Project 1 – Explore Weather Trends

- **Topic: "Are cities at similar latitudes have a similar temperature?"**

I am interested whether latitude would make the difference in temperature across the cities. In other words, the topic is "Are cities at similar latitudes have a similar temperature?"

- **Step1: Data Extraction**

First, SQL is used to extract data from the database. Since living in the UK a few years ago, I choose London (51°30'00"N) to be one of the cities to analyze. Then I choose the other 3 cities in different continents: Victoria in Canada (48°25′59″N), Qiqihar in China (47°14′22″N) and Brussels in Belgium (50°51′0″N).

The SQL queries are as follows:

```
/*Global data*/
select * from global_data

/*Choose city*/
select * from city_list
where 1=1
and city in ('London','Brussels','Victoria','Qiqihar')
```

Output 5 results — Download CSV

city	country
Brussels	Belgium
London	Canada
London	United Kingdom
Qiqihar	China
Victoria	Canada

Figure 1. Query Result

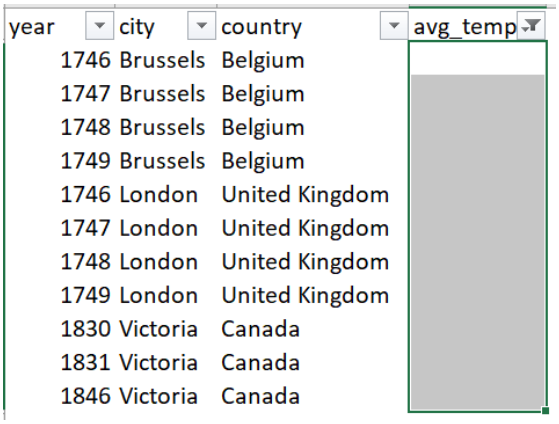
```
/*There are two London in different countries (Figure 1), so I add country='United Kingdom' when selecting the data of London. */

/*Extract city data*/
select * from city_data
where 1=1
and (city='Victoria')
or (city='Brussels')
or (city='Qiqihar')
or (city='London' and country='United Kingdom')
```

● Step2: Data Preparation

After I extract the data from the database and translate them from CSV format to Excel Workbook, I use Excel to do data preparation. There are 2 issues to deal with: missing value and time duration.

1) Missing value

Figure 2 indicates that there are some missing values in **avg_temp** field of different cities. Considering only 11 records are missing of all 922 records, which is a small percentage (nearly 1%), I decide to drop them.



year	city	country	avg_temp
1746	Brussels	Belgium	
1747	Brussels	Belgium	
1748	Brussels	Belgium	
1749	Brussels	Belgium	
1746	London	United Kingdom	
1747	London	United Kingdom	
1748	London	United Kingdom	
1749	London	United Kingdom	
1830	Victoria	Canada	
1831	Victoria	Canada	
1846	Victoria	Canada	

Figure 2. Missing values

2) Data from different cities have a different time duration

The other issue us that each city has a different time duration, which is listed in Table 1:

City/Global	Year Duration	# of missing values
Brussels	1743 - 2013	4 records missing: 1746, 1747, 1748, 1749
London	1743 - 2013	4 records missing: 1746, 1747, 1748, 1749
Qiqihar	1820 - 2013	No missing value
Victoria	1828 - 2013	3 records missing: 1830, 1831, 1846
Global data	1750 - 2015	No missing value

Table 1. Number of missing values in different cities

What I do in this step are: first, only use data (4 cities and global) from 1828 to 2013 to make the time duration be the same; second, because there are 3 missing values in Victoria from 1828 to 2013 (1830, 1831, 1846), I remove the 3-year data from the data of the 4 cities and global.

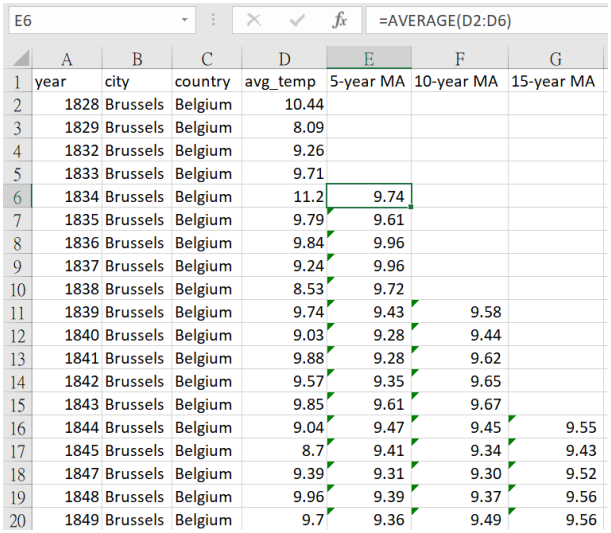
- **Step3: Data Exploration**

After the data preparation, the data population is confirmed to be:

