Udacity course

Importing packages to read and analyze the dataset

In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [30]:
```
result=pd.read_csv('/Users/amit/Desktop/udacity/results.csv')
```

In [31]:
```
result.head()
```

Out[31]:

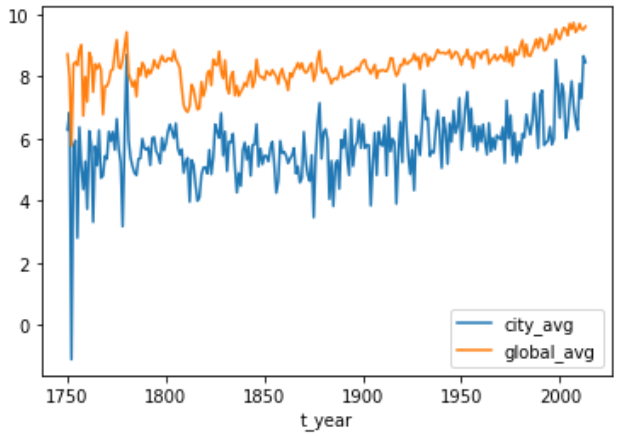
	t_year	city_avg	global_avg
0	1750	6.29	8.72
1	1751	6.84	7.98
2	1752	-1.10	5.78
3	1753	5.76	8.39
4	1754	5.94	8.47

In [32]:
```
result.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 264 entries, 0 to 263
Data columns (total 3 columns):
t_year         264 non-null int64
city_avg       264 non-null float64
global_avg     264 non-null float64
dtypes: float64(2), int64(1)
memory usage: 6.3 KB
```

In [35]:
```
result.plot(x="t_year", y=["city_avg", "global_avg"])
```

Out[35]: `<matplotlib.axes._subplots.AxesSubplot at 0x115e34160>`



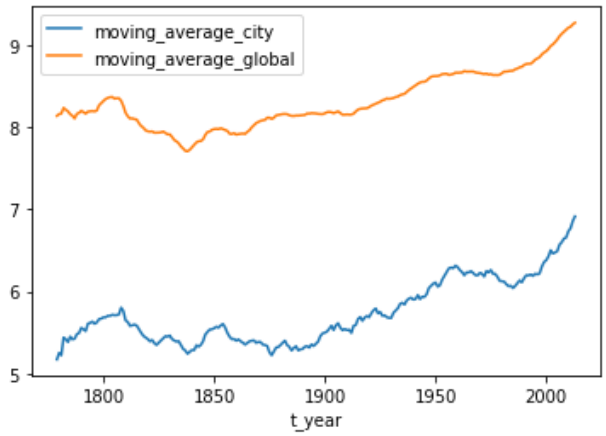
In order to create a line chart with moving average we need to transform the city and global column's moving average

Here I have taken moving average of seven years

In [36]:
```
result['moving_average_city'] = result.city_avg.rolling(window=30).mean()
result['moving_average_global'] = result.global_avg.rolling(window=30).mean()
```

In [39]:
```
result.plot(x="t_year", y=['moving_average_city', 'moving_average_global'])
```

Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x115e8a518>



* Make observations

Answer to question's of first project

1. Yes my city is on an average cooler than global average and the difference is consistent over years

2. It seems that global temperature has correlation with the city temperature i live in

3. The global and city temperature has increased from average of 8 to 9.5 and 5 to 7 respectively (world is become hotter), the trend for city data is not consistent but for the global data it is consistent with an upward trend

4. The global average is average of whole world and that is why the curve is smoother than city average

Outline or Steps taken

Following is the SQL query used to extract dataset

select city_data.year, city_data.avg_temp as city_temp, global_data.avg_temp as global_temp from city_data,global_data where city_data.year=global_data.year and city_data.city='Toronto';

I used python for analysis and sql data extraction

I used rolling mean function from python initially for 5 years , 10 years and then 30 years. 30 years rolling mean showed smoothest curve with minimum distortion of actual trend

Following were my key considerations:

- Trend should reflect variations of the temperature at 10 to 15 percent duration of total range (264 years)
- Lot of changes on shorter moving average durations show undue amount of changes making it difficult to understand the overall trend and that is why I chose 30 years rolling mean
- The imporatant goal was to compare temperature of Toronto with Global average