

项目 1 背景介绍：分析当地和全球温度数据，并将自己所在城市的温度趋势与全球总体温度趋势进行对比。

一、使用 **SQL Workspace 从气温数据库中提取数据**，并把结果下载为 CSV 文件。

- 首先看一下 city_list, city_data, global_data 这三个表大概的表结构是什么样子
- 使用 select 语句发现 city_list 表有 city 和 country 两列，数据类型如下：

Input

HISTORY ▾MENU ▾

SCHEMA ↻

city_data ▾

city_list ▾

global_data ▾

1 SELECT *

2 FROM city_list

3 LIMIT 10;

Success!

EVALUATE

Output 10 results

Download CSV

city	country
Abidjan	Côte D'Ivoire
Abu Dhabi	United Arab Emirates

- 使用 select 语句发现 city_data 表有 year,city,counry,avg_temp 四列，数据类型如下：

SCHEMA ↻

city_data ▾

city_list ▾

global_data ▾

1 SELECT *

2 FROM city_data

3 LIMIT 10;

Success!

EVALUATE

Output 10 results

Download CSV

- 使用 select 语句发现 global_data 表有 year, avg_temp 四列，数据类型如下：

Input		HISTORY ▾	MENU ▾
SCHEMA	↻	<pre> 1 SELECT * 2 FROM global_data 3 LIMIT 10; </pre>	
city_data	▾		
city_list	▾		
global_data	▾		
		Success!	EVALUATE
Output		10 results	Download CSV
year	avg_temp		
1750	8.72		
1751	7.98		
1752	5.78		

2. 找出自己所在的城市温度

- 1) 我自己所在的城市是北京，所以在 city_list 表中用 select 语句查询北京 `SELECT * FROM city_data WHERE city like 'Bei%'` 但并没有发现北京。所以我决定把 city_list 表中 country 是 China 的都查出来看看，并且下载数据，文件名为 China_city_results.csv。

Input		HISTORY ▾	MENU ▾
SCHEMA	↻	<pre> 1 SELECT * 2 FROM city_data 3 WHERE country = 'China'; </pre>	
city_data	▾		
city_list	▾		
global_data	▾		
		Success!	EVALUATE
Output		6172 results	Download CSV
year	city	country	avg_temp
1829	Anshan	China	11.80
1830	Anshan	China	7.12

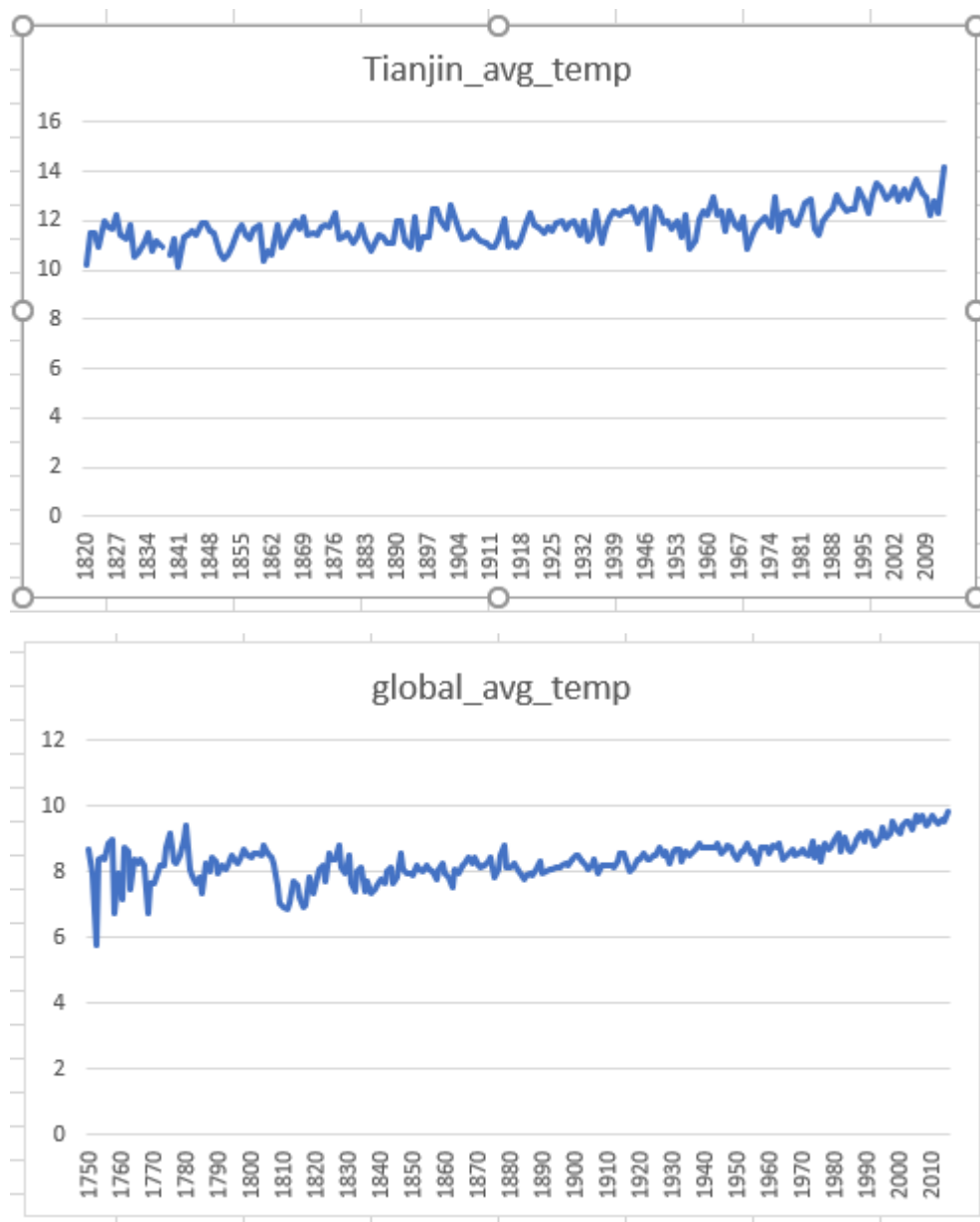
- 2) 通过浏览数据发现数据里不包含北京，因此我想看看数据里是否有离北京比较近的天津。

