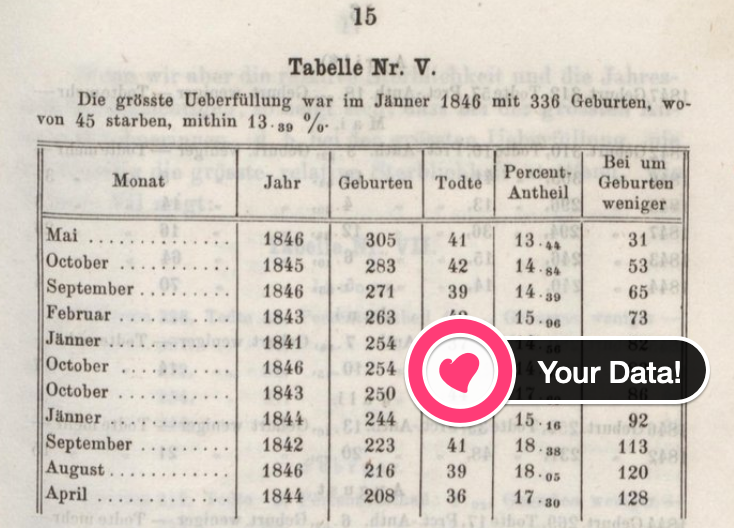
**Preliminary Data Exploration and Visualising Births & Deaths at Vienna Hospital**

You (aka Dr Semmelweis) are working at Vienna General Hospital. Let's take a closer look at the data you've been collecting on the number of births and maternal deaths throughout the 1840s.



Challenge 1: Preliminary Data Exploration

* What is the shape of df\_yearly and df\_monthly? How many rows and columns?
* What are the column names?
* Which years are included in the dataset?
* Are there any NaN values or duplicates?
* What were the average number of births that took place per month?
* What were the average number of deaths that took place per month?

.

.

..

...

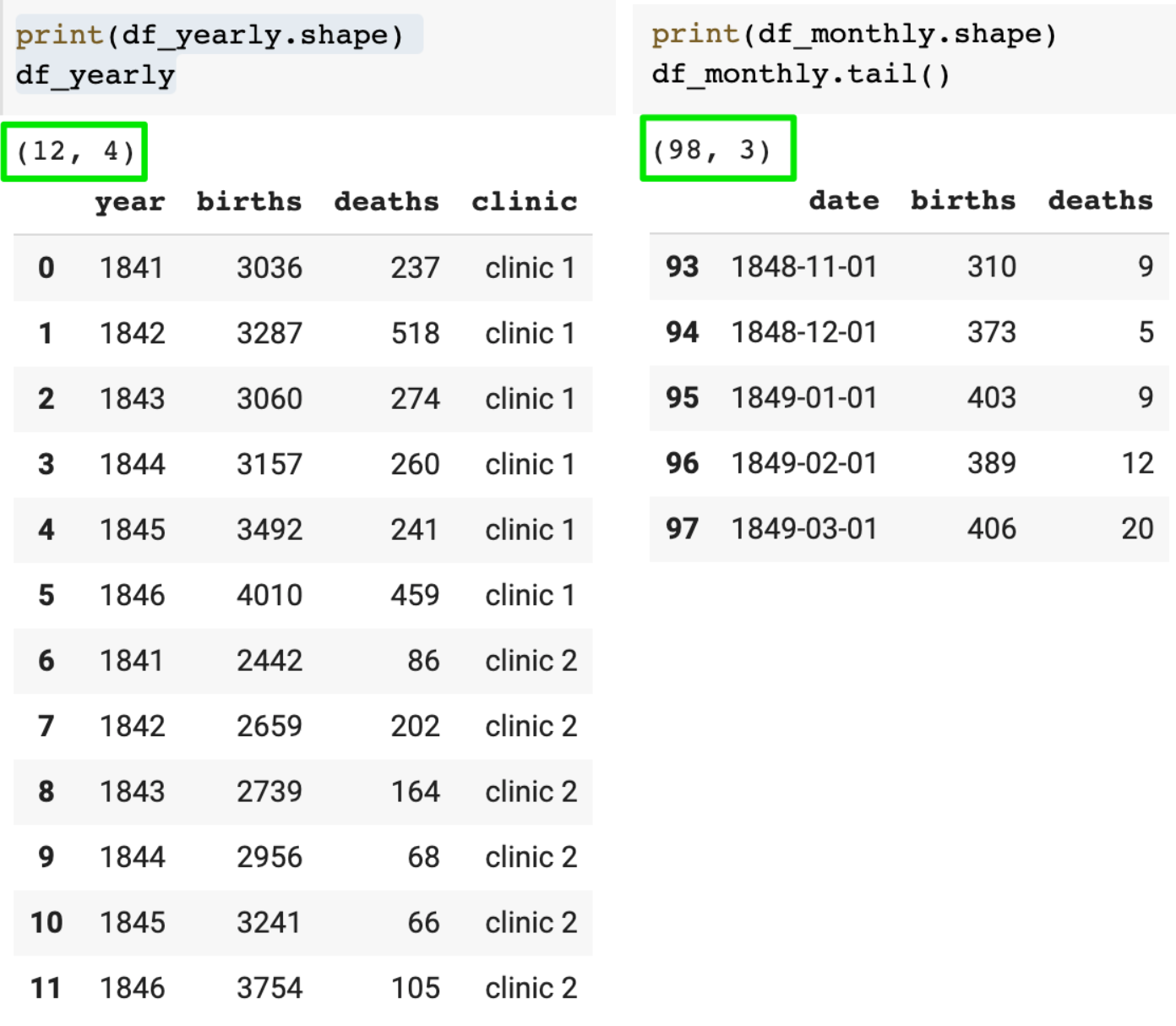
..

