

# 项目一：探索未来气候发展趋势

## Udacity 数据分析入门

### 概览

这是 Udacity 数据分析（入门）纳米学位的第一个项目。在这个项目中，我将通过 SQL 提取数据并运用 Excel 分析本地和全球的气温数据，并比较我的居住地的气温走向与全球气温走向。

### 提取数据

Udacity 数据分析（入门）的课程页面提供了存有此次项目所研究的原始数据的数据库。`global_data.csv` 包含了 1750 年至 2015 年间每年的全球平均气温；`city_data.csv` 包含了每个城市每年的平均气温，然而各个城市的气温数据记录开始和结束的年份并不一致；`city_list.csv` 包含数据库中的城市和国家列表，通过这个文件可以快速查看哪些城市的气温信息可以再数据库中找到。首先通过 SQL 查询 `city_data.csv` 表格得到我所在的城市深圳的气温数据。

Input		HISTORY ▾	MENU ▾
SCHEMA	↻	<pre>1 SELECT * 2 FROM city_data 3 WHERE city = 'Shenzhen';</pre>	
city_data	▾		
city_list	▾		
global_data	▾		
		Success!	EVALUATE
Output		174 results	Download CSV
year	city	country	avg_temp
1840	Shenzhen	Hong Kong	23.71
1841	Shenzhen	Hong Kong	20.76
1842	Shenzhen	Hong Kong	20.96

通过下拉快速查看数据可以发现两点：1. 深圳的这一气温数据开始记录的年份是 1840 年，记录结束的年份为 2013 年，因此为能够平行比较深圳和全球的平均气

温，对于全球的气温数据我也应当截取 1840 年至 2013 年部分；2. 深圳的气温数据中 `country` 一列信息存在错误，Hong Kong 应为 China。但是由于城市所属国家的信息错误并不影响对城市平均气温的探究，我将忽略这一问题，并且只下载深圳的年份和温度数据。

Input

HISTORY ▾

MENU ▾

SCHEMA ↻

city\_data ▾

city\_list ▾

global\_data ▾

1 `SELECT year, avg_temp`

2 `FROM city_data`

3 `WHERE city = 'Shenzhen';`

Success!

EVALUATE

Output

174 results

Download CSV

year	avg_temp
1840	23.71
1841	20.76

接下来通过 SQL 查询得到 `global_data.csv` 的从 1840 年至 2013 年的气温数据，并下载 csv 文件。

Input

HISTORY ▾

MENU ▾

SCHEMA ↻

city\_data ▾

city\_list ▾

global\_data ▾

1 `SELECT *`

2 `FROM global_data`

3 `WHERE year BETWEEN 1840 AND 2013;`

Success!

EVALUATE

Output

174 results

Download CSV

year	avg_temp
1840	7.80
1841	7.69
1842	8.02

# 计算移动平均值

得到数据后，我可以开始用 Excel 分别对全球和深圳的平均气温数据计算移动平均值。选择移动平均值的原因是“移动平均值用于将数据线性化，以便更容易观察长期趋势，也不会因日常波动而错乱。”<sup>1</sup>

在此次探究中，我选择计算的是五年的移动平均值，原因是在折线图中五年的移动平均值相比每年平均气温更加平滑，但同时也不会抹除过多的气温波动。

首先将先前下载的两个 csv 文件用谷歌表格打开，然后将其复制到同一个表格中。接下来利用电子表格的公式计算功能得到已经气温数据的五年移动平均值，具体计算方法是将计算年份和其前四年（即总计共五年）的平均温度相加再除以五。下图截出了在电子表格中使用的公式和得到的前五年的五年移动平均值作为示例。

fx

=AVERAGE(B2:B6)

