

探索未来气候发展趋势

优达数据分析学员：曾艳

气候发展趋势探索？！听起来是不是很高大上呢？没错，对于一名工科小白来说，天文气候领域是神秘而与小曾无关。没想到我也会有这么一次机会，写着跟气候相关的数据分析项目。废话少说，本文将主要利用 SQL 基础语句来查询课程提供的数据库，用表格工具让数据可视化，描述全球气温走向和最接近小曾所在城市气温走向之间的相似性与差异。

本文主要分三个方面来描述，分别为**数据提取**、**数据分析**、**提问总结**。

一、数据提取

使用 **SQL Workspace** 从气温数据库中提取数据。数据库中有三个表：

- city_list - 这个表包含数据库中的城市和国家列表。
- city_data - 这个表包含每个城市每年的平均气温 (°C)。
- global_data - 这个表包含每年的全球平均气温 (°C)。

首先，city_list 只有两个字段，city 和 country。不用多想，马上查询表里中

国都有哪些城市。按照已有经验，China 为大写无疑。

```
HISTORY ▼ MENU ▼  
1 SELECT *  
2 FROM city_list  
3 WHERE country='China';
```

其次，因本人居住在广东珠海，从 34 results 中，查找到了与珠海较为接近

的城市 **Guangzhou**，于是，一切就从 Guangzhou 入手。其实，Foshan 也距离珠海很近，但相比之下我更喜欢广州。

Output 34 results		Download CSV
Fuzhou	China	
Guangzhou	China	
Gulyang	China	

对了，还要提取全球气温数据，并且把以上查询结果都导出 CSV 文件。鉴于以防万一后面要用到别的数据，我又把另外两个表全部内容也下载了（city_list，city_data）

Input		HISTORY	MENU
SCHEMA			
city_data			
1 SELECT *			
2 FROM global_data;			

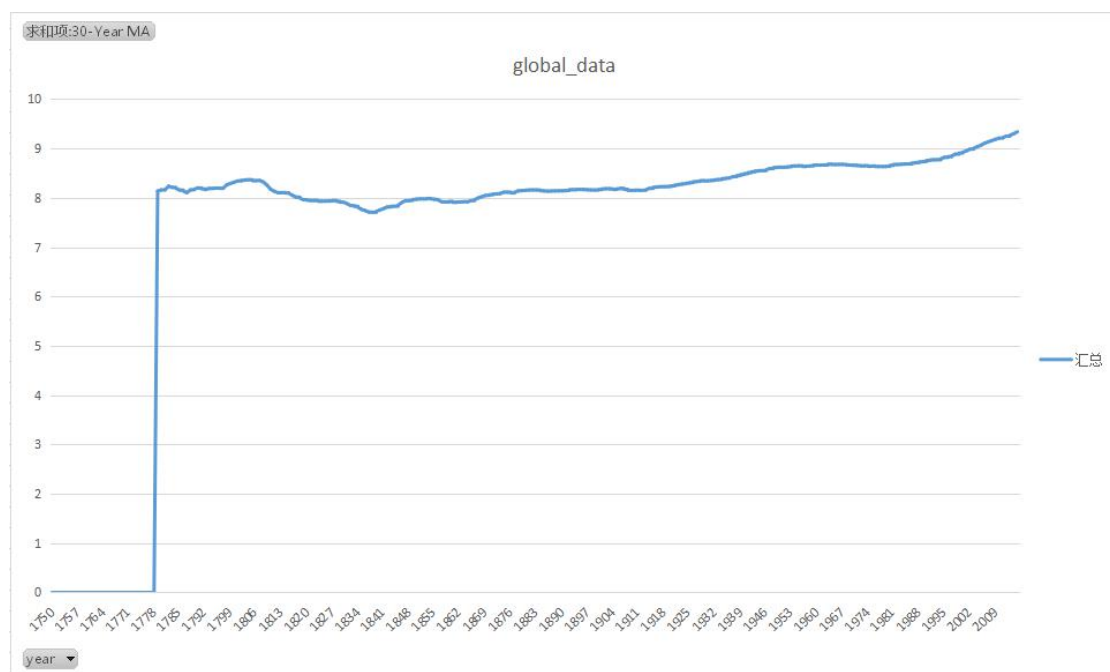
二、数据分析

用 **WPS-表格** 打开 **city_data_Guangzhou.csv**、**global_data.csv**，设置“冻结窗格”和“自动筛选”功能。观察发现两份数据均没有空值，故不需要进行缺失值处理。“year”字段也是连续年份，没有出现年份中断。