.

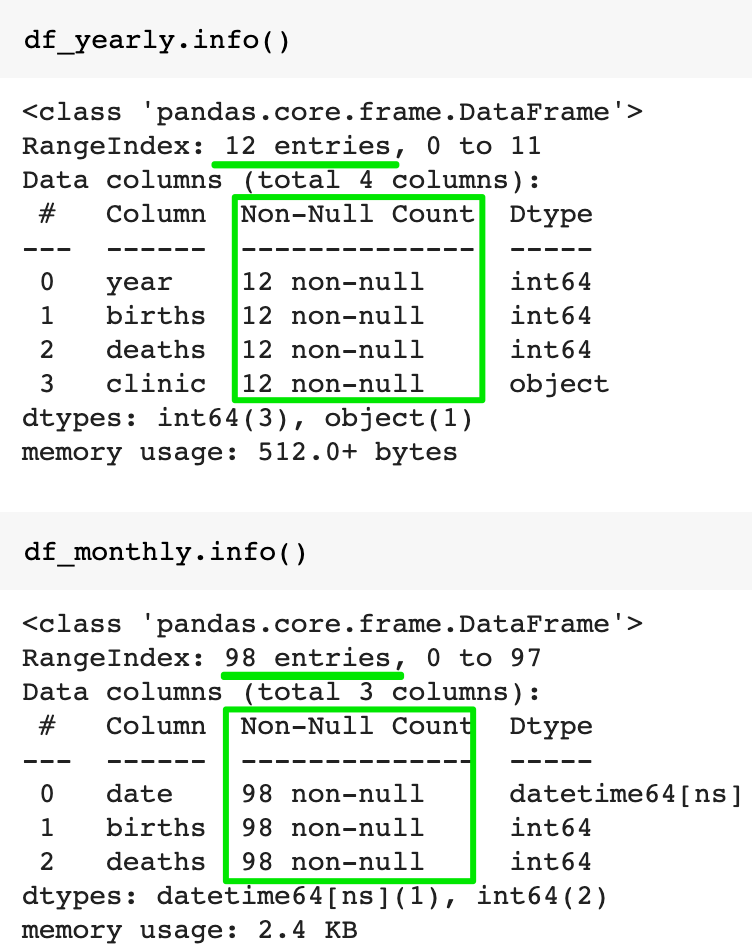
.