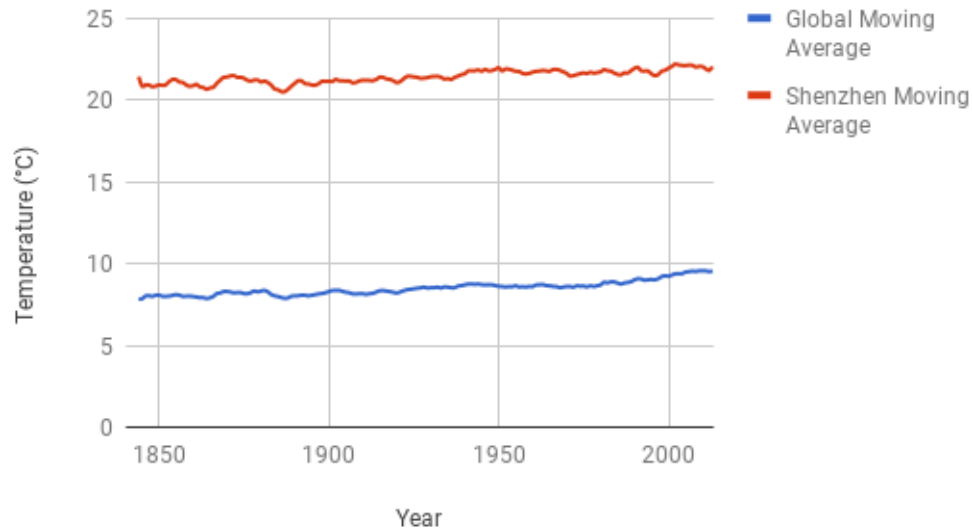
	A	B	C	D	E	F
1	year	globalavg_temp	Global Moving Average	city_avg_temp	Shenzhen Moving Average	
2	1840	7.8		23.71		
3	1841	7.69		20.76		
4	1842	8.02		20.96		
5	1843	8.17		21.05		
6	1844	7.65	7.866	20.66	21.428	
7	1845	7.85	7.876	20.66	20.818	
8	1846	8.55	8.048	21.28	20.922	
9	1847	8.09	8.062	20.91	20.912	
10	1848	7.98	8.024	20.55	20.812	

# 数据可视化

通过计算得到五年移动平均值的数据后，就可以进行数据可视化。这里我将继续使用谷歌表格对两组移动平均值分别绘制折线图。

<sup>1</sup> "SQL 和移动平均值." 优达学城. Accessed February 13, 2018.  
<https://classroom.udacity.com/nanodegrees/nd002-cn-basic/parts/20a82dbe-5a80-4829-b9c8-198f518ed20e/modules/188c878c-5365-4bf3-9fa8-08cf57336fc4/lessons/eb0e4ed3-e04a-4328-aece-da8cf6fc2607/concepts/0f213a2e-ab83-42a5-be47-fd52a193c003>.

## Global and Shenzhen 5 Year Moving Average Temperature



从图中可以看到，长期以来深圳的气温移动平均值基本都在 21 到 22 摄氏度左右，远远高于全球气温移动平均值的 10 度以下范围，其原因是深圳处于亚热带，常年气候温暖，而全球平均温度还要算入极寒地区。

从两组数据在图中的波动情况可以看出，深圳气温的移动平均值没有全球气温的移动平均值稳定，波动相比全球平均值较多。猜测原因可能是全球气温平均值算入了大量城市的平均气温，样本量较大，因此更加稳定。

从随时间变化的趋势来看，深圳的气温移动平均值并没有显著的变化，即不存在较为明显的上升或下降趋势。然而如果分段查看图表，可以看出 1950 年到 2000 年间总体的移动平均值相比 1850 到 1990 年间还是要稍高一些，

全球气温移动平均值随时间变化的趋势要更加明显，特别是在 1950 年后气温方面开始有明显的上升趋势。如果只看折线图中数据开端的 1844 年到数据结束的 2013 年，气温移动平均值的差甚至几乎达到了 2 摄氏度左右。因此，图中蓝色的折线较为直观地展示出了全球变暖的趋势。