Year: 1828 - 2013 (excluding 1830, 1831, 1846)

City: Brussels, London, Qiqihar, Victoria and Global data

In this step, the moving average line would be calculated. According to Moving Average Period and Which Is Best[1], there is no right moving average period. Therefore, I decide to calculate 5-year, 10-year and 15-year moving average lines to see which one works for the analysis. The tool used is the **AVERAGE** function in Excel in Figure 3.



E6 =AVERAGE(D2:D6)

	A	B	C	D	E	F	G
1	year	city	country	avg_temp	5-year MA	10-year MA	15-year MA
2	1828	Brussels	Belgium	10.44			
3	1829	Brussels	Belgium	8.09			
4	1832	Brussels	Belgium	9.26			
5	1833	Brussels	Belgium	9.71			
6	1834	Brussels	Belgium	11.2	9.74		
7	1835	Brussels	Belgium	9.79	9.61		
8	1836	Brussels	Belgium	9.84	9.96		
9	1837	Brussels	Belgium	9.24	9.96		
10	1838	Brussels	Belgium	8.53	9.72		
11	1839	Brussels	Belgium	9.74	9.43	9.58	
12	1840	Brussels	Belgium	9.03	9.28	9.44	
13	1841	Brussels	Belgium	9.88	9.28	9.62	
14	1842	Brussels	Belgium	9.57	9.35	9.65	
15	1843	Brussels	Belgium	9.85	9.61	9.67	
16	1844	Brussels	Belgium	9.04	9.47	9.45	9.55
17	1845	Brussels	Belgium	8.7	9.41	9.34	9.43
18	1847	Brussels	Belgium	9.39	9.31	9.30	9.52
19	1848	Brussels	Belgium	9.96	9.39	9.37	9.56
20	1849	Brussels	Belgium	9.7	9.36	9.49	9.56

Figure 3. Moving average calculation

Moreover, I choose using "**Data analysis**" add-in in Excel rather than inputting **CORREL** function to generate correlation coefficients between multiple variables. First, I merge the average temperature (**avg_temp**) of 4 cities and global data (Figure 4) and follow the steps in Figure 5 and Figure 6.

year	Brussels	London	Qiqihar	Victoria	Global
1828	10.44	10.1	2.41	6.83	8.17
1829	8.09	7.95	2.29	6.58	7.94
1832	9.26	9.24	1.67	3.25	7.45
1833	9.71	9.4	2.03	7.27	8.01
1834	11.2	10.59	2.65	6.81	8.15
1835	9.79	9.48	1.71	5.35	7.39
1836	9.84	9.04	2.15	6.52	7.7

Figure 4. The 4 cities and global data after merge

[1] Macroption, 2020. *Moving Average Period and Which Is Best*. Available at: <https://www.macroption.com/moving-average-period/>