**Solution to Challenge 1**

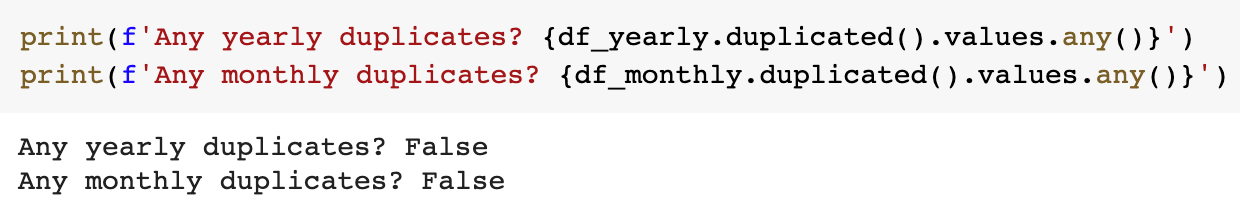
Using .shape, .head(), .tail() we see that the dataset covers the years 1841 to 1849. The two tables report the total number of births and the total number of deaths. Interestingly, the yearly data breaks the number of birthds and deaths down by clinic.



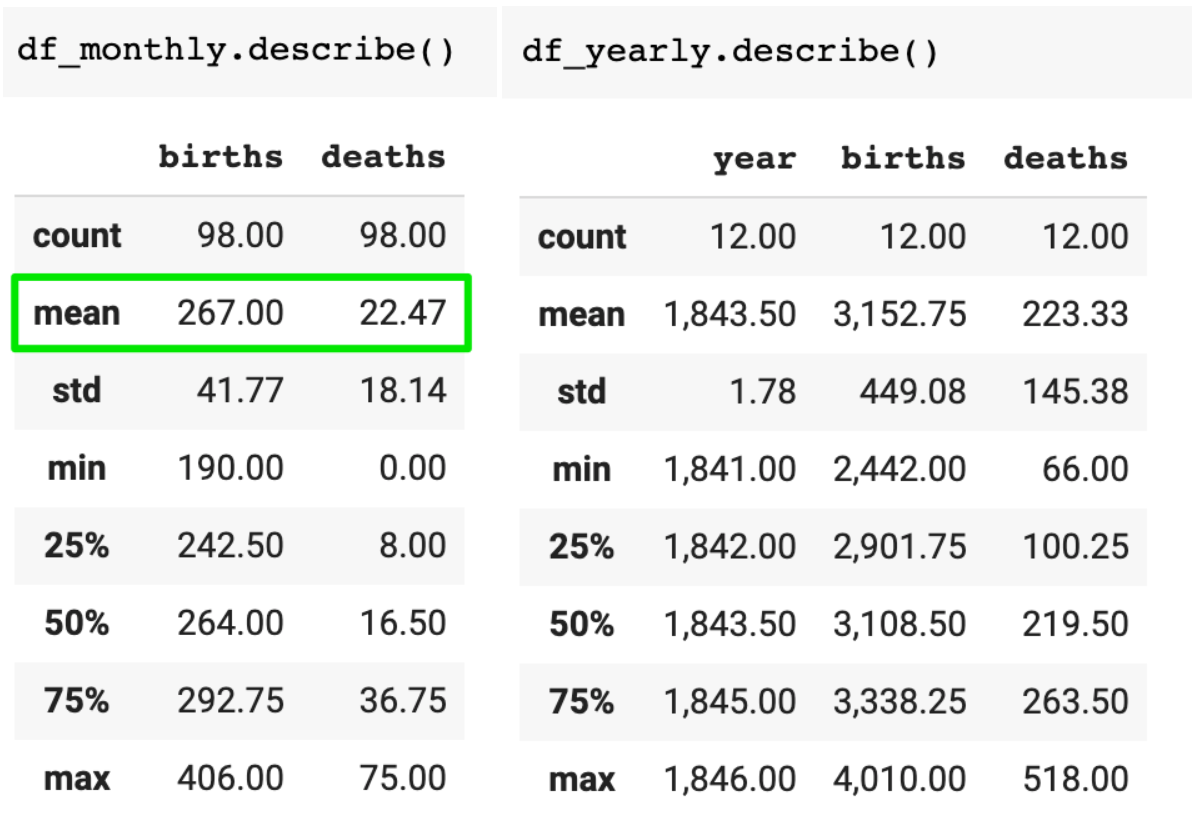
We see that there are no NaN values in either of the DataFrames. We can verify this either with using .info() or using .isna().values.any().



