-01-17

```
url1 <- "https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1.txt"
dd <- read.table(url1, header = TRUE, skip = 2)
dd$Female <- as.double(dd$Female)
dd$Male <- as.double(dd$Male)
dd$Year <- as.double(dd$Year)
dd$Total <- as.double(dd$Total)
```

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

```
dd_Age <- dd |>
             filter(Age == 10|Age == 20|Age == 30|Age == 40)|>
             mutate(mf_ratio = Male/Female)
dd_Age |>
       ggplot(aes(x = Year, y = mf_ratio, color = Age)) +
       geom_line() +
       scale_color_brewer(palette = "Set1") +
       labs(title = "The ratio of male to female mortality rates over time",
       subtitle = "for ages 10,20,30 and 40",
       y = "Male to female mortality ratios") +
       theme_classic(base_size = 10) # change theme
```



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

```
dd_HighestF <- dd |>
                  select(Year:Female) |>
                  merge(aggregate(Female~Year, dd, max, na.rm=TRUE))
dd_HighestF
```

```
##      Year   Female  Age
## 1    1921 6.000000  106
## 2    1922 0.603404   98
## 3    1923 0.524211  104
## 4    1924 6.000000  107
## 5    1925 0.513654   98
## 6    1926 4.164706  106
## 7    1927 6.000000  106
## 8    1928 2.131980  104
## 9    1929 1.323263  104
## 10   1930 6.000000  105
## 11   1931 1.250000  104
## 12   1932 6.000000  105
## 13   1933 0.583120  104
## 14   1934 6.000000  106
## 15   1935 0.508961  104
## 16   1936 4.247788  106
## 17   1937 6.000000  105
## 18   1938 0.891892  104
```

```
## 19  1939 6.000000  105
## 20  1940 1.066059  104
## 21  1941 4.275862  105
## 22  1942 0.787234  104
## 23  1943 1.592920  105
## 24  1944 0.455084   98
## 25  1945 0.644571  104
## 26  1946 1.010050  105
## 27  1947 0.557979  104
## 28  1948 0.676512   99
## 29  1949 0.572759  102
## 30  1950 0.628931  102
## 31  1951 2.649007 110+
## 32  1952 0.675676  107
## 33  1953 1.162791  106
## 34  1954 4.081633 110+
## 35  1955 1.020408  107
## 36  1956 4.081633 110+
## 37  1957 2.083333  107
## 38  1958 4.081633 110+
## 39  1959 4.081633  108
## 40  1960 1.801802  107
## 41  1961 0.877193  106
## 42  1962 1.960784  108
## 43  1963 2.157303  109
## 44  1964 2.250000  109
## 45  1965 4.105263  109
## 46  1966 1.242236  105
## 47  1967 1.343785  107
## 48  1968 0.512564   97
## 49  1969 4.095563  109
## 50  1970 2.790698  107
## 51  1971 1.421801  107
## 52  1972 1.345291  107
## 53  1973 0.771208  105
## 54  1974 1.500000  109
## 55  1974 1.500000  107
## 56  1975 6.000000  108
## 57  1975 6.000000 110+
## 58  1976 0.857143  106
## 59  1976 0.857143  108
## 60  1977 1.100000  103
## 61  1978 3.000000  109
## 62  1979 0.857143  109
## 63  1980 1.500000 110+
## 64  1981 1.384615  107
## 65  1982 1.500000  109
## 66  1983 1.500000 110+
## 67  1984 6.000000 110+
## 68  1985 6.000000 110+
## 69  1986 1.500000  109
## 70  1987 1.000000  109
## 71  1988 0.857143  108
## 72  1989 0.720000  108
```

```
## 73   1990 0.642857  108
## 74   1991 0.527415  103
## 75   1992 1.235294  108
## 76   1993 1.636364  109
## 77   1994 1.333333  109
## 78   1995 0.978593  107
## 79   1996 1.600000  109
## 80   1997 0.885609  107
## 81   1998 1.121495 110+
## 82   1999 1.675978 110+
## 83   2000 1.120000  106
## 84   2001 1.090909 110+
## 85   2002 1.463415  107
## 86   2003 3.000000  109
## 87   2004 1.000000  108
## 88   2005 2.117647  108
## 89   2006 3.000000 110+
## 90   2007 0.737619  107
## 91   2008 1.041215  109
## 92   2009 1.521555 110+
## 93   2010 0.986610  108
## 94   2011 2.227679 110+
## 95   2012 0.981461  109
## 96   2013 0.980926 110+
## 97   2014 1.205424 110+
## 98   2015 1.829268 110+
## 99   2016 1.049869 110+
## 100  2017 1.214575 110+
## 101  2018 0.859195 110+
## 102  2019 1.477105 110+
```

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

```
dd_SD <- dd |>
            group_by(Age) |>
            summarize(across(Female:Total, sd, na.rm = TRUE))
dd_SD
```

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

```
## 10 106   1.21     1.13     1.20
## # ... with 101 more rows
```

