Dr. Semmelweis and the Importance of Handwashing

Elvin Abdullayev

03-02-2024

1. Loading data

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
yearly <- read_csv("yearly_deaths_by_clinic.csv")
```

```
## Rows: 12 Columns: 4
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): clinic
## dbl (3): year, births, deaths
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
monthly <- read_csv("monthly_deaths.csv")
```

```
## Rows: 98 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl  (2): births, deaths
## date (1): date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
print(yearly)
```

```
## # A tibble: 12 x 4
##     year births deaths clinic
##    <dbl>  <dbl>  <dbl> <chr>
##  1  1841   3036    237 clinic 1
##  2  1842   3287    518 clinic 1
```

```
##  3  1843   3060    274 clinic 1
##  4  1844   3157    260 clinic 1
##  5  1845   3492    241 clinic 1
##  6  1846   4010    459 clinic 1
##  7  1841   2442     86 clinic 2
##  8  1842   2659    202 clinic 2
##  9  1843   2739    164 clinic 2
## 10  1844   2956     68 clinic 2
## 11  1845   3241     66 clinic 2
## 12  1846   3754    105 clinic 2
```

```
print(monthly)
```

```
## # A tibble: 98 x 3
##    date       births deaths
##    <date>      <dbl>  <dbl>
##  1 1841-01-01    254     37
##  2 1841-02-01    239     18
##  3 1841-03-01    277     12
##  4 1841-04-01    255      4
##  5 1841-05-01    255      2
##  6 1841-06-01    200     10
##  7 1841-07-01    190     16
##  8 1841-08-01    222      3
##  9 1841-09-01    213      4
## 10 1841-10-01    236     26
## # i 88 more rows
```

2. Add new column

```
yearly <- yearly %>%
mutate(proportion_deaths = deaths / births)

monthly <- monthly %>%
mutate(proportion_deaths = deaths / births)

print(yearly)
```

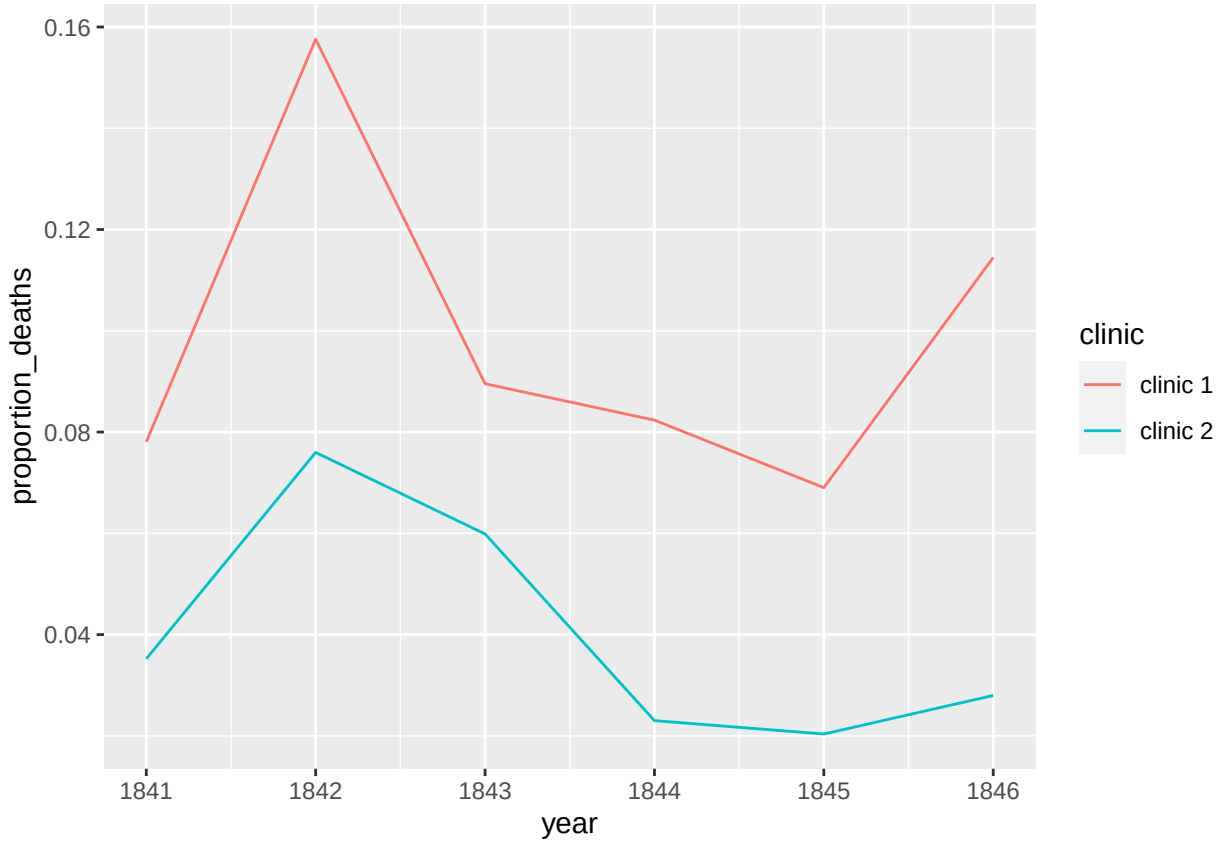
```
## # A tibble: 12 x 5
##     year births deaths clinic   proportion_deaths
##    <dbl>  <dbl>  <dbl> <chr>                <dbl>
##  1  1841   3036    237 clinic 1            0.0781
##  2  1842   3287    518 clinic 1            0.158
##  3  1843   3060    274 clinic 1            0.0895
##  4  1844   3157    260 clinic 1            0.0824
##  5  1845   3492    241 clinic 1            0.0690
##  6  1846   4010    459 clinic 1            0.114
##  7  1841   2442     86 clinic 2            0.0352
##  8  1842   2659    202 clinic 2            0.0760
##  9  1843   2739    164 clinic 2            0.0599
## 10  1844   2956     68 clinic 2            0.0230
## 11  1845   3241     66 clinic 2            0.0204
## 12  1846   3754    105 clinic 2            0.0280
```

