**The Effect of Handwashing**

In June 1846, Dr Semmelweis ordered everyone on his medical staff to start cleaning their hands and instruments not just with soap and water but with a chlorine solution (he didn't know it at the time, but chlorine is an amazing disinfectant). The reason Dr Semmelweis actually chose the chlorine was that he wanted to get rid of any smell on doctors' hands after an autopsy. No one knew anything about bacteria, germs or viruses at the time.

Challenge 1: The Effect of Handwashing

* Add a column called "pct\_deaths" to df\_monthly that has the percentage of deaths per birth for each row.
* Create two subsets from the df\_monthly data: before and after Dr Semmelweis ordered washing hand.
* Calculate the average death rate prior to June 1846.
* Calculate the average death rate after June 1846.

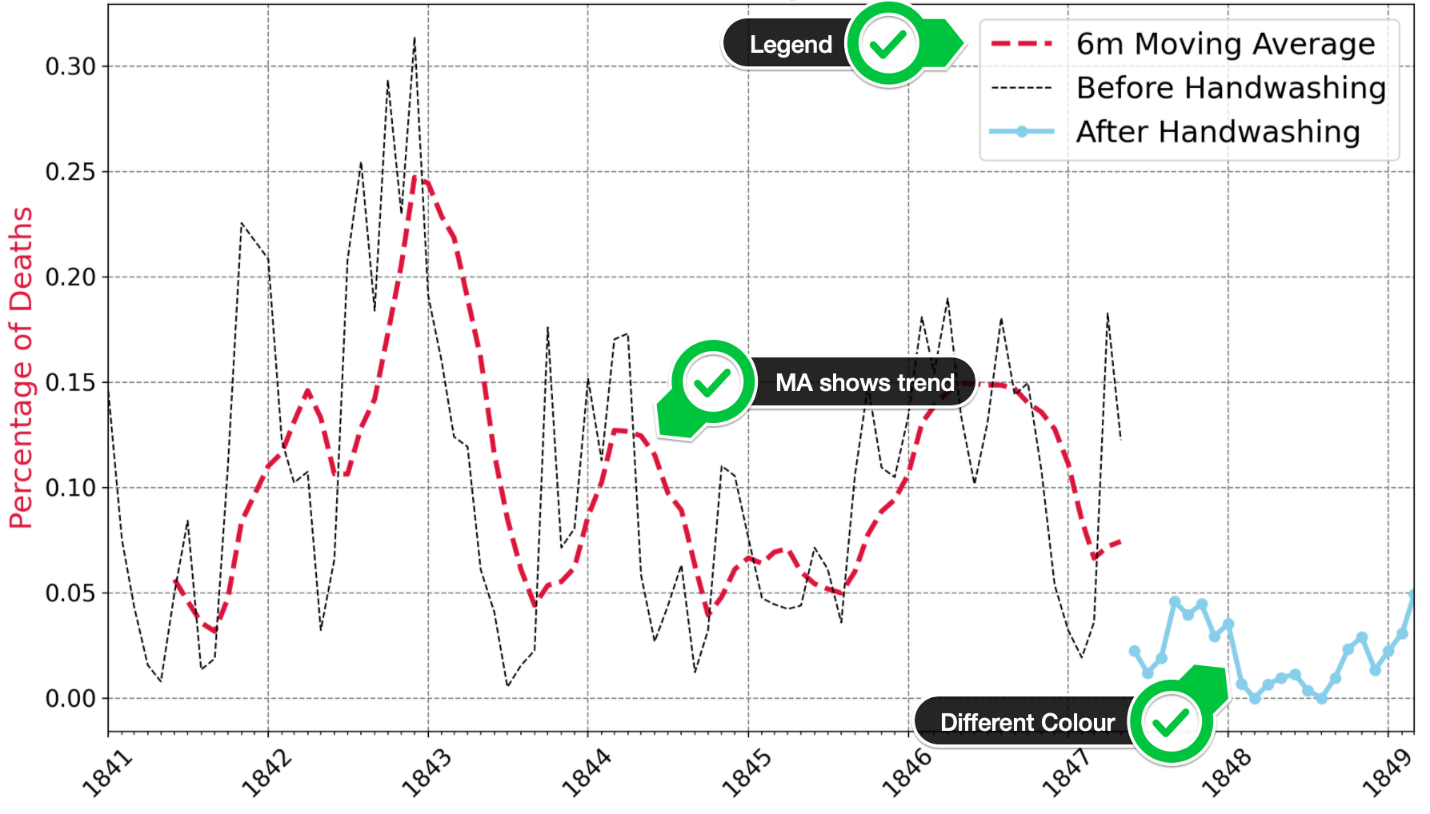
Challenge 2: Calculate a Rolling Average of the Death Rate

Create a DataFrame that has the 6-month rolling average death rate prior to mandatory handwashing.