然后，问题来了。**移动平均值的时间间隔怎么选才合适？**

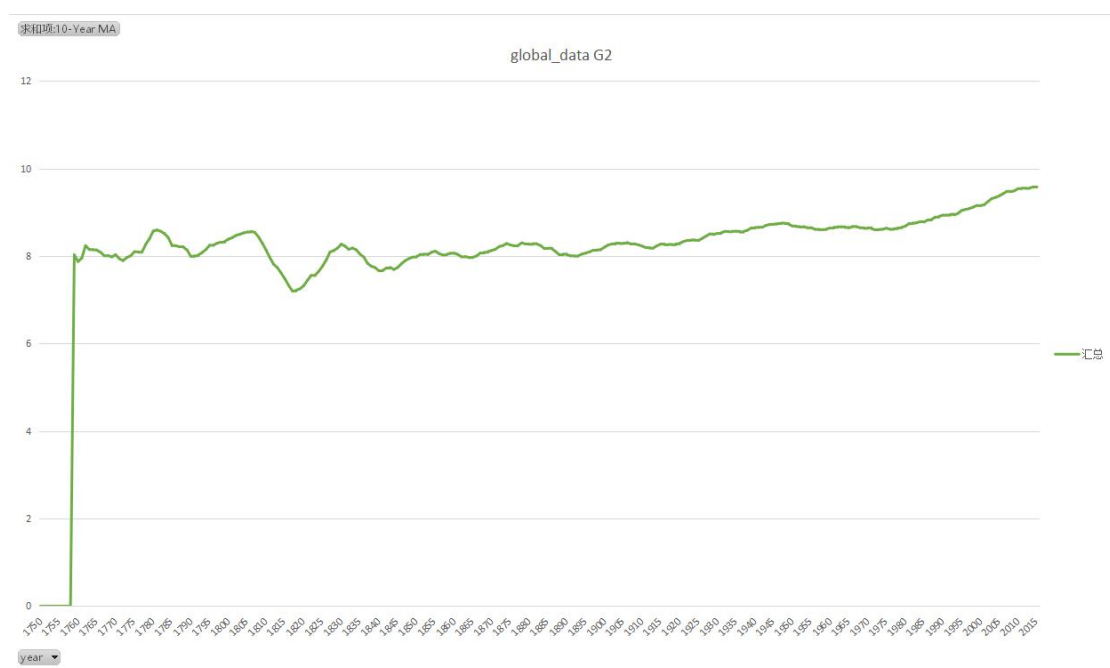
首先 我把间隔设置为 30 年。在 **global_data.csv** 表中创建一个名为“30-Year MA”的列，计算**移动平均值**，通过数据透视表设置 Y 轴、值，插入“堆积折线图”G1；

year	avg tem	30-Year MA
1777	8.26	
1778	8.54	
1779	8.98	=AVERAGE(B2:B31)



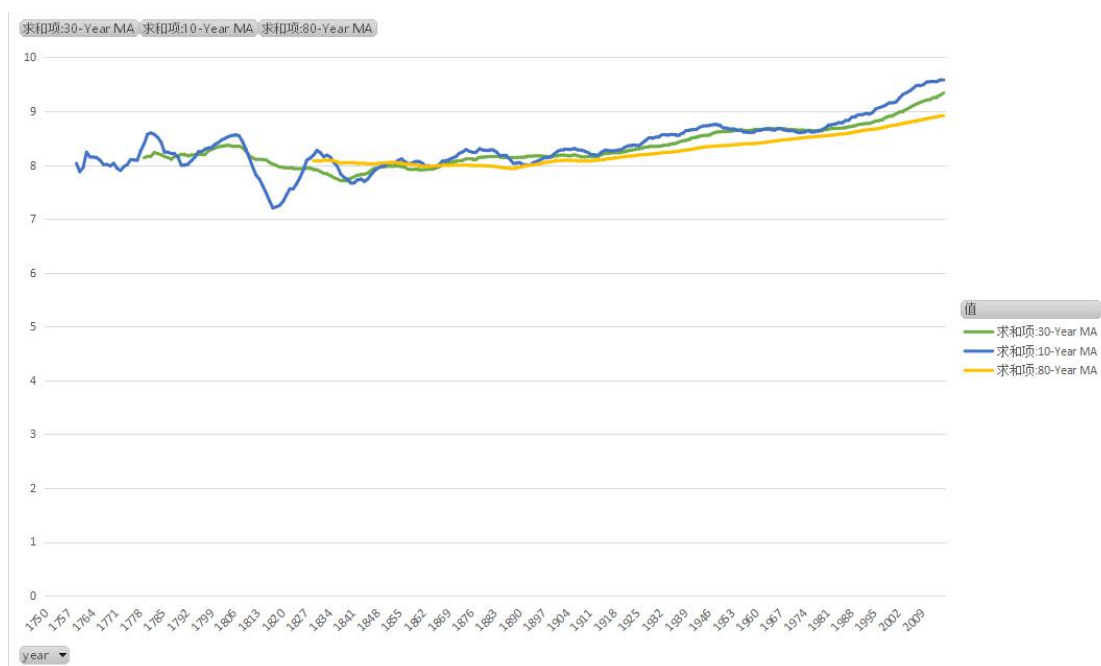
可以看出，**曲线较平缓，有上升趋势**。但，这看不出什么...