There are also no duplicate entries. In other words, the dataset appears to be clean.



Using .describe() allows us to view some interesting statistics at a glance. We see that on average there were about 267 births and 22.47 deaths per month.



Challenge 2: Percentage of Women Dying in Childbirth

How dangerous was childbirth in the 1840s in Vienna?

* Using the annual data, calculate the percentage of women giving birth who died throughout the 1840s at the hospital.

In comparison, the United States recorded 18.5 maternal deaths per 100,000 or 0.018% in 2013 [(source).](https://en.wikipedia.org/wiki/Maternal_death#:~:text=The%20US%20has%20the%20%22highest,17.8%20per%20100%2C000%20in%202009)

.

.

..

...

..

.

.

**Solution to Challenge 2**

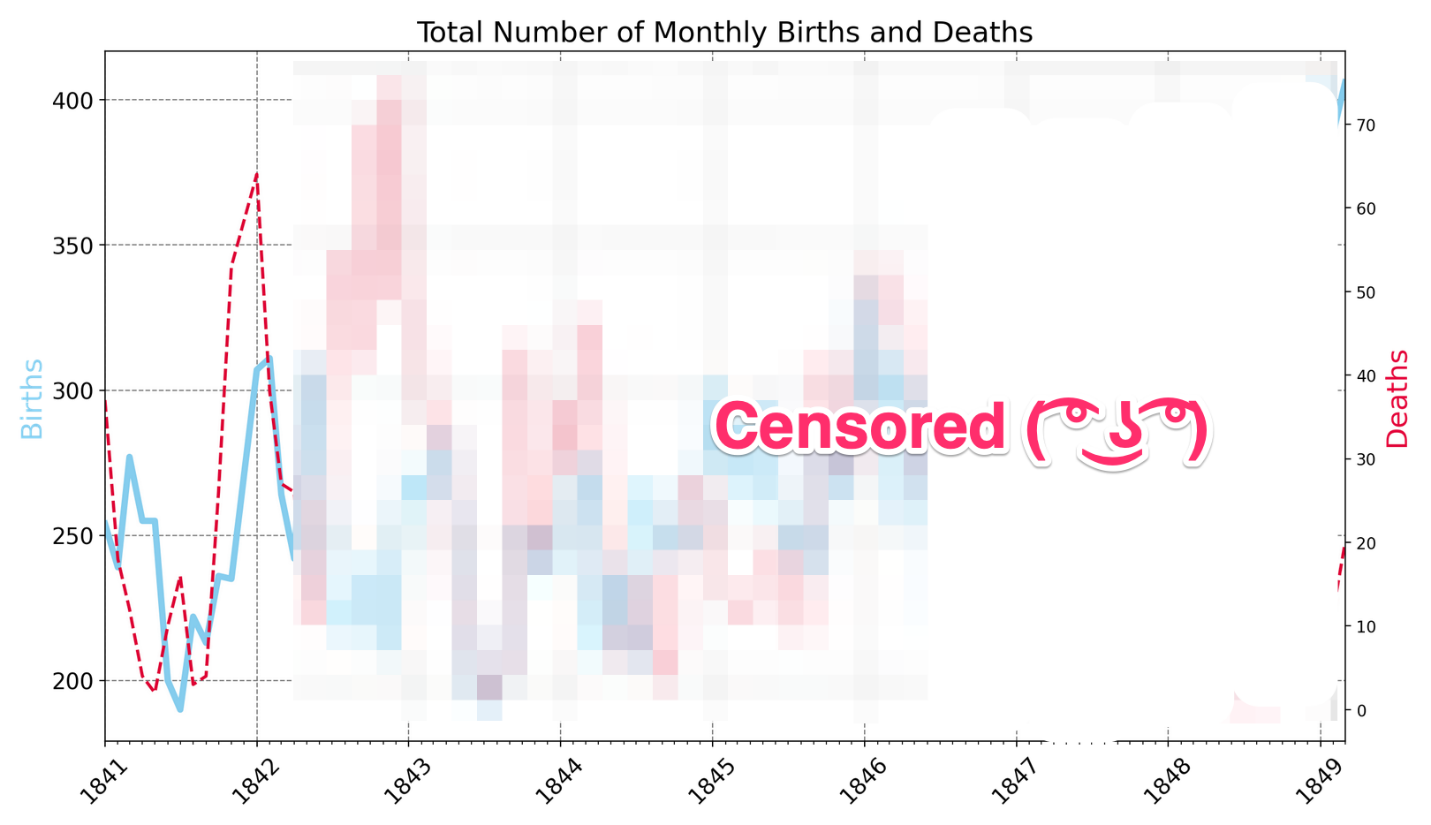
Childbirth was very risky! About 7.08% of women died 💀 in the 1840s (compared to 0.018% in the US in 2013).