*Hint*: You'll need to set the dates as the index in order to avoid the date column being dropped during the calculation

Challenge 3: Highlighting Subsections of a Line Chart

Copy-paste and then modify the Matplotlib chart from before to plot the monthly death rates (instead of the total number of births and deaths). The chart should look something like this:



* Add 3 separate lines to the plot: the death rate before handwashing, after handwashing, and the 6-month moving average before handwashing.
* Show the monthly death rate before handwashing as a thin dashed black line.
* Show the moving average as a thicker, crimson line.
* Show the rate after handwashing as a skyblue line with round markers.
* Look at the [code snippet in the documentation to see how you can add a legend](https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.legend.html) to the chart.

.

.

..

...

..

.

.

**Solution to Challenge 1**

We can add a column with the proportion of deaths per birth like so:

1. df\_monthly['pct\_deaths'] = df\_monthly.deaths/df\_monthly.births

Then we can create two subsets based on the handwashing\_start date.

1. before\_washing = df\_monthly[df\_monthly.date < handwashing\_start]
2. after\_washing = df\_monthly[df\_monthly.date >= handwashing\_start]

The death rate per birth dropped dramatically after handwashing started - from close to 10.53% to 2.15%. We can use the colon and dot inside a print statement to determine the number of digits we'd like to print out from a number.

1. bw\_rate = before\_washing.deaths.sum() / before\_washing.births.sum() \* 100
2. aw\_rate = after\_washing.deaths.sum() / after\_washing.births.sum() \* 100
3. print(f'Average death rate before 1847 was {bw\_rate:.4}%')
4. print(f'Average death rate AFTER 1847 was {aw\_rate:.3}%')

**Solution to Challenge 2**

To work out the moving 6-month average we first set the date column as the index. Then we can use the same Pandas functions as in the Google Trends notebook.

1. roll\_df = before\_washing.set\_index('date')
2. roll\_df = roll\_df.rolling(window=6).mean()

**Solution to Challenge 3**

After copy-pasting the previous code for the Matplotlib chart, we just need to change a few things. First, we remove the twin axes. And instead, we plot the three different lines on the same axis. To create the legend, we supply a label to the .plot() function and capture return value in a variable. It's important to notice that .plot() returns more than one thing, so we need to use a comma (,) since we're only grabbing the first item. We can then feed these handles into plt.legend().

1. plt.figure(figsize=(14,8), dpi=200)
2. plt.title('Percentage of Monthly Deaths over Time', fontsize=18)
3. plt.yticks(fontsize=14)
4. plt.xticks(fontsize=14, rotation=45)
6. plt.ylabel('Percentage of Deaths', color='crimson', fontsize=18)
8. ax = plt.gca()
9. ax.xaxis.set\_major\_locator(years)
10. ax.xaxis.set\_major\_formatter(years\_fmt)
11. ax.xaxis.set\_minor\_locator(months)
12. ax.set\_xlim([df\_monthly.date.min(), df\_monthly.date.max()])
14. plt.grid(color='grey', linestyle='--')
16. ma\_line, = plt.plot(roll\_df.index,
17. roll\_df.pct\_deaths,
18. color='crimson',
19. linewidth=3,
20. linestyle='--',
21. label='6m Moving Average')
22. bw\_line, = plt.plot(before\_washing.date,
23. before\_washing.pct\_deaths,
24. color='black',
25. linewidth=1,
26. linestyle='--',
27. label='Before Handwashing')
28. aw\_line, = plt.plot(after\_washing.date,
29. after\_washing.pct\_deaths,
30. color='skyblue',
31. linewidth=3,
32. marker='o',
33. label='After Handwashing')
35. plt.legend(handles=[ma\_line, bw\_line, aw\_line],
36. fontsize=18)
38. plt.show()

