Exploring Weather Trends

Name: Mattias Johansson
Company: Volvo Penta

Udacity Nano Degree

Outline

- What tools did you use for each step? (Python, SQL, Excel, etc)

    - SQL was used, as instructed, on the website to extract the data. Python was used for calculations and visualizations. Packages used are pandas, numpy, matplotlib, scipy.

      SQL queries:

        1. Select *
           FROM city_data
           WHERE 'Stockholm' in (city_data.city)
        2. Select *
           FROM global_data

- How did you calculate the moving average?

    - The MA is calculated by using a user-defined *n* window size for each year that averages (calculates the mean) values *n* number of years back.

- What were your key considerations when deciding how to visualize the trends?

    - I want to visualize mainly two things: comparison of *trend* between the two measurements together with correlation. To visualize a trend, a trend line will be a good addition. Therefore, a line plot is sufficient, but also using a simple regression to fit the trend line. Comparisons of the slope coefficient for each model for each dataset can tell us if the temperature is rising in average.

    - I want an easy way to see the *distributions* of the two datasets to see if my city is warmer or colder than the global average. A histogram with sufficient bin size can tell us that (a line plot may also be obvious, but a histogram is good at presenting this).

Analysis

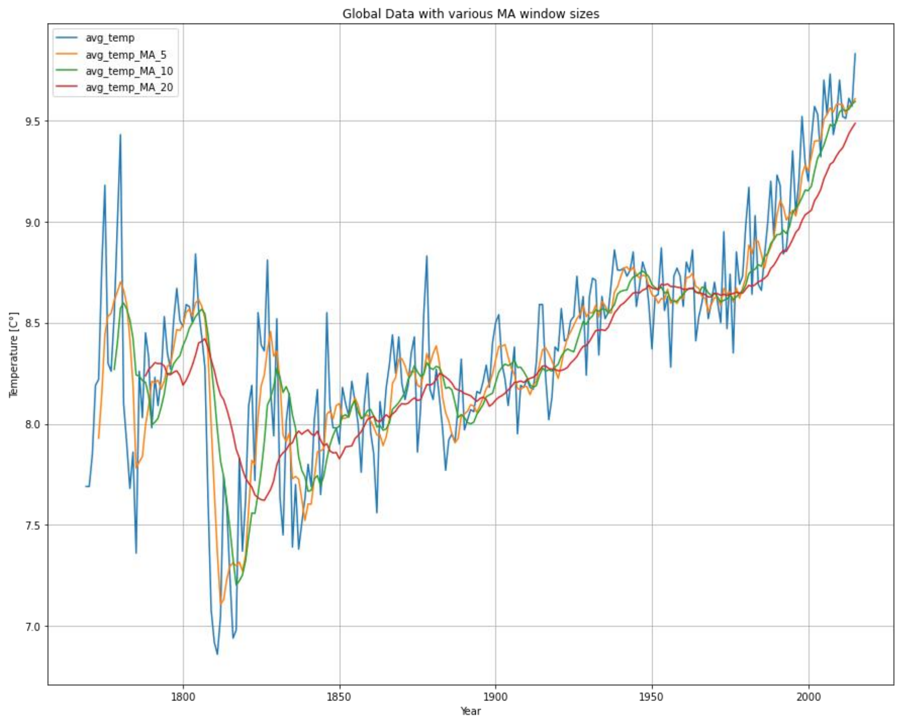


Figure 1: Global dataset with MA for different window sizes, for comparison.