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.

```
url2 <- "https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt"
dp <- read.table(url2, header = TRUE, skip = 2)
```

```
dp_Male <- dp |>
              select(Year, Age, Male) |> # subset Male and group by year
              group_by(Year)|>
              mutate(TotalMale = sum(Male)) |>
              mutate(MWeight = Male/TotalMale) |>
              filter(Year < 2020) # we don't have data on 2020 mortality rates
dp_Male$M_mortality <- dd$Male
Male_weightedmean <- dp_Male |>
                             group_by(Year) |>
                             summarise(weighted.mean(M_mortality, MWeight))
```

```
dp_Female <- dp |>
                select(Year, Age, Female) |> # subset Female and group by year
                group_by(Year)|>
                mutate(TotalFemale = sum(Female)) |>
                mutate(FWeight = Female/TotalFemale) |>
                filter(Year < 2020) # we don't have data on 2020 mortality rates
dp_Female$F_mortality <- dd$Female
Female_weightedmean <- dp_Female |>
                             group_by(Year) |>
                             summarise(weighted.mean(F_mortality, FWeight))
```

The population weighted average mortality rate for males and females:

```
weightedMean <- merge(Female_weightedmean, Male_weightedmean)
colnames(weightedMean) <- c("Year", "Female", "Male")
weightedMean
```

```
##    Year      Female        Male
## 1  1921 0.011325796 0.012220142
## 2  1922 0.010980856 0.011786651
## 3  1923 0.011388715 0.012148074
## 4  1924 0.010355531 0.011140613
## 5  1925 0.010443200 0.011279815
## 6  1926 0.010791811 0.011827697
## 7  1927 0.010206070 0.011306594
## 8  1928 0.010827705 0.011756738
```

