Applied Stats 2 Lab 1

Lars Kutschinski

```
library(tidyverse)
```

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

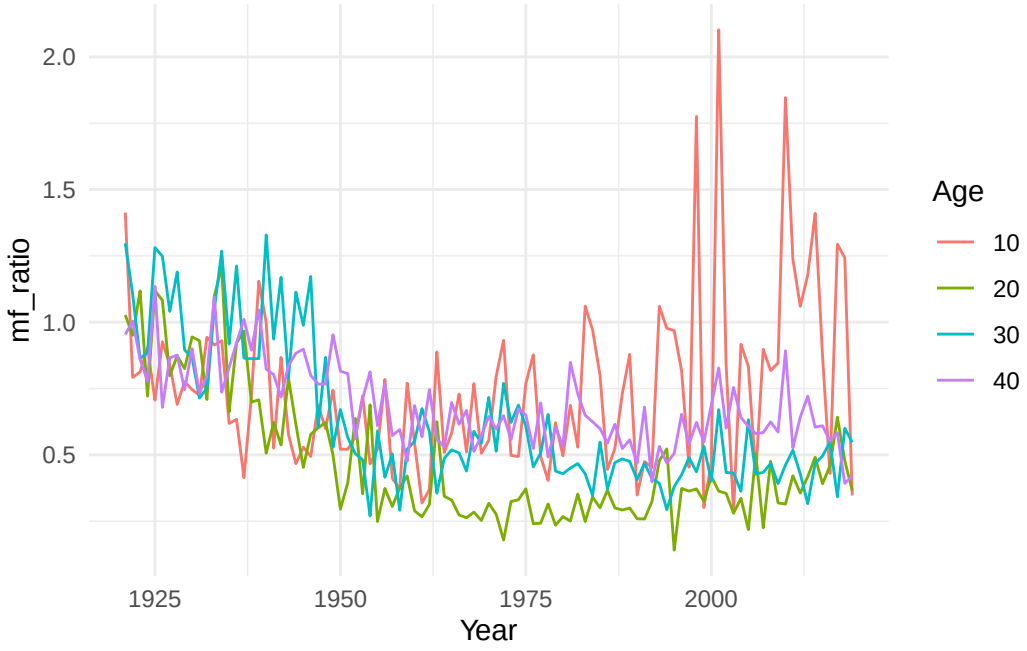
```
Warning: 494 parsing failures.
row    col               expected actual
108 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
109 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Female no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Male   no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
110 Total  no trailing characters      . 'https://www.prdh.umontreal.ca/BDLC/data/ont/Mx_1x1
... ...... ..................... ...... ..............................................
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
```

1. Plot the female to male ratios

```
dm <- dm |> mutate(mf_ratio = Female/Male)
plot_data <- dm |>
  filter(Age == 10|Age == 20|Age == 30|Age == 40) |>
  select(Year, Age, mf_ratio)
ggplot(data = plot_data, aes(x = Year, y = mf_ratio, color = Age)) +
  geom_line() +
  theme_minimal()
```



2. Find the minimum age for each year

```
dm |> group_by(Year) |> summarize(Age[which.min(Female)])
```

```
# A tibble: 99 x 2
    Year `Age[which.min(Female)]`
   <dbl> <chr>
 1  1921 13
 2  1922 104
 3  1923 105
```

```
 4  1924 14
 5  1925 105
 6  1926 11
 7  1927 9
 8  1928 9
 9  1929 10
10  1930 13
# i 89 more rows
```

3. Find the standard deviation

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

```
Warning: There was 1 warning in `summarize()`.
i In argument: `across(c(Male, Female, Total), sd, na.rm = TRUE)`.
i In group 1: `Age = "0"`.
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `.fns` through an anonymous function instead.

  # Previously
  across(a:b, mean, na.rm = TRUE)

  # Now
  across(a:b, \(x) mean(x, na.rm = TRUE))
```