我再次尝试把间隔设置为 10 年，粒度更细，看看是什么效果。通过用同样方式，得出 G2。图表同等放大 200%对比，**这次的曲线波动稍微大了一点，有上升趋势**。**但加大粒度后区别不大。**



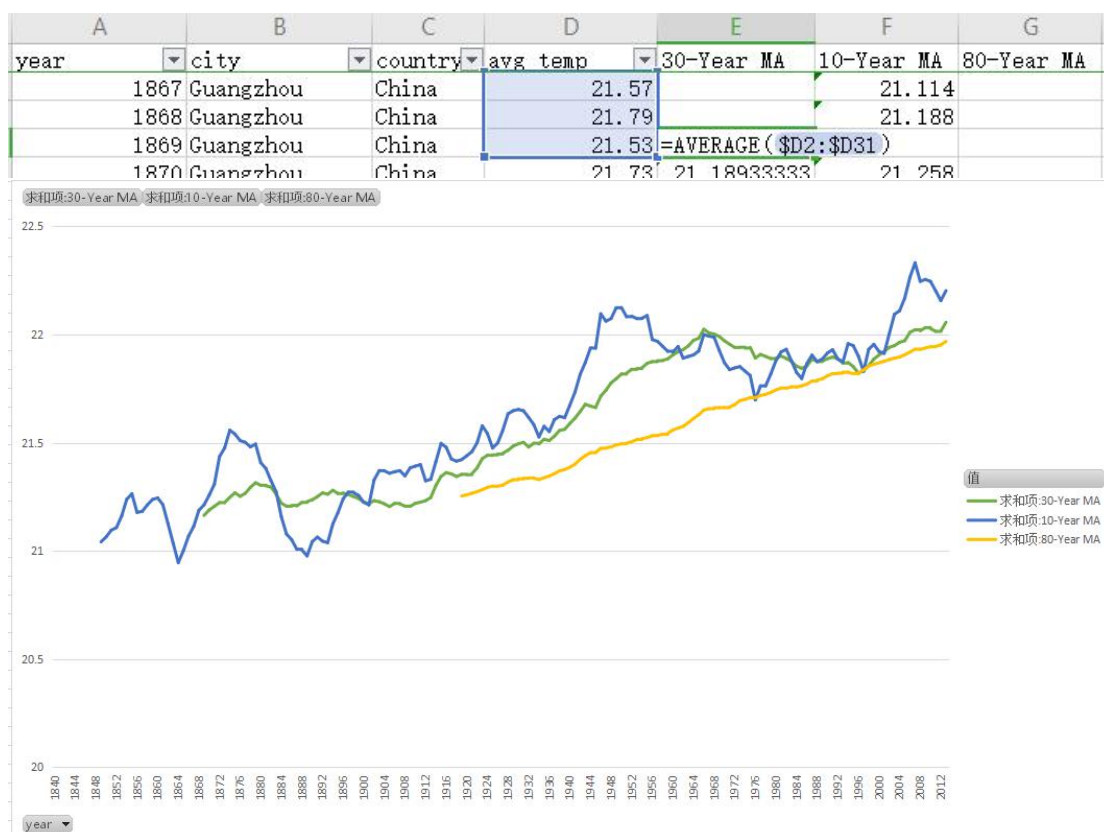
于是，第三次把间隔设置为 80 年，以同样方式操作，加上前两次的值，得出 G3。

