Karina Condeixa - 20/May/2019

Exploring Weather Trends - Project Instructions

Extracting the Data

SQL is used to extract the data and EXCEL to prepare the spreadsheets and graphs.

I calculate the Moving Average in Excel, using the plugin Data Analysis for 10 years, and calculation and average for the last 100 years and dragging until the first year. The former to see the averages in function of the 10 previous years and the latter to see the variations on average from the posterior 100 years.

Procedures for extract data and generate csv files:

First of all, I analyze the contents of the tables and the years they have data of temperature averages and the columns they have.

Since the city_data has the two columns (city, country) like in city_list, it is not needed to join them, I can use just city_data.

To have a table with averages of temperature from Berlin (city I live) and global, the given query is used:

```
SELECT g.year,c.avg_temp AS avg_temp_berlin,g.avg_temp AS avg_temp_global
FROM global_data g
LEFT JOIN city_data c
ON c.year=g.year
WHERE city LIKE 'Berlin';
EVALUATE
Download CSV
```

I use and intersection left between global_area and city_data and put the global as left because this data is more complete (in cities, there are null cells). Thus, the results are similar of union of tables (outer join);

To have a table with averages of temperature from cities in Germany, the given query is used:

```
SELECT year,city,avg_temp
FROM city_data
WHERE country LIKE 'Germany';
EVALUATE
Download CSV
```

Karina Condeixa - 20/May/2019

To have a table with averages of temperature from cities from Brazilians cities that are used for correlation coefficient, the given query is used:

```
SELECT year,city,avg_temp
FROM city_data
WHERE country LIKE 'Brazil';
EVALUATE
Download CSV
```

Analyzing the Data and preparing graphs

The csv files for Berlin vs global and for Berlin and other German cities are opened in Excel.

These processes are made in both csv files:
- Create new columns calculating the moving average for 10 years and insert a line graph.
- Create new columns calculating the retroactive moving average for the last 100 years and insert a line graph.

The file with data for Berlin and other German cities has to be adapted putting results for Hamburg and Munich in columns aligned with Berlin for the corresponding year. I made it manually using a filter by city and using ctrl+x and ctrl+v.

The approach for the last 100 years retroactively is examined to test if it is really true that global warming has been steadily increasing in the last hundred years.

Correlation coefficient:

Correlation coefficient is a numerical value that can be used to compare the relationship between variables, which in our case, are the trends of temperature variations.

The global correlation coefficient is compared with those of German cities and also with Brazilian cities (Rio de Janeiro, Porto Alegre and Recife). The goal is to compare the country I live and the country I came from.

I create a new sheet in Excel and insert the Brazilian data manually using ctrl+x and ctrl+v.

I use the temperatures as they were retrieved (not with moving average) to calculate the coefficients using the toolpak Data Analysis – Correlation.

Karina Condeixa - 20/May/2019

Graphs and key observations:

Figures 1a and 1b compare Berlin and global moving averages. The 1a shows variations from the average of temperature (°C) on the previous 10 years and the 1b shows variations from the average of temperature (°C) on a last 100 years-based approach, retroactively.
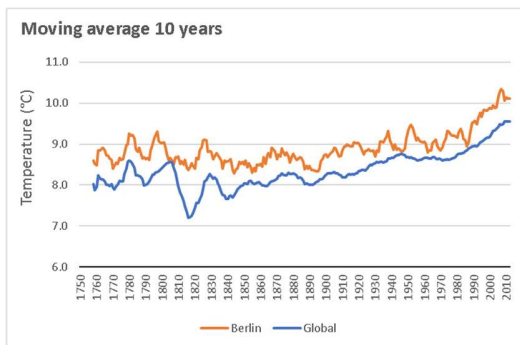

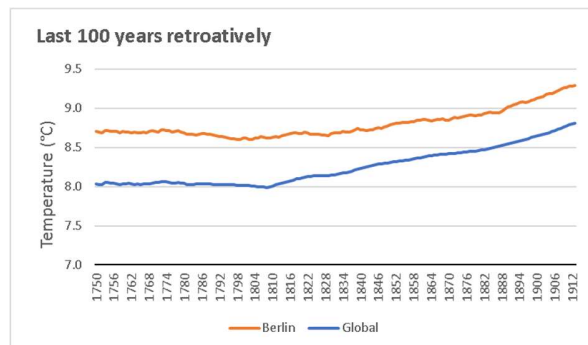*Figure 1a – Berlin vs Global: Moving average 10 years*

Figure 1b – Berlin vs Global: Moving average last 100 years

Figure 1a shows that:

- It used to occur variations in temperature, such as cycles of approximately 20 years, until about 1910, when the temperature increase became more pronounced and more continuous.
- An exception was near 1810 when the global temperature fell by 2 degrees Celsius, such an occurrence was not observed in Berlin.
- In addition, Berlin's temperature is shown to be much more unstable than in the rest of the world, as expected.