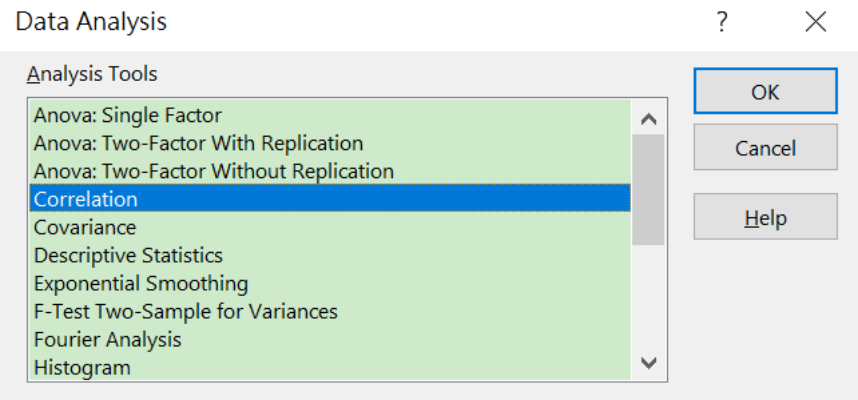


Figure 5. Data Analysis add-in in Excel

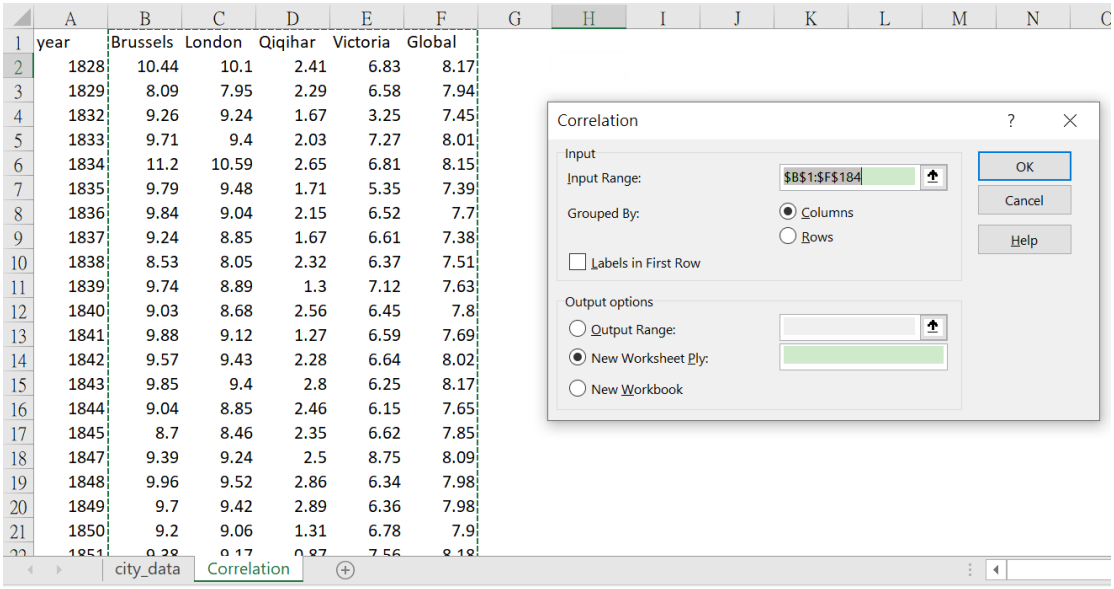


Figure 6. Data range selection

- **Step4: Data Visualization**

In this step, I combine the city data and the global data to import into Tableau Desktop. For global data, instead of being NULL, the value “Global” is filled into **city** and **country** field (Figure 7).

year	city	countr	avg_te	5-year M	10-year MA	15-year MA
1828	Global	Global	8.17			
1829	Global	Global	7.94			
1832	Global	Global	7.45			
1833	Global	Global	8.01			
1834	Global	Global	8.15		7.944	

Figure 7. The value “Global” filled into related fields

Then I import the merged data into Tableau Desktop (Figure 8).

