THE DISCOVERY OF HANDWASHING

In 1847, the Hungarian physician Ignaz Semmelweis made a breakthough discovery: he discovers handwashing. Contaminated hands was a major cause of childbed fever and by enforcing handwashing at his hospital he saved hundreds of lives.

Import Libraries

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Load The Dataset

In [3]:
```
df=pd.read_csv(r'C:\Users\user\Downloads\yearly_deaths_by_clinic.csv')
df
```

Out[3]:

	year	births	deaths	clinic
0	1841	3036	237	clinic 1
1	1842	3287	518	clinic 1
2	1843	3060	274	clinic 1
3	1844	3157	260	clinic 1
4	1845	3492	241	clinic 1
5	1846	4010	459	clinic 1
6	1841	2442	86	clinic 2
7	1842	2659	202	clinic 2
8	1843	2739	164	clinic 2
9	1844	2956	68	clinic 2
10	1845	3241	66	clinic 2
11	1846	3754	105	clinic 2

Basic Chacks

In [4]:
```
df.head()
```

Out[4]:

	year	births	deaths	clinic
0	1841	3036	237	clinic 1
1	1842	3287	518	clinic 1
2	1843	3060	274	clinic 1
3	1844	3157	260	clinic 1
4	1845	3492	241	clinic 1

In [5]:
```
df.tail()
```

Out[5]:

	year	births	deaths	clinic
7	1842	2659	202	clinic 2
8	1843	2739	164	clinic 2
9	1844	2956	68	clinic 2
10	1845	3241	66	clinic 2
11	1846	3754	105	clinic 2

In [6]:
```
df.describe()
```

Out[6]:

	year	births	deaths
count	12.000000	12.000000	12.000000
mean	1843.500000	3152.750000	223.333333
std	1.783765	449.078476	145.383089
min	1841.000000	2442.000000	66.000000
25%	1842.000000	2901.750000	100.250000
50%	1843.500000	3108.500000	219.500000
75%	1845.000000	3338.250000	263.500000
max	1846.000000	4010.000000	518.000000

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
year      0
births    0
deaths    0
clinic    0
dtype: int64
```

In [9]:
```
df.dtypes
```

Out[9]:
```
year       int64
births     int64
deaths     int64
clinic    object
dtype: object
```

Alarming number of deaths

In [11]:
```
df["proportion_deaths"] = df['deaths'] / df['births']
```

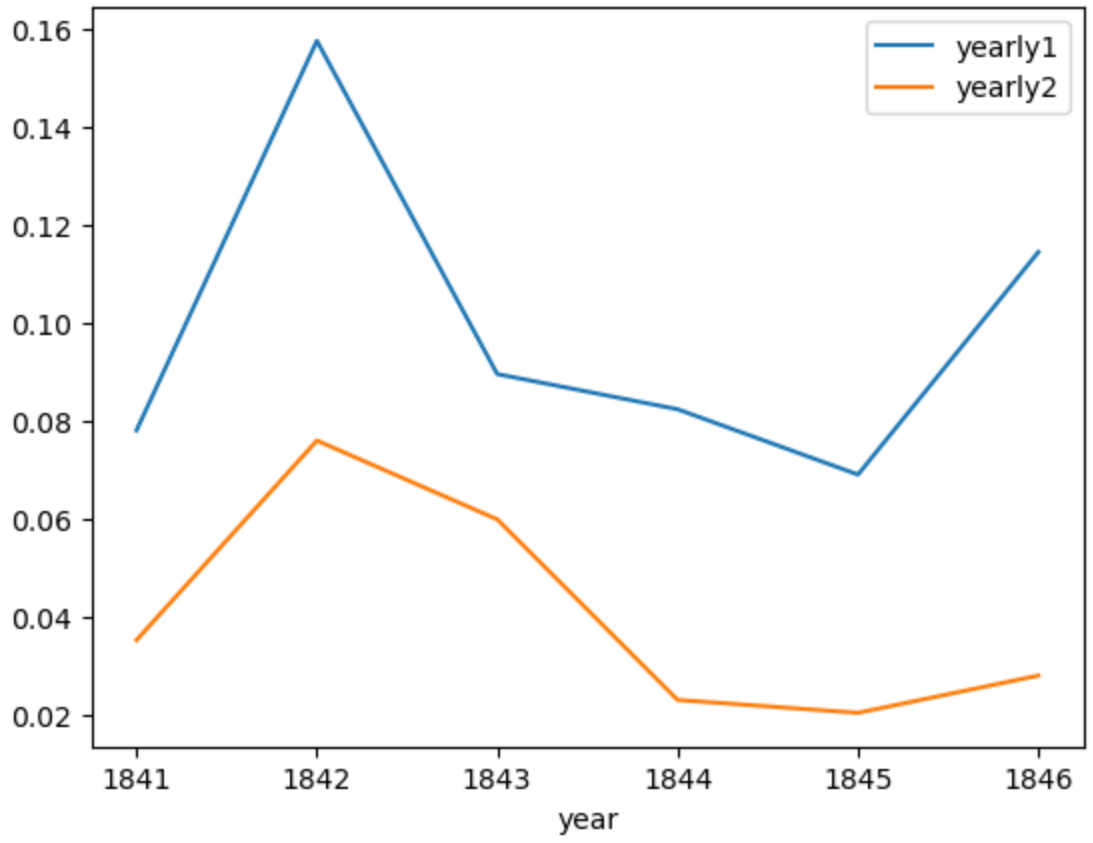
In [13]:
```
df1 = df.iloc[0:6]
df2 = df.iloc[6:12]
print(df1)
```