```
print(monthly)
```

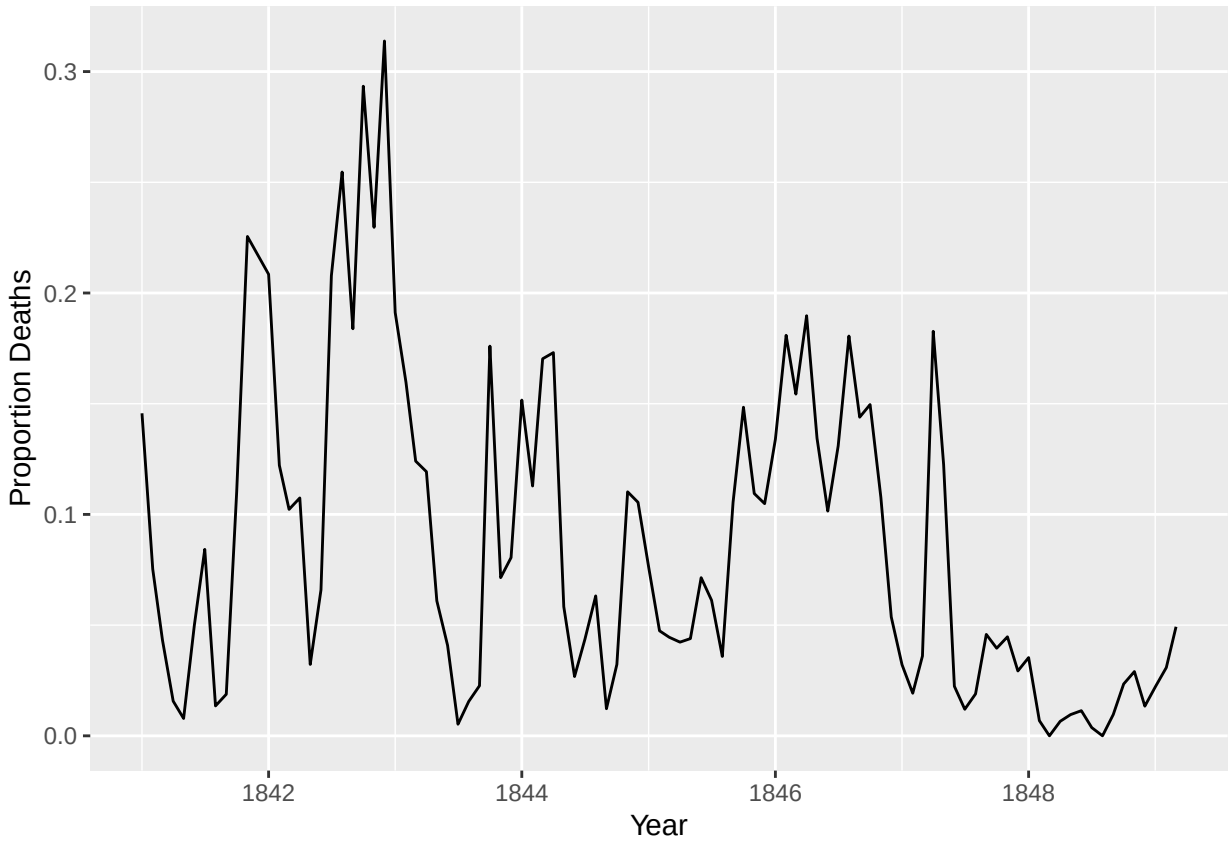
```
## # A tibble: 98 x 4
##    date       births deaths proportion_deaths
##    <date>      <dbl>  <dbl>             <dbl>
##  1 1841-01-01    254     37           0.146
##  2 1841-02-01    239     18           0.0753
##  3 1841-03-01    277     12           0.0433
##  4 1841-04-01    255      4           0.0157
##  5 1841-05-01    255      2           0.00784
##  6 1841-06-01    200     10           0.05
##  7 1841-07-01    190     16           0.0842
##  8 1841-08-01    222      3           0.0135
##  9 1841-09-01    213      4           0.0188
## 10 1841-10-01    236     26           0.110
## # i 88 more rows
```

3. Make a line plot