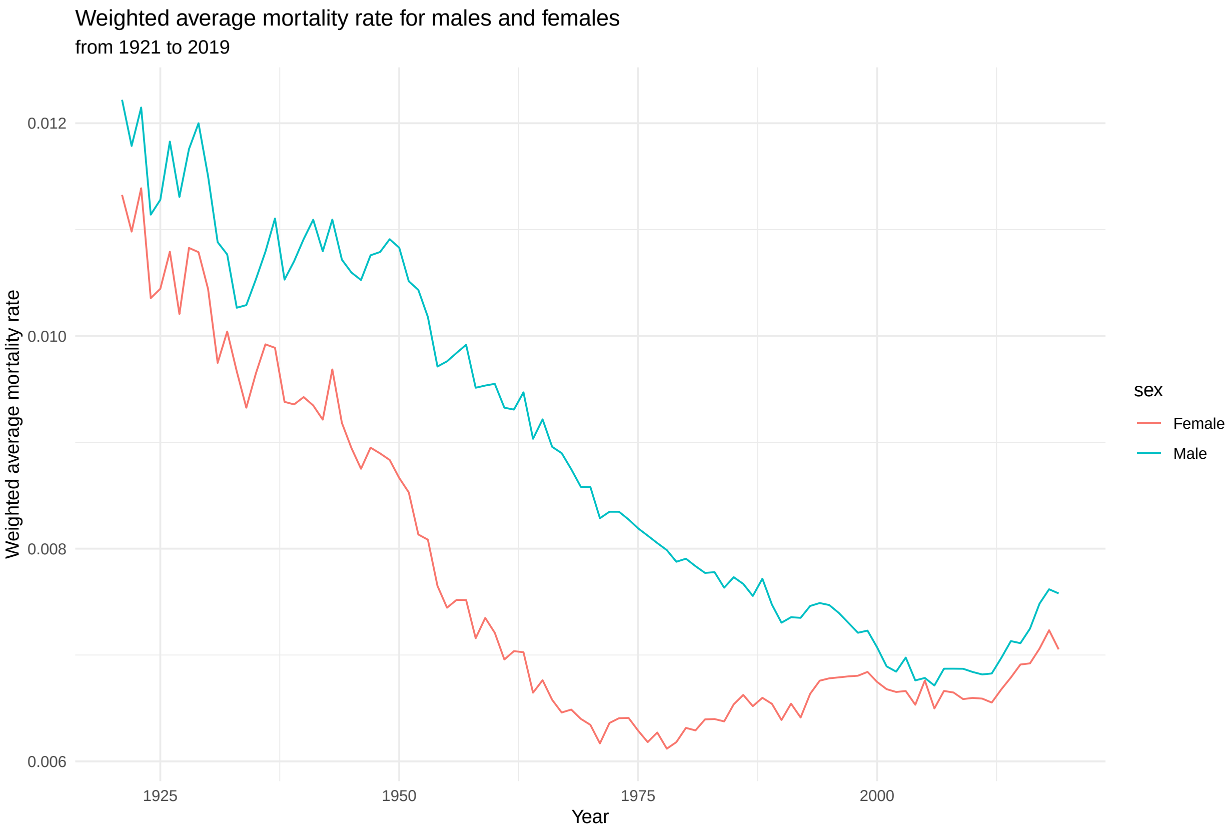
```
## 9  1929 0.010787682 0.011999642
## 10 1930 0.010440854 0.011501583
## 11 1931 0.009746587 0.010881929
## 12 1932 0.010041492 0.010766138
## 13 1933 0.009666937 0.010264978
## 14 1934 0.009325380 0.010289879
## 15 1935 0.009647106 0.010532033
## 16 1936 0.009920940 0.010791736
## 17 1937 0.009888581 0.011104573
## 18 1938 0.009380425 0.010529354
## 19 1939 0.009356136 0.010702494
## 20 1940 0.009425049 0.010910648
## 21 1941 0.009346245 0.011092690
## 22 1942 0.009212997 0.010795760
## 23 1943 0.009684846 0.011093705
## 24 1944 0.009182679 0.010717220
## 25 1945 0.008946195 0.010596214
## 26 1946 0.008750802 0.010525813
## 27 1947 0.008949747 0.010758398
## 28 1948 0.008895455 0.010789256
## 29 1949 0.008834123 0.010909576
## 30 1950 0.008664632 0.010829554
## 31 1951 0.008528869 0.010513888
## 32 1952 0.008133128 0.010432913
## 33 1953 0.008084390 0.010177697
## 34 1954 0.007650458 0.009712427
## 35 1955 0.007444980 0.009761095
## 36 1956 0.007518059 0.009840321
## 37 1957 0.007517151 0.009916586
## 38 1958 0.007157927 0.009512735
## 39 1959 0.007348880 0.009533546
## 40 1960 0.007208729 0.009549593
## 41 1961 0.006957777 0.009325406
## 42 1962 0.007036301 0.009307514
## 43 1963 0.007026484 0.009469654
## 44 1964 0.006645727 0.009032921
## 45 1965 0.006763255 0.009215656
## 46 1966 0.006579092 0.008957463
## 47 1967 0.006459522 0.008897932
## 48 1968 0.006487001 0.008747111
## 49 1969 0.006399882 0.008581111
## 50 1970 0.006343572 0.008579693
## 51 1971 0.006168827 0.008285794
## 52 1972 0.006360885 0.008347010
## 53 1973 0.006405991 0.008346822
## 54 1974 0.006408222 0.008274412
## 55 1975 0.006288947 0.008190711
## 56 1976 0.006181239 0.008122683
## 57 1977 0.006271285 0.008052660
## 58 1978 0.006119284 0.007986374
## 59 1979 0.006181102 0.007876822
## 60 1980 0.006314860 0.007905656
## 61 1981 0.006291226 0.007835429
## 62 1982 0.006395103 0.007771803
```

```
## 63 1983 0.006397921 0.007779006
## 64 1984 0.006375816 0.007632986
## 65 1985 0.006536511 0.007731978
## 66 1986 0.006624870 0.007668443
## 67 1987 0.006519139 0.007555427
## 68 1988 0.006597563 0.007717754
## 69 1989 0.006540967 0.007473754
## 70 1990 0.006389031 0.007303942
## 71 1991 0.006542816 0.007355351
## 72 1992 0.006412417 0.007349749
## 73 1993 0.006635369 0.007461245
## 74 1994 0.006758491 0.007488547
## 75 1995 0.006780910 0.007469755
## 76 1996 0.006789700 0.007395549
## 77 1997 0.006799199 0.007302579
## 78 1998 0.006805188 0.007209105
## 79 1999 0.006841197 0.007229552
## 80 2000 0.006747247 0.007072570
## 81 2001 0.006678811 0.006892941
## 82 2002 0.006652632 0.006843698
## 83 2003 0.006661587 0.006976017
## 84 2004 0.006532326 0.006761312
## 85 2005 0.006760140 0.006784345
## 86 2006 0.006497701 0.006713236
## 87 2007 0.006662134 0.006871433
## 88 2008 0.006647475 0.006871517
## 89 2009 0.006585504 0.006870746
## 90 2010 0.006596625 0.006840626
## 91 2011 0.006590114 0.006817226
## 92 2012 0.006553040 0.006826342
## 93 2013 0.006676823 0.006972525
## 94 2014 0.006788516 0.007130816
## 95 2015 0.006910715 0.007111311
## 96 2016 0.006921469 0.007247882
## 97 2017 0.007060244 0.007483179
## 98 2018 0.007233095 0.007618084
## 99 2019 0.007053412 0.007579056
```

The weighted average mortality rate across 111 age groups of males was consistently higher than that of females from 1921 to 2019, though both of which appear to decrease over time.

```r
weightedMean_long <- weightedMean |>
                                          pivot_longer(Female:Male, names_to = "sex",
                                                       values_to = "WeightedMean")

weightedMean_long |>
                  ggplot(aes(x = Year, y = WeightedMean, color = sex)) +
                  geom_line()+
                  labs(title = "Weighted average mortality rate for males and females",
                       subtitle = "from 1921 to 2019",
                       y = "Weighted average mortality rate") +
                  theme_minimal()
```


Weighted average mortality rate for males and females
from 1921 to 2019
Weighted average mortality rate
0.012
0.010
0.008
0.006
1925
1950
1975
2000
Year
sex
Female
Male