Input		HISTORY ▾	MENU ▾
SCHEMA	↻	<pre> 1 SELECT * 2 FROM city_list 3 WHERE city = 'Tianjin'; </pre>	
city_data	▾		
city_list	▾		
global_data	▾		
		Success!	EVALUATE
Output		1 results	Download CSV
city	country		
Tianjin	China		

发现有天津，因此到 city_data 表中把城市为天津的数据下载下来，命名为 Tianjin_results.csv

Input		HISTORY ▾	MENU ▾
SCHEMA	↻	<pre> 1 SELECT * 2 FROM city_data 3 WHERE city = 'Tianjin'; </pre>	
city_data	▾		
city_list	▾		
global_data	▾		
		Success!	EVALUATE
Output		194 results	Download CSV
year	city	country	avg_temp
1820	Tianjin	China	10.18
1821	Tianjin	China	11.51
1822	Tianjin	China	11.50

3. 把全球温度下载下来，并命名为 global_results.csv



2. 因移动平均值比平均值会使折线图平滑一些，我们选用温度移动平均值取代温度平均值，再做一次折线图，看看是否有改善。因这两张图没有特别明显的趋势循环，我们暂选每 10 年的移动平均值。具体做法如下

1) 在 Tianjin_results.csv 添加一列 10_year_avg_temp 表示 10 年移动平均温度，并计算数值。

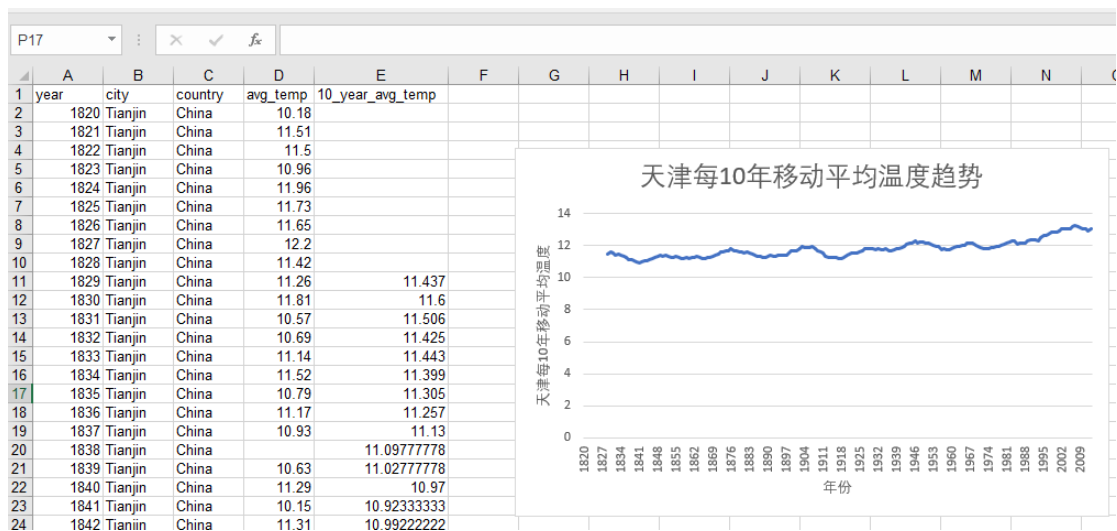
- 先计算一个 10 年的移动平均值

SUM		:	X	✓	<i>f_x</i>	=AVERAGE(D2:D11)
	A	B	C	D	E	F
1	year	city	country	avg temp	10_year_avg_temp	
2	1820	Tianjin	China	10.18		
3	1821	Tianjin	China	11.51		
4	1822	Tianjin	China	11.5		
5	1823	Tianjin	China	10.96		
6	1824	Tianjin	China	11.96		