```
ggplot(yearly, aes(x = year, y = proportion_deaths, color = clinic)) + geom_line()
```




```
ggplot(monthly, aes(date, proportion_deaths)) + geom_line() + labs(x = "Year", y = "Proportion Deaths")
```

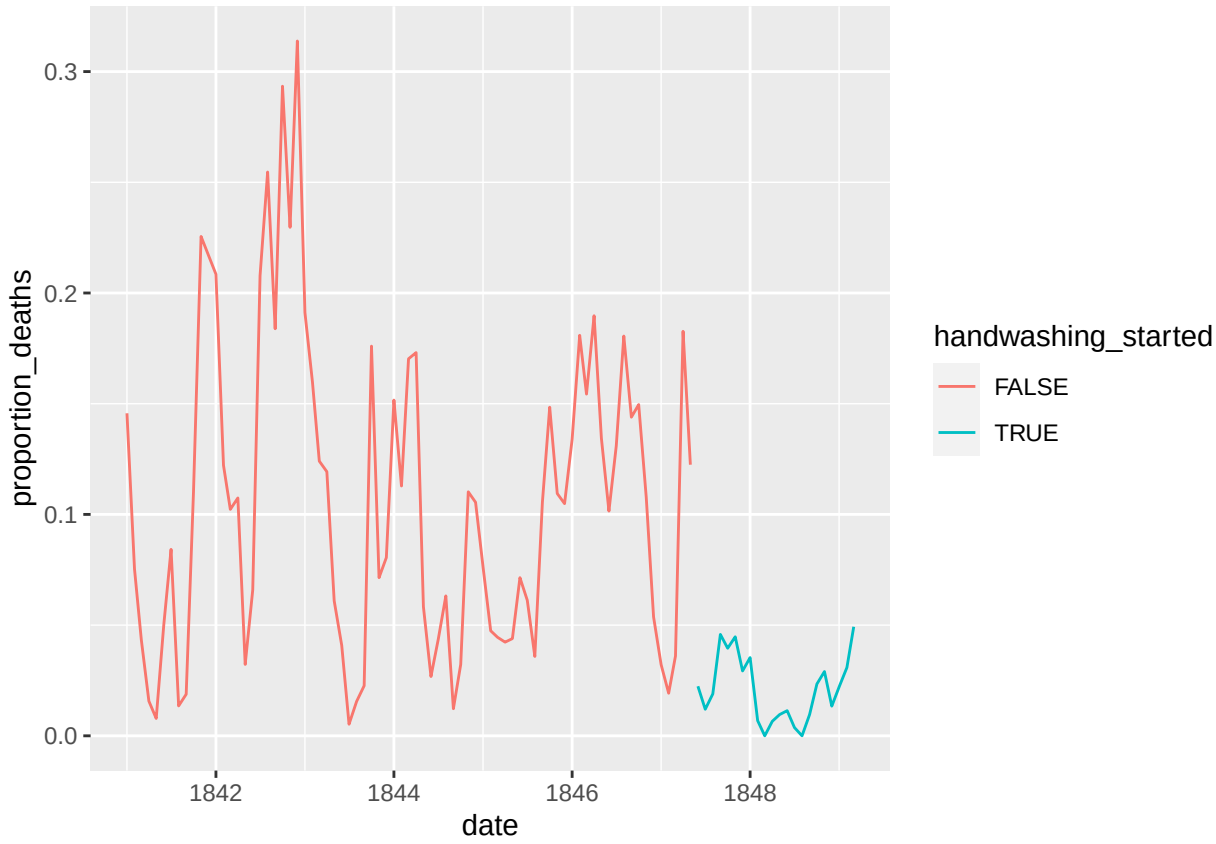


4. Add the threshold and flag and plot again

```
handwashing_start = as.Date('1847-06-01')

monthly <- monthly %>%
mutate(handwashing_started = date >= handwashing_start)

ggplot(monthly, aes(x = date, y = proportion_deaths, color = handwashing_started)) + geom_line()
```



#

5. Find the mean

```
monthly_summary <- monthly %>%
  group_by(handwashing_started) %>%
  summarize(mean_proportion_deaths = mean(proportion_deaths))

monthly_summary
```

```
## # A tibble: 2 x 2
##   handwashing_started mean_proportion_deaths
##   <lgl>                                <dbl>
## 1 FALSE                               0.105
## 2 TRUE                                0.0211
```