```
   year  births  deaths    clinic  proportion_deaths
0  1841    3036     237  clinic 1           0.078063
1  1842    3287     518  clinic 1           0.157591
2  1843    3060     274  clinic 1           0.089542
3  1844    3157     260  clinic 1           0.082357
4  1845    3492     241  clinic 1           0.069015
5  1846    4010     459  clinic 1           0.114464
```

Death at the clinic

In [14]:
```
%matplotlib inline
ax = df1.plot(x='year', y='proportion_deaths', label='yearly1')
df2.plot(x='year', y='proportion_deaths', label='yearly2', ax=ax)
```

Out[14]:
```
<Axes: xlabel='year'>
```



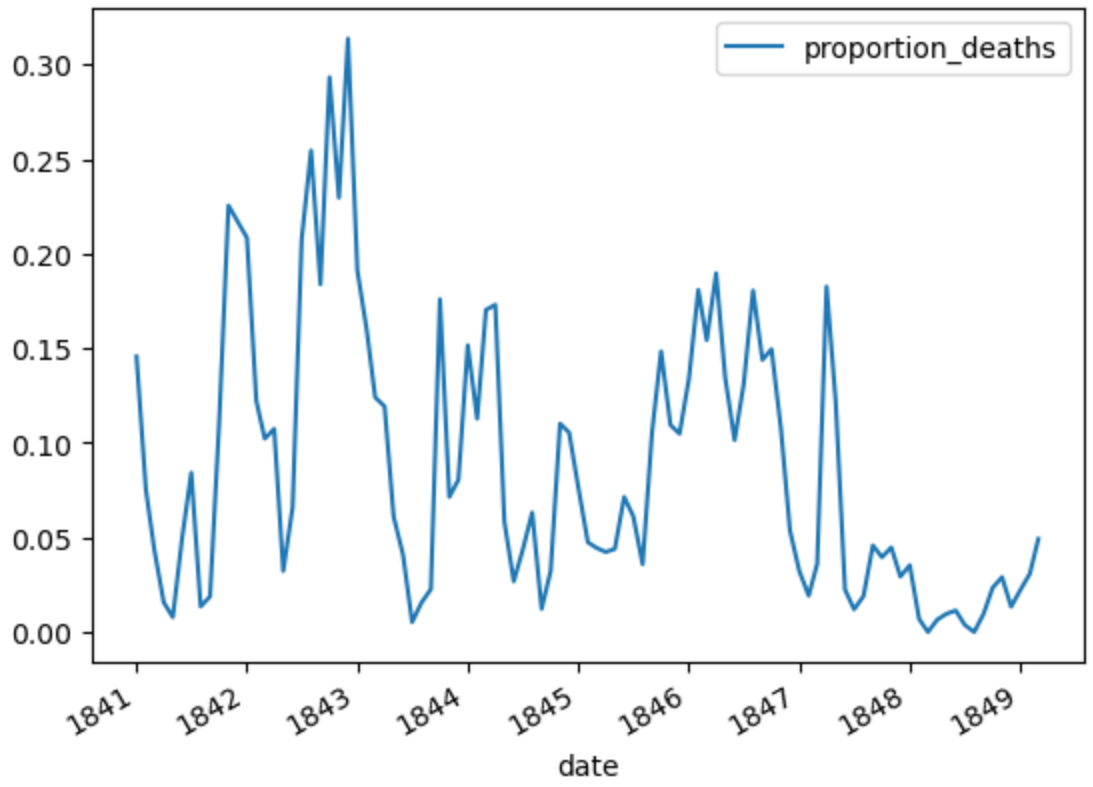
The handwashing begins

In [16]:
```
monthly = pd.read_csv(r'C:\Users\user\Downloads\monthly_deaths.csv', parse_dates=['date'

monthly["proportion_deaths"] = monthly['deaths'] / monthly['births']