1. prob = df\_yearly.deaths.sum() / df\_yearly.births.sum() \* 100
2. print(f'Chances of dying in the 1840s in Vienna: {prob:.3}%')

If someone gave me a bag of 100 M&Ms and told me that 7 of them would kill me, I'd (probably) pass on those M&Ms 🤭. Just saying.

Challenge 3: Visualise the Total Number of Births 🤱 and Deaths 💀 over Time

Create a [Matplotlib chart](https://matplotlib.org/3.3.2/api/_as_gen/matplotlib.pyplot.plot.html) with twin y-axes. It should look something like this:



* Format the x-axis using locators for the years and months (Hint: we did this in the Google Trends notebook)
* Set the range on the x-axis so that the chart lines touch the y-axes
* Add gridlines
* Use skyblue and crimson for the line colours
* Use a dashed line style for the number of deaths
* Change the line thickness to 3 and 2 for the births and deaths respectively.
* Do you notice anything in the late 1840s?

.

.

..

...

..

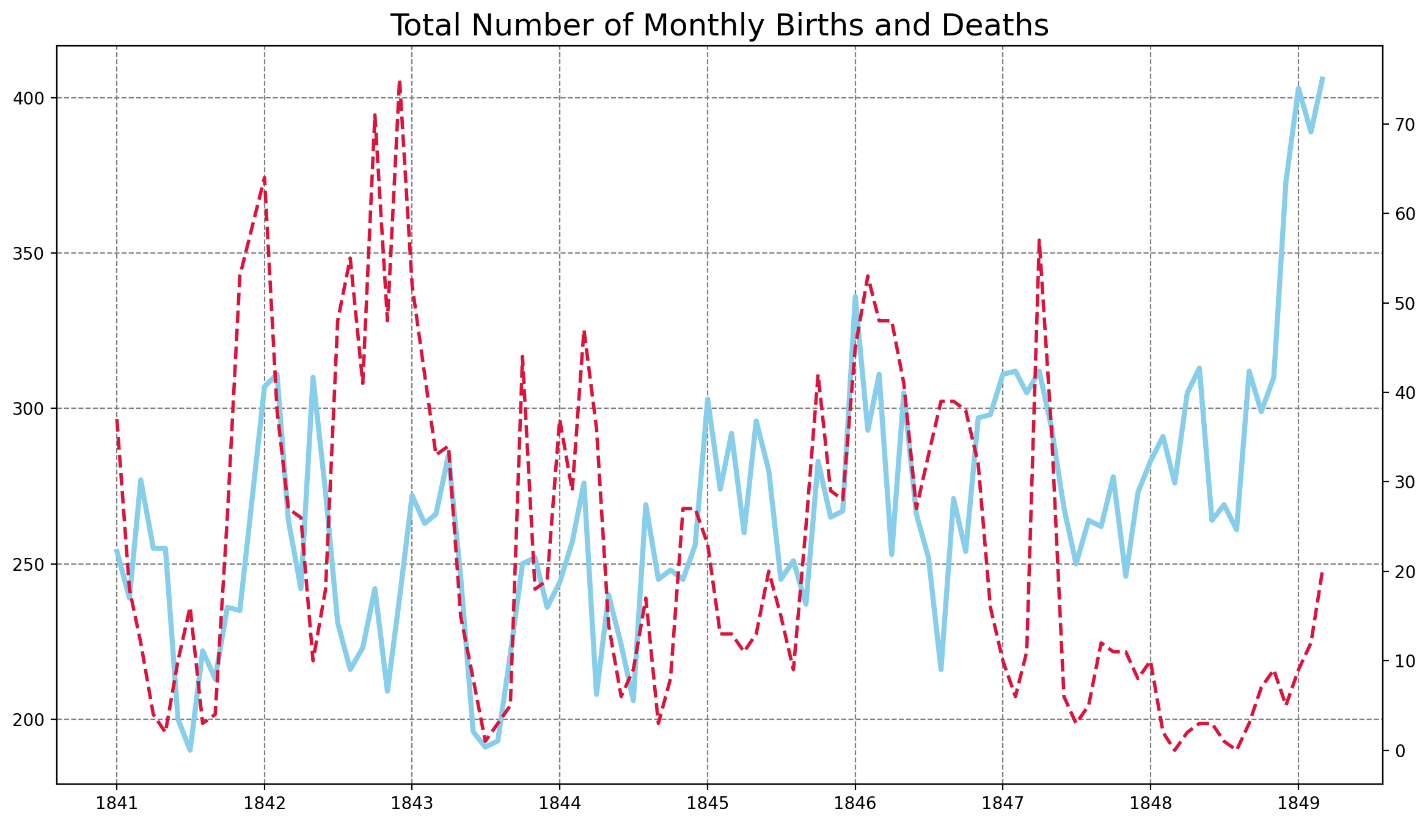
.