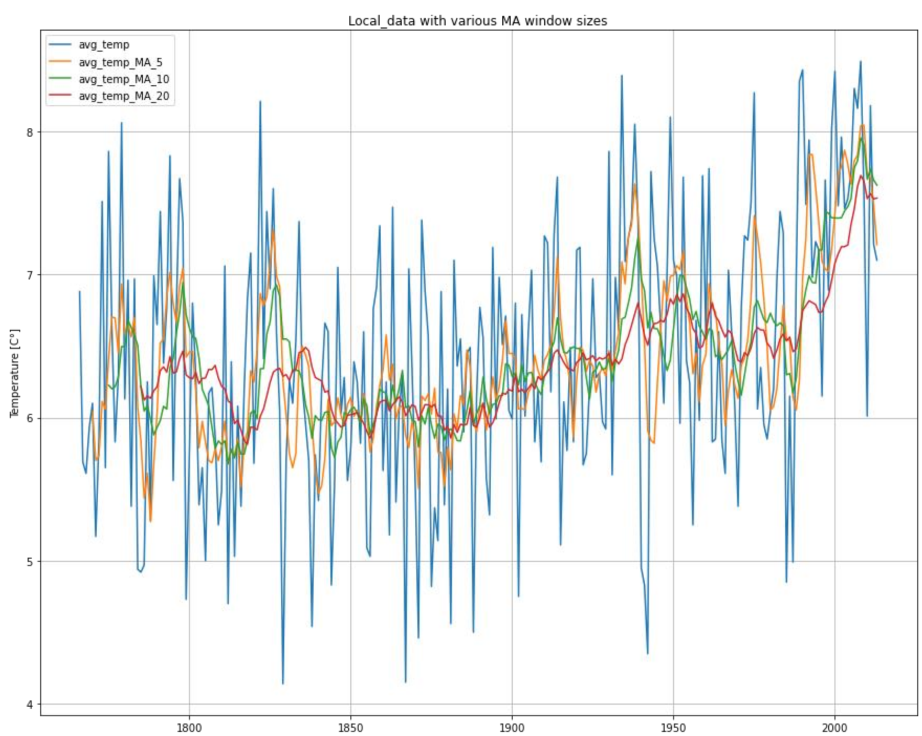


Figure 2: Local dataset with MA for different window sizes, for comparison.

Initially to “remove” the noise in the data, we compare different implementations of a moving average presented in Figure 1 & Figure 2 for each dataset. The higher values on the window size, the smoother the line. Though, when increasing the window size, we also see that the measurement is lagged. When moving further in the report, window size of 20 is used.