7	1825	Tianjin	China	11.73		
8	1826	Tianjin	China	11.65		
9	1827	Tianjin	China	12.2		
10	1828	Tianjin	China	11.42		
11	1829	Tianjin	China	11.26	=AVERAGE(D2:D11)	
12	1830	Tianjin	China	11.81		
13	1831	Tianjin	China	10.57		

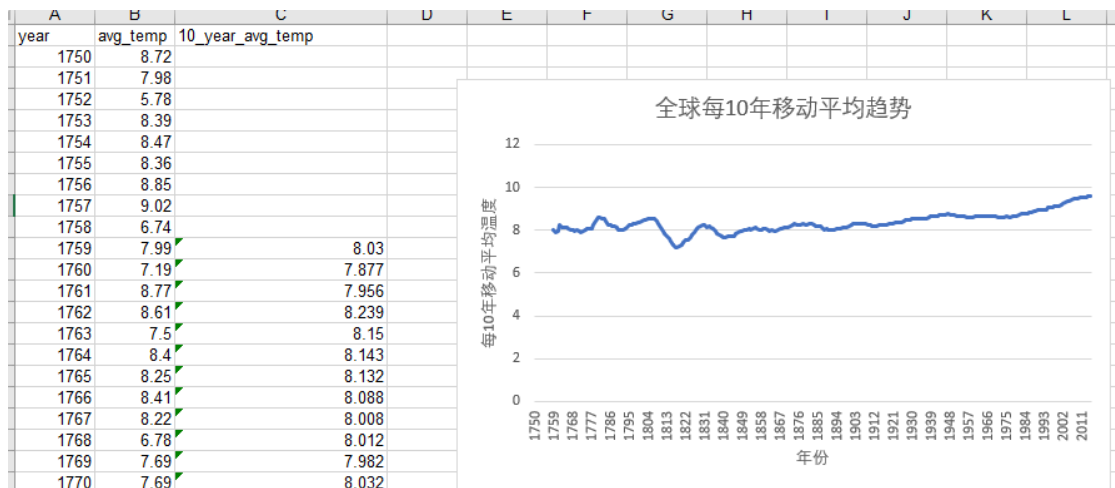
- 然后往下拉，得出全部的移动平均值。

剪贴板		常用		字体		
E11				=AVERAGE(D2:D11)		
	A	B	C	D	E	F
1	year	city	country	avg_temp	10_year_avg_temp	
2	1820	Tianjin	China	10.18		
3	1821	Tianjin	China	11.51		
4	1822	Tianjin	China	11.5		
5	1823	Tianjin	China	10.96		
6	1824	Tianjin	China	11.96		
7	1825	Tianjin	China	11.73		
8	1826	Tianjin	China	11.65		
9	1827	Tianjin	China	12.2		
10	1828	Tianjin	China	11.42		
11	1829	Tianjin	China	26	11.437	
12	1830	Tianjin	China	11.81	11.6	
13	1831	Tianjin	China	10.57	11.506	
14	1832	Tianjin	China	10.69	11.425	
15	1833	Tianjin	China	11.14	11.443	
16	1834	Tianjin	China	11.52	11.399	
17	1835	Tianjin	China	10.79	11.305	
18	1836	Tianjin	China	11.17	11.257	
19	1837	Tianjin	China	10.93	11.13	
20	1838	Tianjin	China		11.09777778	
21	1839	Tianjin	China	10.63	11.02777778	
22	1840	Tianjin	China	11.29	10.97	
23	1841	Tianjin	China	10.15	10.92333333	