Figure 1b shows that the global and Berlin temperature averages tended to increase gradually from 1810, and it seems that cycles are increasing to around 50 years.

Both figures 1a and 1b show that the increase in the average temperature start to happen from earlies 1800, but the increase in overall averages has been shown to be faster than that of Berlin.

Figures 2a and 2b compare Berlin and other German cities' moving average of temperature (Hamburg and Munich). Figure 2a shows variations from the average of temperature (°C) on the previous 10 years and the Figure 2b shows variations from the average of temperature (°C) on a last 100 years-based approach, retroactively.
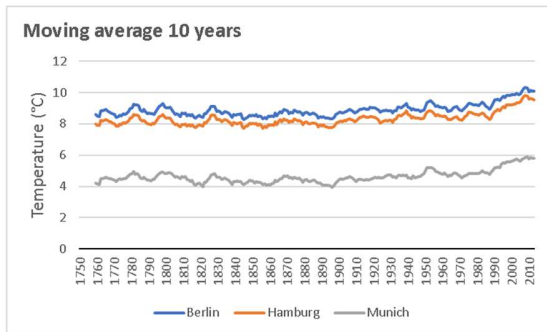
Karina Condeixa - 20/May/2019

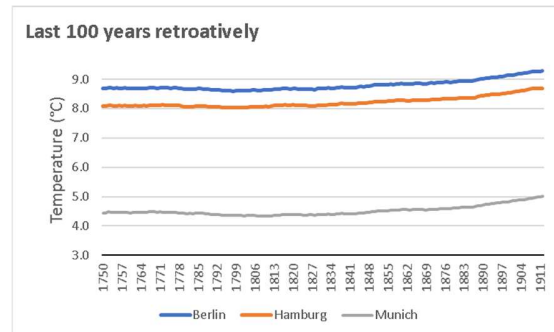Figure 2a – German cities: Moving average 10 years

Figure 2b – German cities: Moving average last 100 years

Figure 2a shows that between 1750 and 2010, there was an increase of 1 to 2 degrees Celsius in the average temperature, with a sharp increase from 1980.

Figure 2b shows that the temperature is no longer stable around 1820.

Both figures show that Berlin is hotter than the other cities and Munich is much colder than the others, but the variations are in line with each other.

Correlation coefficient:

Table below shows the founded coefficients. Where *Column 1* represents the global, Column 1 represents Berlin, Column 2 represents Hamburg, Column 3 represents Munich, Column 4 represents Rio de Janeiro, Column 5 represents Porto Alegre, and Column 6 represents Recife.

Table 1 - Correlation coefficients

	Column 1
Column 1	1
Column 2	0.548006
Column 3	0.544942
Column 4	0.615142
Column 5	0.798931
Column 6	0.777356
Column 7	0.876041

Considering the global temperatures as comparison parameter, the coefficient from the Brazilians cities were demonstrated to be closer that the global than the Germans. That is, those cities had higher increase of temperature, but still smaller than the global approach.

Conclusion

The main similarity found are as follows:

Karina Condeixa - 20/May/2019

- Among the German's cities and the global approach, it is shown to be the trend in increasing temperature and increase in periods between cycles of usual variations from the early 1800s.

As main differences found are:

- An increase in temperature in Berlin (and probably in the other German cities investigated) was slightly lower than the global one.
- A more significant instability in temperature in Berlin than in the rest of the world.
- The global temperature averages showed a great and inexperienced reduction around 1810 that was not observed in Berlin

Based on my analysis, I could recommend an investigation of historical events which may have increased the temperature. Particularly, about events around 1810, when the global temperature was demonstrated to decrease strongly, different from Berlin.

Moreover, a comparison of correlation coefficients shows that the Brazilians cities suffer more from global warming than the cities in Germany, however global warming averages are even more significant.

Self-analysis

 SQL:
+ I was able to extract the data in two csv files.

- I could not redesign a table to have years in rows and German cities in in columns.

Spreadsheet:
+ I was able to calculate moving averages in Excel.

- I do not feel confident working on a large amount of data in Excel. I think it could be still harder if I were trying using tools that are new to me, like Python and SQL.

? I made a retroactive analysis based on empirical knowledge and feeling, without theoretical foundation. I would like to have a feedback about it.

Graphs:
+ I was able to prepare graphs in Excel.

- I do not know how to include all this information in only one graph without it get confuse.

Interpret data visualization:
+ I made my best, using complete sentences and explain better as possible.

? You will say me if they are accurate. I had to interpreted data from concepts that are new for me, like moving average.