print(monthly.head())
```

```
        date  births  deaths  proportion_deaths
0 1841-01-01     254      37           0.145669
1 1841-02-01     239      18           0.075314
2 1841-03-01     277      12           0.043321
3 1841-04-01     255       4           0.015686
          255       2           0.007843
```

The effect of handwashing

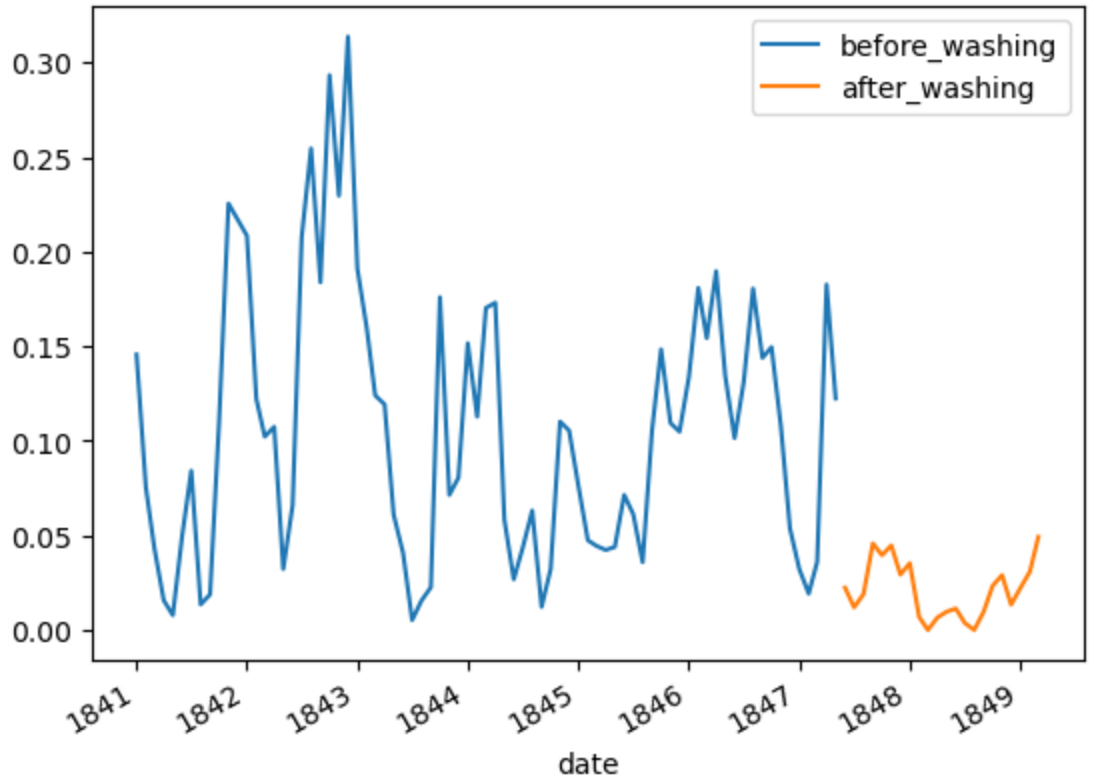
In [17]:
```
ax = monthly.plot(x='date', y='proportion_deaths')
```



The effect of handwashing highlighted

In [18]:
```
import pandas as pd
handwashing_start = pd.to_datetime('1847-06-01')

before_washing = monthly[monthly['date'] < handwashing_start]
after_washing = monthly[monthly['date'] >= handwashing_start]

ax = before_washing.plot(x='date', y='proportion_deaths', label='before_washing')
after_washing.plot(x='date', y='proportion_deaths', label='after_washing', ax=ax)
```

Out[18]:
```
<Axes: xlabel='date'>
```



More handwashing, fewer deaths?

In [19]:
```
before_proportion = before_washing['proportion_deaths']
after_proportion = after_washing['proportion_deaths']
mean_diff = after_proportion.mean() - before_proportion.mean()
mean_diff
```

Out[19]:
```
-0.08395660751183336
```

A Bootstrap analysis of Semmelweis handwashing data

In [20]:
```
boot_mean_diff = []
for i in range(3000):
    boot_before = before_proportion.sample(frac=1, replace=True)
    boot_after = after_proportion.sample(frac=1, replace=True)
    boot_mean_diff.append(boot_after.mean() - boot_before.mean())

confidence_interval = pd.Series(boot_mean_diff).quantile([0.025, 0.975])
confidence_interval
```

Out[20]:
```
0.025   -0.100992
0.975   -0.066386
dtype: float64
```

The fate of Dr. Semmelweis

In [21]:
```
# The data Semmelweis collected points to that:
doctors_should_wash_their_hands = True
```

In [ ]:
```

```