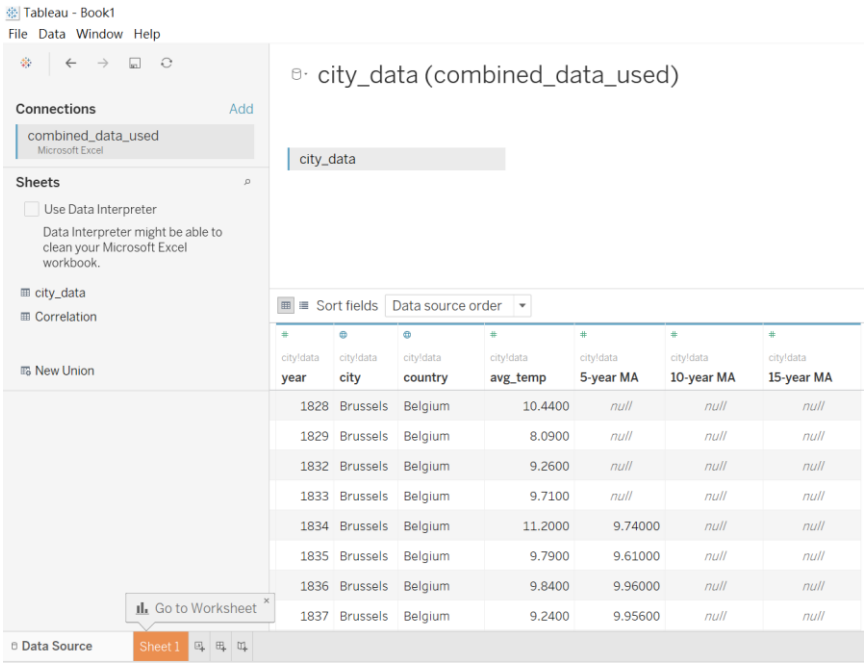


Figure 8. Data import in Tableau Desktop

In the data visualization step, the key consideration is 1) to clearly show the trend of the moving average temperature and 2) to clarify the difference between the 4 cities and the global data. Therefore, the line graph is the most suitable.

I drag **year** field into **Columns** and 3 types of moving average into **Rows**. Figure 9 illustrates that the 15-year moving average line is smoother than the others. Hence, I decide to use the 15-year moving average line to do data visualization. After I drag **15-year MA** field into **Rows** and put **city** into **color mark card**, the final view is finished. Figure 10 has **year** on X-axis and 15-year moving average values on Y-axis with different colors indicating cities and global data.