- 做出 10 年移动平均值的折线图

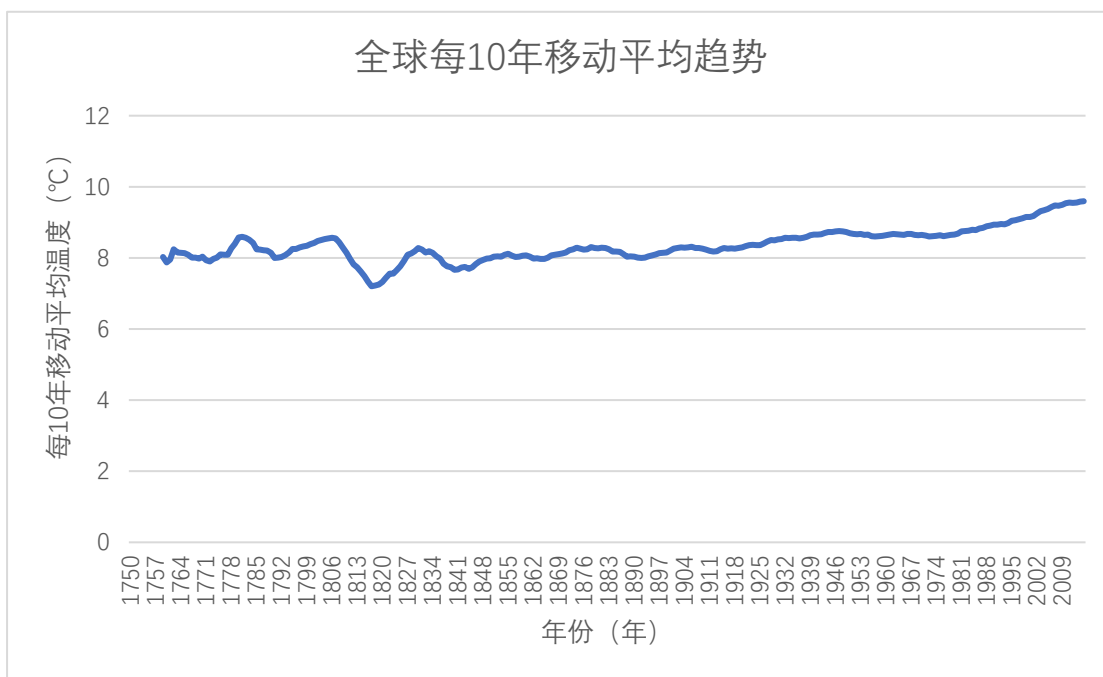
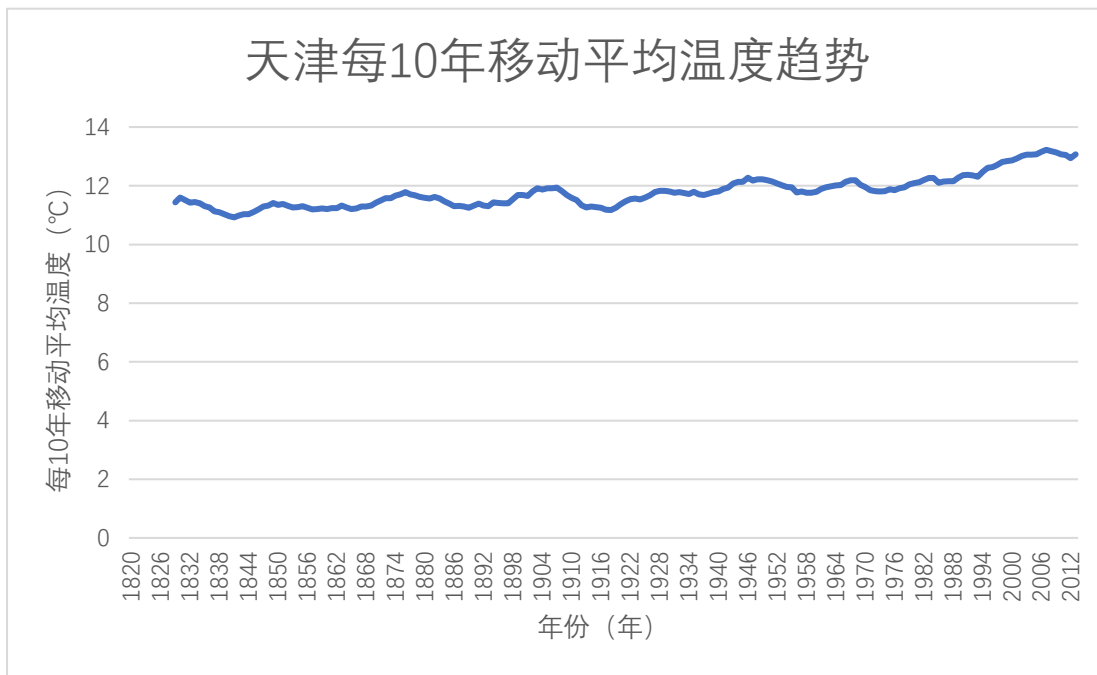


