Hungarian physician Dr. Ignaz Semmelweis worked at the Vienna General Hospital with childbed fever patients. Childbed fever is a deadly disease affecting women who have just given birth, and in the early 1840s, as many as 10% of the women giving birth died from it at the Vienna General Hospital. Dr.Semmelweis discovered that it was the contaminated hands of the doctors delivering the babies, and on June 1st, 1847, he decreed that everyone should wash their hands, an unorthodox and controversial request; nobody in Vienna knew about bacteria.

You will reanalyze the data that made Semmelweis discover the importance of handwashing and its impact on the hospital.

The data is stored as two CSV files within the `data` folder.

`yearly_deaths_by_clinic.csv` contains the number of women giving birth at the two clinics at the Vienna General Hospital between the years 1841 and 1846.

Column	Description
`year`	Years (1841-1846)
`births`	Number of births
`deaths`	Number of deaths
`clinic`	Clinic 1 or clinic 2

`monthly_deaths.csv` contains data from 'Clinic 1' of the hospital where most deaths occurred.

Column	Description
`date`	Date (YYYY-MM-DD)
`births`	Number of births
`deaths`	Number of deaths

```
# Imported libraries
library(tidyverse)
```

```
# Start coding here..
# 1. Load and inspect the data
yearly = as.data.frame(read_csv("data/yearly_deaths_by_clinic.csv"))
monthly =  as.data.frame(read_csv("data/monthly_deaths.csv"))

print(yearly)
print(monthly)
```

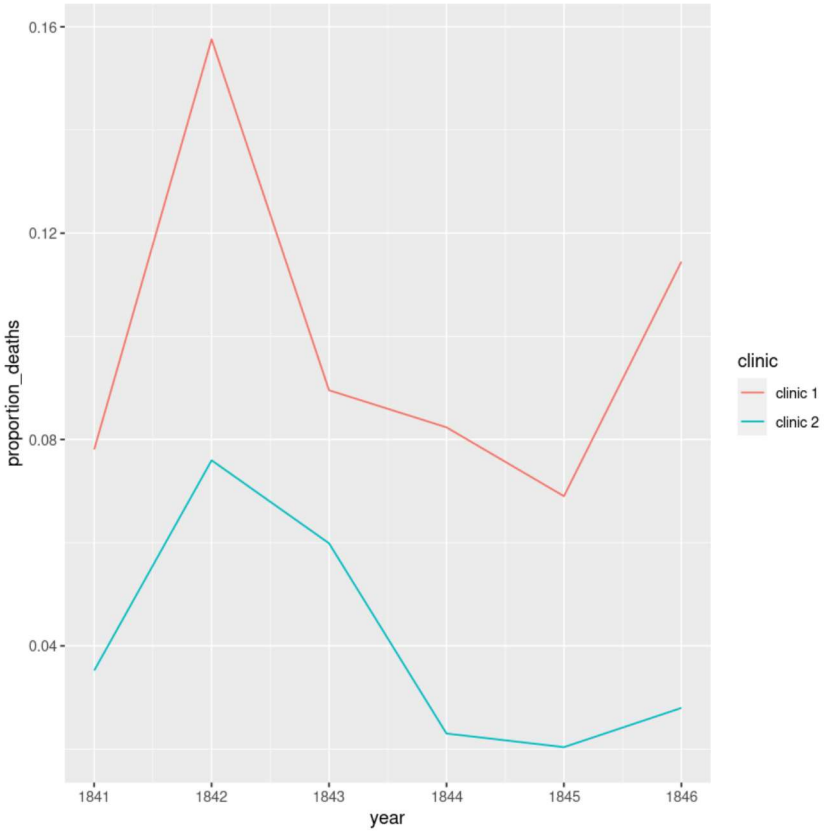
```
Rows: 12 Columns: 4
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (1): clinic
dbl (3): year, births, deaths

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.
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.
```