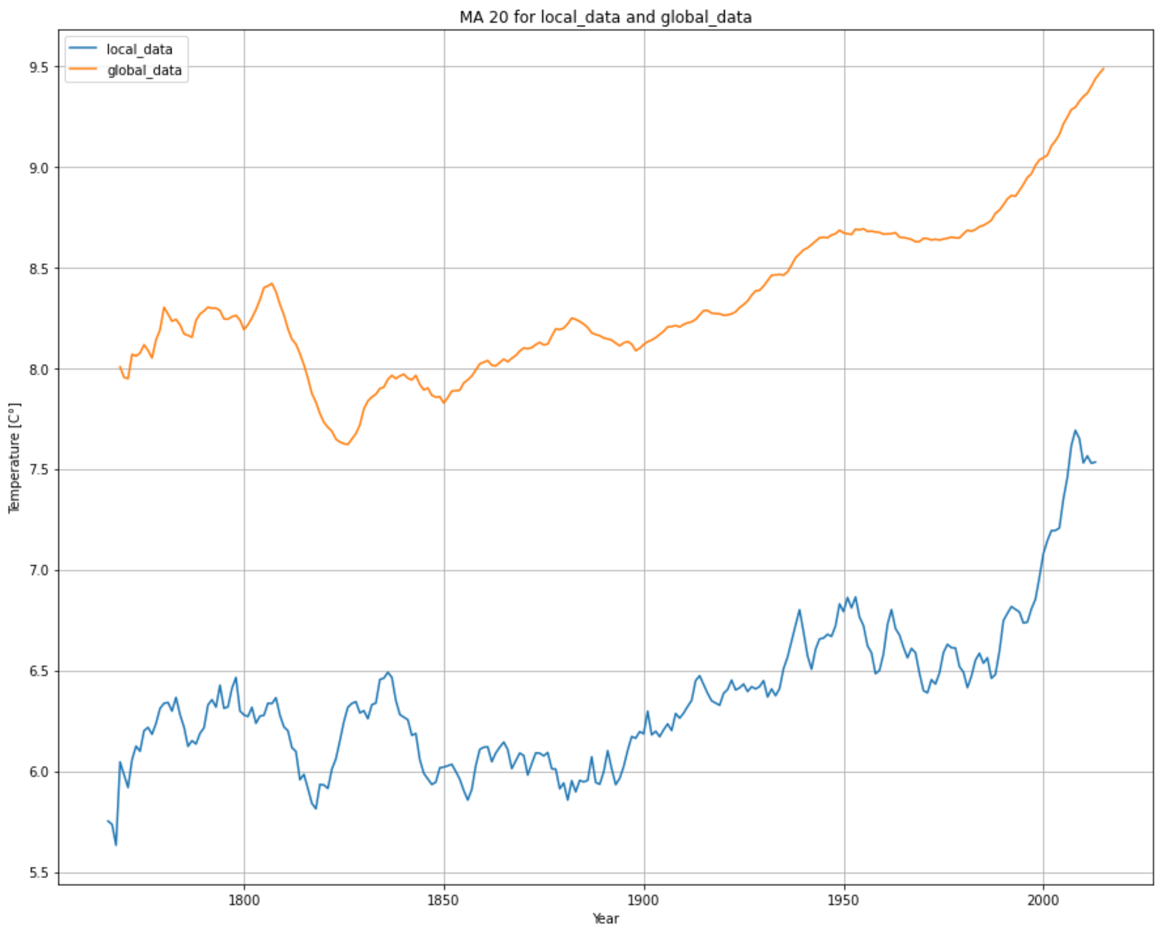
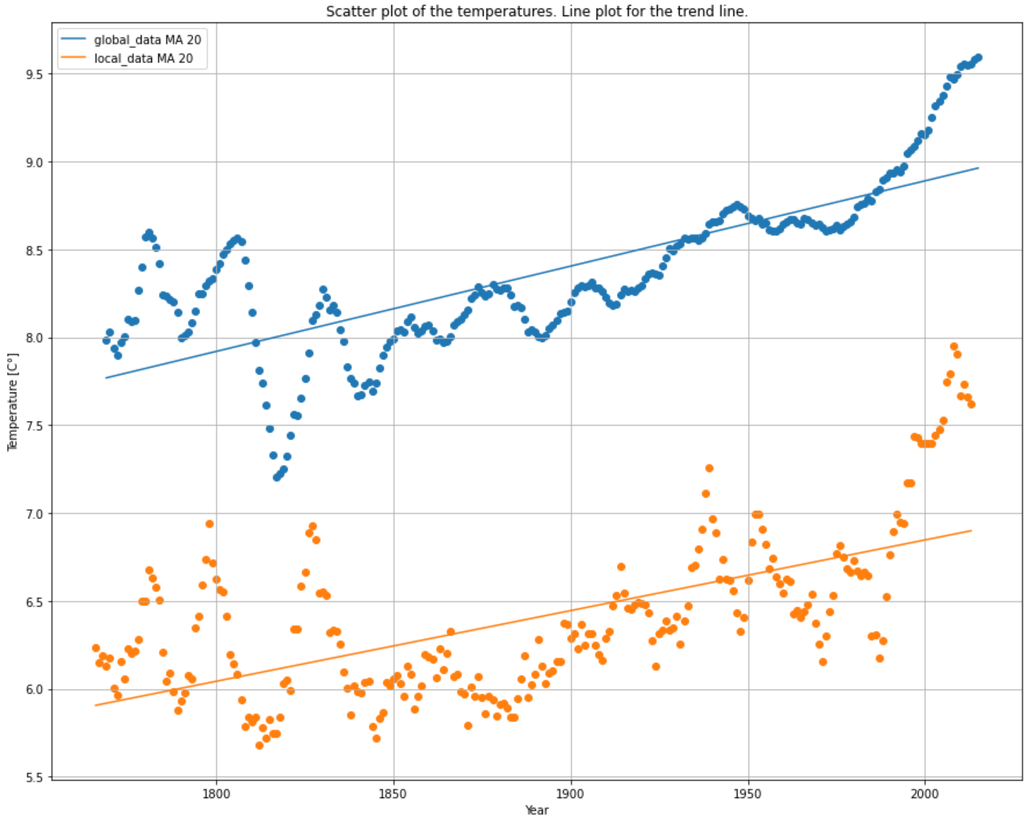


Figure 3: Moving average of both datasets where window size = 20 for MA.

Figure 3 presents the moving average for both the datasets where window size = 20. It is obvious that the global average temperature is higher than the temperature in Stockholm through the years. Noticeable are also the dips and peaks throughout the history that seem to correlate even without using a correlation test.




```
global_data has a slope coefficient of 0.004852178940360643
local has a slope coefficient of 0.004020609930908171
```

Figure 4: Line/Scatter plot of both datasets with a linear regression implemented. The slope coeffiecient is presented under the plot.