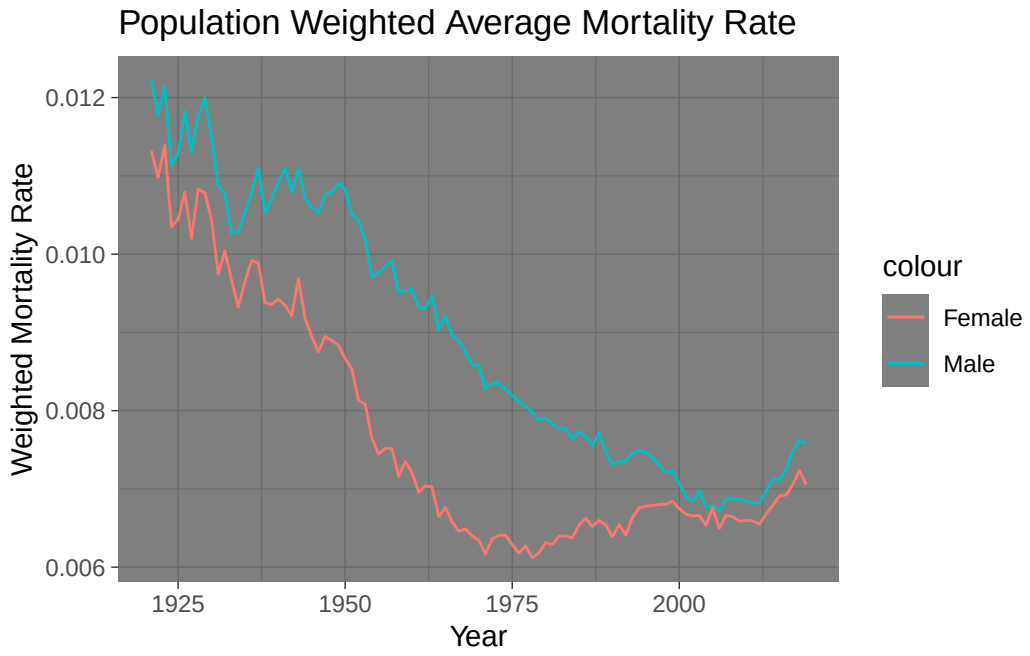
```
# A tibble: 111 x 4
   Age       Male   Female    Total
   <chr>    <dbl>    <dbl>    <dbl>
 1 0     0.0330   0.0256   0.0294
 2 1     0.00396  0.00352  0.00374
 3 10    0.000561 0.000474 0.000509
 4 100   0.138    0.0928   0.0729
 5 101   0.158    0.125    0.0995
 6 102   0.214    0.143    0.114
 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
# i 101 more rows
```

4. Calculate the weighted average mortality rate

```
dp <- read_table("https://www.prdh.umontreal.ca/BDLC/data/ont/Population.txt", skip = 1, c

total_data <-left_join(dm,dp, by = c("Year","Age"))|>
  group_by(Year) |>
  drop_na() |>
  summarize(Avg_Male_Mortality = weighted.mean(Male.x, w=Male.y, na.rm = TRUE),
            Avg_Female_Mortality = weighted.mean(Female.x, w=Female.y, na.rm = TRUE))

ggplot(data = total_data, aes(x = Year)) +
  geom_line(aes(y = Avg_Male_Mortality, color = "Male")) +
  geom_line(aes(y = Avg_Female_Mortality, color = "Female")) +
  labs(title = "Population Weighted Average Mortality Rate",
       x = "Year",
       y = "Weighted Mortality Rate") +
  theme_dark()
```



We notice that the male weighted average mortality rate is higher than the female weighted average mortality rate. However both rates have significantly decreased over the years.

5. Linear Regression

```
dm$Age <- as.numeric(dm$Age)
```

```
Warning: NAs introduced by coercion
```

```
model_data <- dm |>
  filter(Age < 106,Year==2000) |>
  select(Female,Age)
linear_model <- lm(log(Female) ~ Age, data = model_data)
summary(linear_model)
```

```
Call:
lm(formula = log(Female) ~ Age, data = model_data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9692 -0.3194 -0.1341  0.2734  4.7993

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.062281   0.121345  -82.92   <2e-16 ***
Age           0.086891   0.001997   43.51   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6291 on 104 degrees of freedom
Multiple R-squared:  0.9479,    Adjusted R-squared:  0.9474
F-statistic:  1893 on 1 and 104 DF,  p-value: < 2.2e-16
```

The estimated coefficient for Age in our model is 0.086891. This coefficient can be interpreted as the rate of change of female mortality byage.