80-Year MA (黄色线) 坡度稍有增加, 呈上升趋势。



用同样方式操作 city_data_Guangzhou.csv 表中数据, 分别用 30、10、80 间隔

计算其移动平均值, 创建折线图 G4。



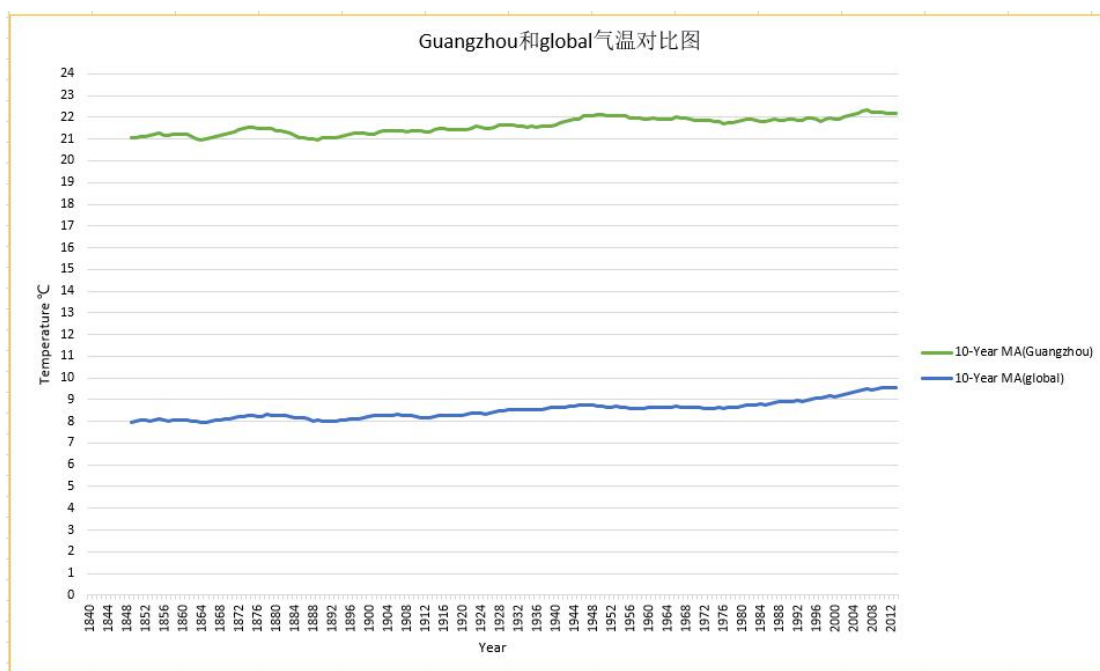
根据前三次经验, 最后选择粒度小的间隔 10 年作为 global_data 和

city_data_Guangzhou 对比，是否两者近百年来气温变化趋势是一致的？

此时需创建一份新表 **union_data.xlsx**，整合两方数据，保留 year，avg_temp，重新计算 10-Year MA 字段。因 city_data_Guangzhou 统计年份 1840-2013，而 global_data 为 1750-2015，为方便对比，统计年份为 1840-2013。

重新修改 SQL 语句，导出 CSV 文件，并创建折线图 G5。

```
1 SELECT city_data.year,city_data.avg_temp as  
   guangzhou_temp,global_data.avg_temp as global_temp  
2 FROM city_data,global_data  
3 WHERE city_data.year = global_data.year  
4 AND city_data.city = 'Guangzhou'
```



根据 G5 可看出，两者波动平缓，均有缓慢的上升趋势，坡度接近，Guangzhou 线图的波动相对明显一点点。

三、提问总结

经过一番折腾，最终发现近百年来，全球和广州的气温均在缓慢变暖。

继续思考：

- 从哪个年份开始，两者的气温变化明显加剧？

-根据 G5 分析看出，从 2000 年左右开始，global 和 Guangzhou 曲线坡度增加。

- 加剧变化的年份开始到近年，哪一年广州平均气温最热？

-根据 SQL 语句查询，从 2000 年开始到 2013 年，广州平均气温 2013 年最热，高达 22.93 摄氏度。

The screenshot shows a SQL query interface. On the left, a schema for 'city_data' is listed with columns: year, city, country, and avg_temp. The SQL query in the center is:

```
1 SELECT year, avg_temp
2 FROM city_data
3 WHERE city = 'Guangzhou' AND year >= 2000
4 ORDER BY avg_temp DESC;
```

Below the query, a green 'Success!' message and a blue 'EVALUATE' button are visible. The 'Output' section shows 14 results. The first two rows are highlighted with a red box:

year	avg_temp
2013	22.93
2007	22.54

- 广州平均气温最热的那年，是否也为全球气温最热的一年？

-根据 SQL 语句查询，2013 年并非为全球气温最热的一年。全球平均最高气温最高为 2007 年，高达 9.73 摄氏度。

SCHEMA

city_data

year

city

country

avg_temp

1 SELECT year,avg_temp

2 FROM global_data

3 WHERE year >= 2000 AND year <= 2013

4 ORDER BY avg_temp DESC;

Success!

EVALUATE

Output 14 results

Download CSV

year	avg_temp
2007	9.73
2005	9.70
2010	9.70
2013	9.61