Figure 4 presents both the global average temperature and local average temperature with MA with window size 20. The lines that are fitted to the data points have different slopes. The global data line has a higher slope coefficient which implies that the global average temperature is rising faster than the local average temperature in Stockholm. Both datasets present a rise in the average temperature as well, which implies that the temperature increases over time.

The change has increased the last 150 years. Before that, the temperatures seem to only show a seasonal pattern, more than a rising pattern.

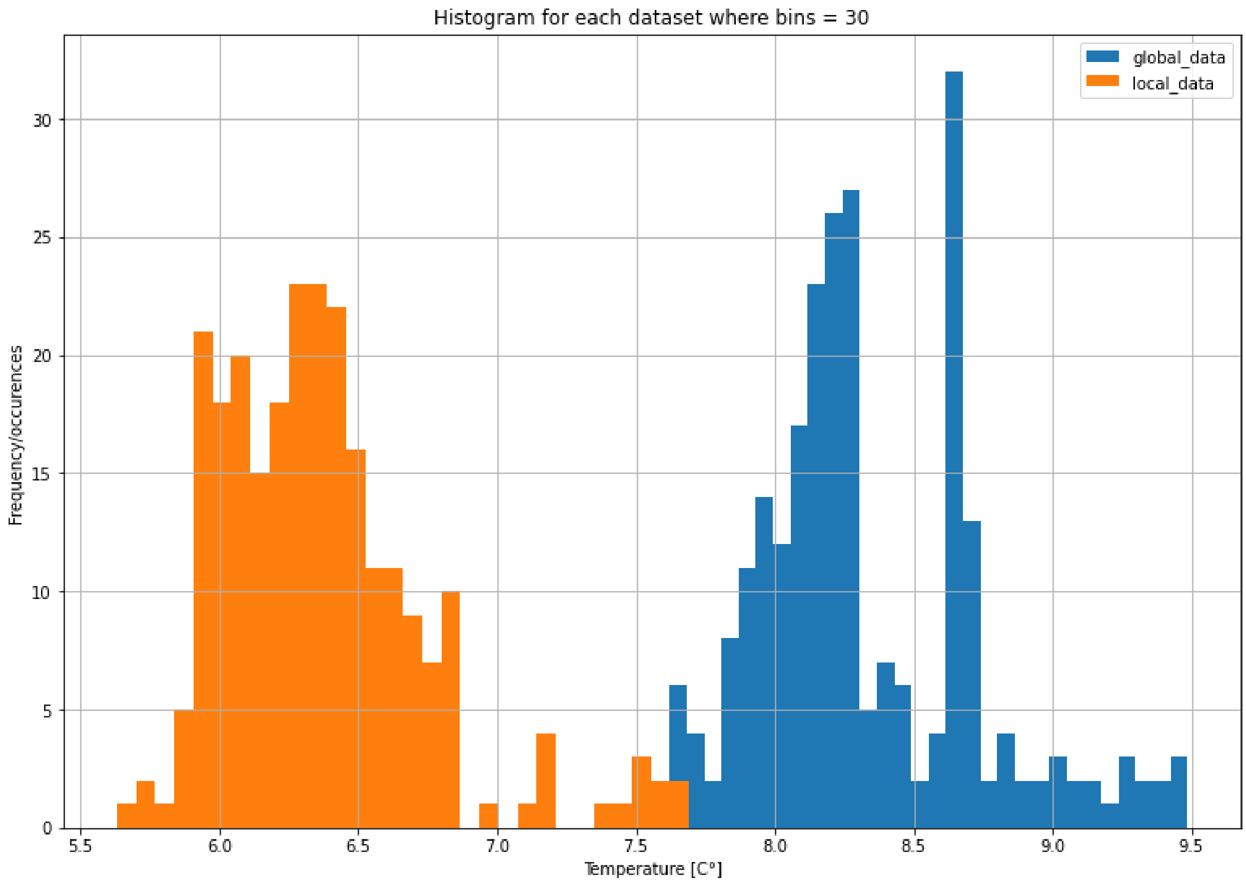


Figure 5: Histogram for each dataset with 30 bins.

Figure 5 presents a histogram for the datasets. The distributions are clearly separated in to two modes, where the local data clearly are distributed around a lower temperature through the years compared to the global data.