2) 在 global_results.csv 添加一列 10_year_avg_temp 表示 10 年移动平均温度，并计算数值，步骤与天津的例子是一样的。

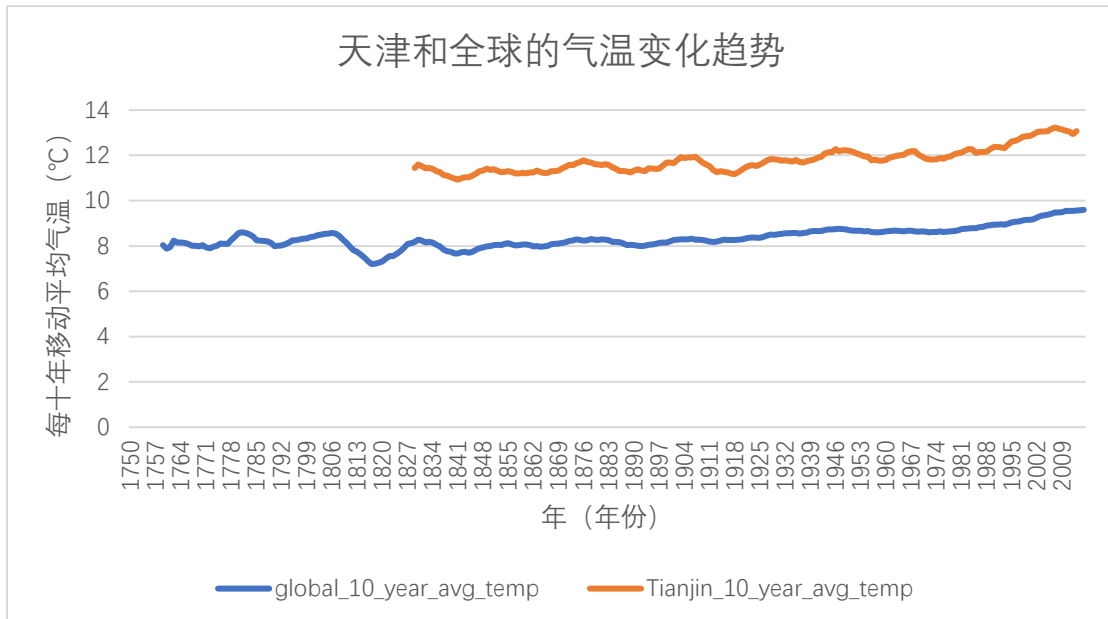


三、分析天津与全球的温度变化趋势

观察天津和全球的温度变化折线图：



把天津与全球的温度变化趋势放在一起，并得出一下结论，可以在文件 Tianjin_global.xlsx 看到具体数据：



1. **相似性**：

- 1) 天津和全球的气温趋势都是逐渐上升的。
- 2) 天津和全球气温差这么多年是基本一致的。
- 3) 天津和全球气温的相似度很高，高达 0.96。用的 excel 数据分析里的相关系数做的。

2. **差异性**：

- 1) 天津的平均气温比全球的平均要高，并且是天津长期比全球的温度高。

天津的每十年移动平均气温大概在 10 到 14 摄氏度，而全球的每十年移动平均气温大概在 8 到 10 摄氏度。对每十年移动平均气温求平均值，天津的大概是 11.79°C，全球大概是 8.35°C。

- 2) 全球气温离散程度相对天津较小。天津温度的标准差为 0.53，全球温度的标准差 0.45。用的是 excel 中 stdev 求得。