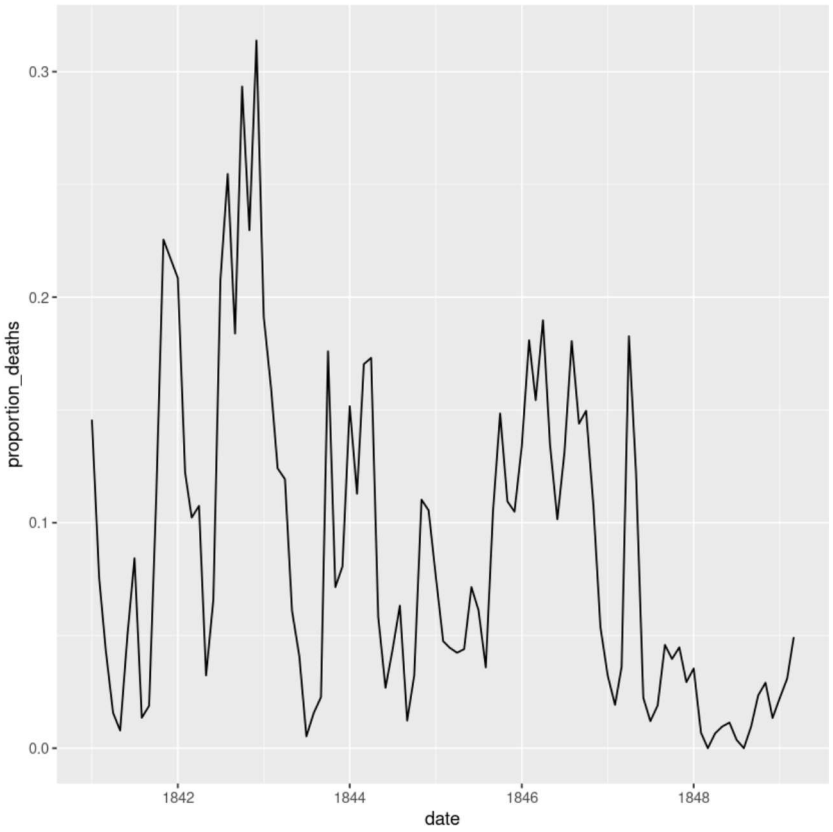
```
   year births deaths   clinic
1  1841   3036    237 clinic 1
2  1842   3287    518 clinic 1
3  1843   3060    274 clinic 1
4  1844   3157    260 clinic 1
5  1845   3492    241 clinic 1
6  1846   4010    459 clinic 1
7  1841   2442     86 clinic 2
8  1842   2659    202 clinic 2
9  1843   2739    164 clinic 2
10 1844   2956     68 clinic 2
11 1845   3241     66 clinic 2
12 1846   3754    105 clinic 2
         date births deaths
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
11 1841-11-01    235     53
12 1842-01-01    307     64
13 1842-02-01    311     38
14 1842-03-01    264     27
15 1842-04-01    242     26
```

```
# 2. Add a new column with the proportions
yearly <- yearly %>%
    mutate(proportion_deaths = deaths/births)

monthly <- monthly %>%
```

```
    mutate(proportion_deaths = deaths/births)
```

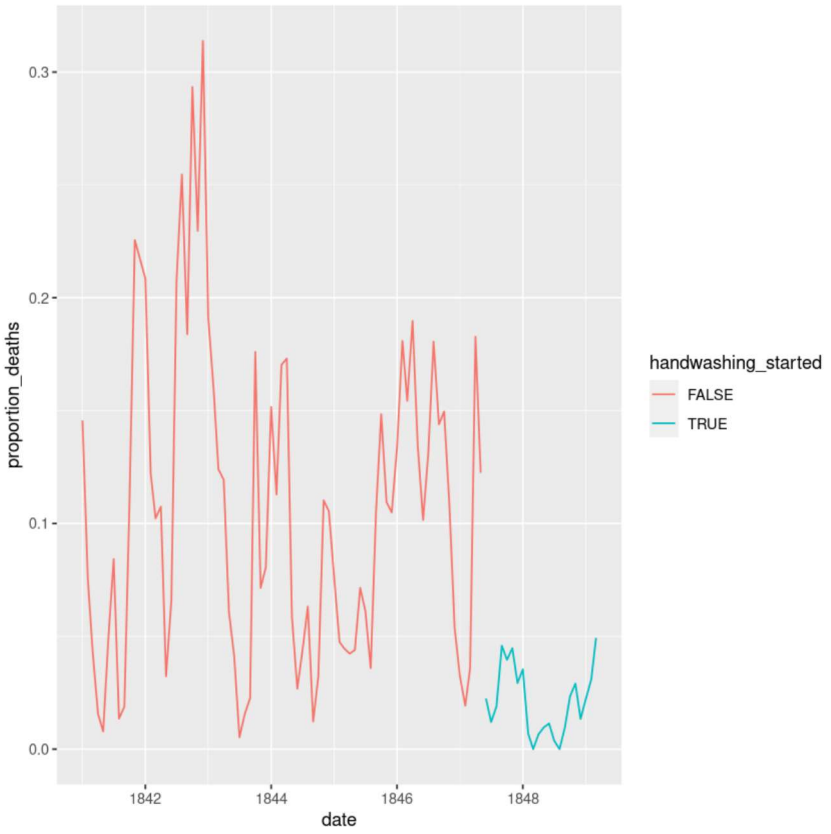
```
# 3. Make a line plot for each data frame
ggplot(yearly, aes(x = year, y = proportion_deaths, color = clinic)) +
    geom_line()

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






```
# 4. Visualize the threshold
```

```
monthly <- monthly %>%
    mutate(handwashing_started = date >= as.Date("1847-06-01"))

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




```
# 5. Calculate the mean proportion of deaths
monthly_summary <- monthly %>%
    group_by(handwashing_started) %>%
    summarise(mean_prop_deaths = mean(proportion_deaths))
monthly_summary
```