.

**Solution to Challenge 3**

Just as in previous notebooks we can use .twinx() to create to y-axes. Then it's just a matter of adding a gird with .grid() and configuring the look of our plots with the color, linewidth, and linestyle parameters.

1. plt.figure(figsize=(14,8), dpi=200)
2. plt.title('Total Number of Monthly Births and Deaths', fontsize=18)
4. ax1 = plt.gca()
5. ax2 = ax1.twinx()
7. ax1.grid(color='grey', linestyle='--')
9. ax1.plot(df\_monthly.date,
10. df\_monthly.births,
11. color='skyblue',
12. linewidth=3)
14. ax2.plot(df\_monthly.date,
15. df\_monthly.deaths,
16. color='crimson',
17. linewidth=2,
18. linestyle='--')
20. plt.show()

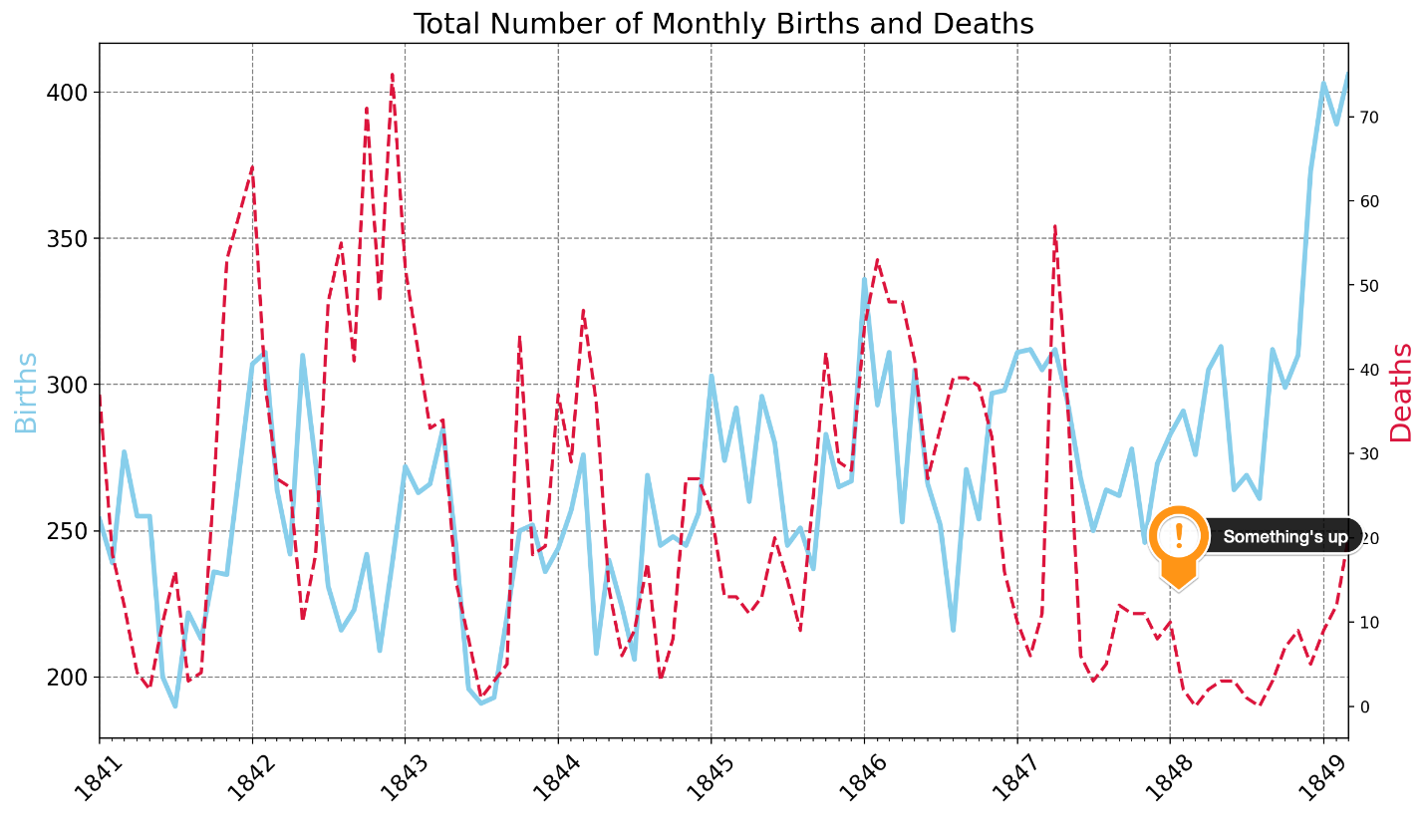


To get the tickmarks showing up on the x-axis, we need to use mdates and Matplotlib's locators.

1. # Create locators for ticks on the time axis
2. years = mdates.YearLocator()
3. months = mdates.MonthLocator()
4. years\_fmt = mdates.DateFormatter('%Y')

We can then use the locators in our chart:

1. plt.figure(figsize=(14,8), dpi=200)
2. plt.title('Total Number of Monthly Births and Deaths', fontsize=18)
3. plt.yticks(fontsize=14)
4. plt.xticks(fontsize=14, rotation=45)
6. ax1 = plt.gca()
7. ax2 = ax1.twinx()
9. ax1.set\_ylabel('Births', color='skyblue', fontsize=18)
10. ax2.set\_ylabel('Deaths', color='crimson', fontsize=18)
12. # Use Locators
13. ax1.set\_xlim([df\_monthly.date.min(), df\_monthly.date.max()])
14. ax1.xaxis.set\_major\_locator(years)
15. ax1.xaxis.set\_major\_formatter(years\_fmt)
16. ax1.xaxis.set\_minor\_locator(months)
18. ax1.grid(color='grey', linestyle='--')
20. ax1.plot(df\_monthly.date,
21. df\_monthly.births,
22. color='skyblue',
23. linewidth=3)
25. ax2.plot(df\_monthly.date,
26. df\_monthly.deaths,
27. color='crimson',
28. linewidth=2,
29. linestyle='--')
31. plt.show()



What we see is that something happened after 1847. The total number of deaths seems to have dropped, despite an increasing number of births! 🤔