

Project 1: Exploring Weather Trends

Lukas Stylianou

September 2018

1 Extracting the data

To extract the data for Berlin, following SQL query was used:

```
1 select year, avg_temp
   from city_data
3 where city = 'Berlin'
```

To extract global data, all data from the global_data table was extracted using following query:

```
1 select *
   from global_data
```

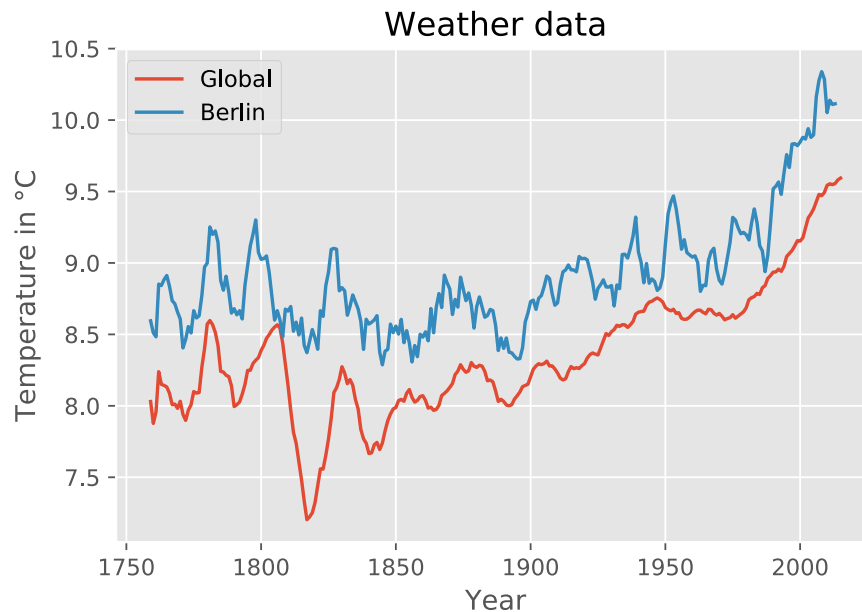
2 Visualizing the data

For the purpose of this task, Python was used with the packages **pandas** and **matplotlib**. The moving average was calculated using the **rolling()** function of **DataFrame** with a window size of **window=10**. Every window was averaged using the **mean()** function. Subsequently, the Python code for calculating moving averages for Berlin and global looks as follows:

```
1 bd['ma']=bd.rolling(window=10)['avg_temp'].mean()
2 gd['ma']=gd.rolling(window=10)['avg_temp'].mean()
```

Key considerations for visualizing the data included how to establish a clear comparison between Berlin and global and how to visualize the trends. As suggested in the task description, a line diagram was chosen for visualization. It was also considered, what window size is reasonable for showing the trends and establishing a good comparison between plots. A window size of 10 years

was chosen, as it seemed a good trade-off between preserving details and, at the same time, keeping the plot smooth enough. Additionally, decades are a quite intuitive time frame when it comes to interpreting the visualization.



3 Observations

Here are four observations about the similarities and/or differences in the trends:

3.1 Is your city hotter or cooler on average compared to the global average?

Yes, on average, Berlin is hotter compared to the global average.

3.2 Has the difference been consistent over time?

No, this can be seen e.g. around 1820 where global temperatures dropped significantly while Berlin temperatures did not drop as much.

3.3 How do the changes in your city's temperatures over time compare to the changes in the global average?

Berlin temperatures behave similar to the global temperatures. Most peaks can be found in both time series. The link between global and city can also be

seen when looking at the trends: An overall rise may be observed in both global and Berlin temperatures.

3.4 What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?

The trend shows that the world is getting hotter. This trend is not consistent, as can be seen from year 1900 on. Before 1900, temperatures have some ups and downs but stay constant in the long run. However, as of 1900 a significant uptrend can be observed.

4 More insights

4.1 What's the correlation coefficient?

The year correlates with the temperature. However, it correlates *strongly* with its moving average.

Berlin	avg_temp	ma
year	0.336616	0.633135

Global	avg_temp	ma
year	0.336616	0.633135

4.2 Can you estimate the average temperature in your city based on the average global temperature?

Yes, this would be possible using a supervised machine learning algorithm to predict a city's temperature from global temperature, using historic temperatures of both city and global as predictors for training the model.

However, as an easy approximation, we can calculate the overall average in the city and calculate its difference to the global average

Global average [°C]	8.35
Berlin average [°C]	8.90

Therefore we can estimate, that on average Berlin is 0.55 °C hotter than the global temperature.

4.3 Multiple cities - Add your favorite cities from around the globe to your visualization. What do you learn about them?

Shanghai and San Francisco are much hotter than Berlin (and the global average). Also it can be seen, that historic data does not reach back as far.

However, the upward trend of temperature is clearly visible across all visualized cities.