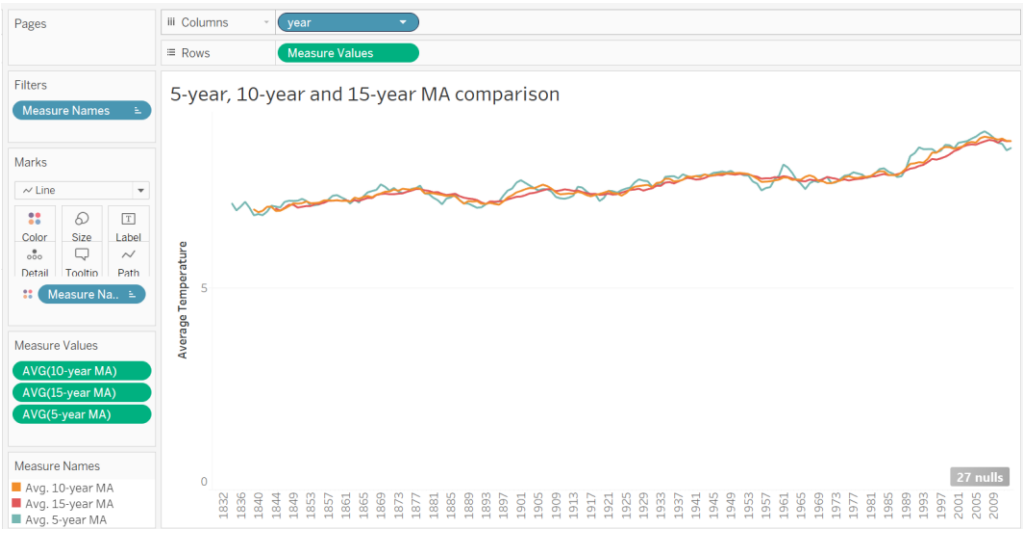


Figure 9. 3 types of moving average period comparison

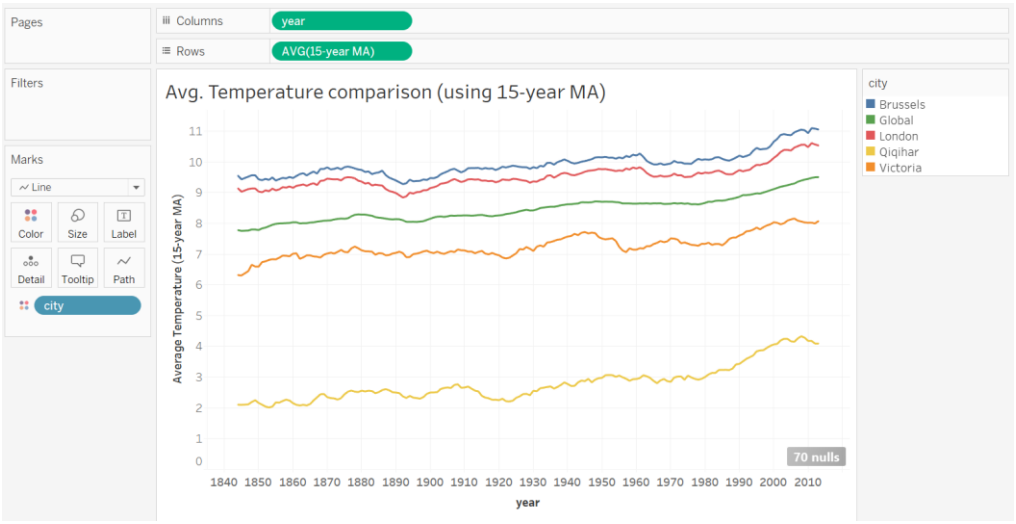


Figure 10. Final view

- **Findings**

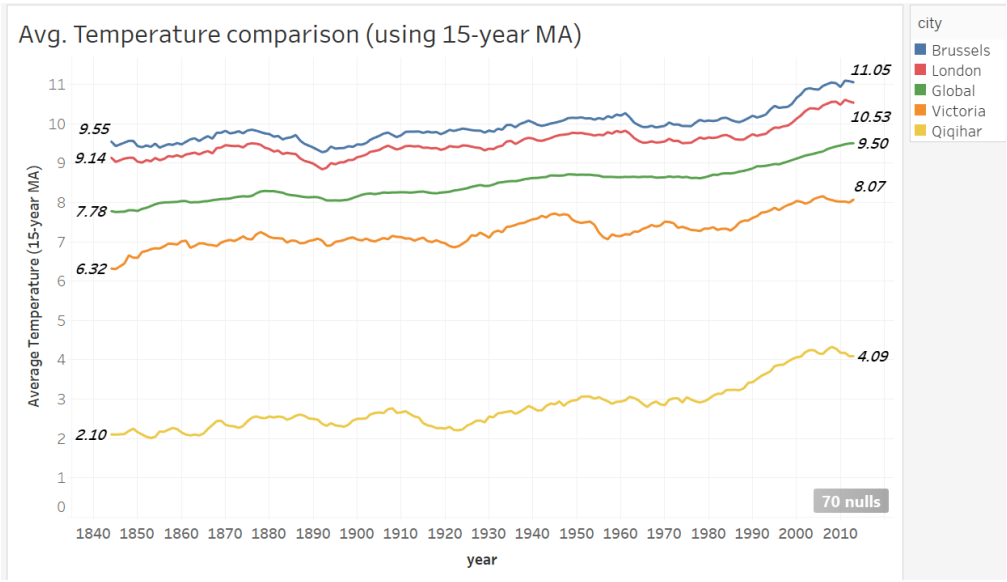


Figure 11. Moving average value of temperature comparison

	Brussels	London	Qiqihar	Victoria	Global
Brussels	1				
London	0.96	1			
Qiqihar	0.56	0.58	1		
Victoria	0.28	0.27	0.38	1	
Global	0.58	0.59	0.70	0.64	1

Table 2. Correlation coefficient matrix

According to the line graph (Figure 11) and the correlation coefficient matrix (Table 2), I have the following findings:

1) There are upward trends for global moving average temperatures and for which of each of the 4 cities. There is also an increase in all the 4 cities after 2000, showing that the climate changes dramatically.

2) The moving average temperatures of Brussels and London have been higher than the global over time. On the other hand, the moving average temperature of Qiqihar has been the lowest. It illustrates that latitude is not the only indicator that influences temperature. The temperature varies even if they have similar latitudes. In FlexBooks® 2.0[2], there is a similar result that the continental position might affect the climate.

3) Overall, the global moving average temperature has a positive correlation with all of the 4 cities. The moving average temperature of Qiqihar rises substantially among all the cities. In 2013, it increased by 94.39%. The rest of the cities also rose by different percentage: Brussels (+15.75%), London (+15.27%), Victoria (+27.76%). There was a 22.21% increase in the global average temperature.

4) There is a high correlation between the average temperature of Brussels and London, with a correlation coefficient of 0.96. Conversely, the correlation between London and Victoria is the lowest, with a coefficient of 0.27. However, it cannot be concluded that the increase in temperature of one city would have an impact on the change in temperature of the other.

[2] FlexBooks® 2.0, 2018. *Effect of Continental Position on Climate*. Available at: <https://flexbooks.ck12.org/cbook/ck-12-middle-school-earth-science-flexbook-2.0/section/12.4/primary/lesson/effect-of-continental-position-on-climate-ms-es>