

Udacity-Project--Exploring Weather Trends

Step 1

Extracting data from the provided Database

I want to know how many Belgium cities are in the list

```
SELECT * FROM city_list WHERE country like 'Belgium';
```

Only Brussels is in the cities list.

Step 2

I renamed columns for joining

```
ALTER TABLE global_data RENAME COLUMN avg_temp to global_avg_temp;
```

```
ALTER TABLE city_data RENAME COLUMN avg_temp to city_avg_temp;
```

Step 3

-- Downloaded the joined tables

```
SELECT global_data.year, global_data.global_avg_temp, city_avg_temp FROM global_data INNER JOIN city_data ON  
global_data.year=city_data.year WHERE city like 'Brussels';
```

-- I saved as weather_project.csv

In [21]:

```
# Now I use python  
# Importing libraries  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
# Import the dataset  
df = pd.read_csv("weather_project.csv")
```

In [22]:

```
# look the head of the dataframe  
  
df.head(10)
```

Out[22]:

	year	global_avg_temp	city_avg_temp
0	1750	8.72	10.73
1	1751	7.98	10.52
2	1752	5.78	6.55
3	1753	8.39	9.86
4	1754	8.47	9.54
5	1755	8.36	9.30
6	1756	8.85	9.77
7	1757	9.02	9.61
8	1758	6.74	9.31
9	1759	7.00	10.06

9	1739	7.99	10.20
	year	global_avg_temp	city_avg_temp

In [23]:

```
# look the tail of the dataframe
df.tail(10)
```

Out[23]:

	year	global_avg_temp	city_avg_temp
254	2004	9.32	10.94
255	2005	9.70	11.16
256	2006	9.53	11.48
257	2007	9.73	11.52
258	2008	9.43	10.90
259	2009	9.51	11.00
260	2010	9.70	9.74
261	2011	9.52	11.69
262	2012	9.51	10.68
263	2013	9.61	10.36

In [24]:

```
#To learn the shape of the data
df.shape
```

Out[24]:

```
(264, 3)
```

In [25]:

```
#Find description of the dataset
df.describe()
```

Out[25]:

	year	global_avg_temp	city_avg_temp
count	264.000000	264.000000	264.000000
mean	1881.500000	8.359394	9.876932
std	76.354437	0.575184	0.733044
min	1750.000000	5.780000	6.550000
25%	1815.750000	8.077500	9.435000
50%	1881.500000	8.365000	9.880000
75%	1947.250000	8.700000	10.332500
max	2013.000000	9.730000	11.690000

In [26]:

```
#Defined Calculating function
def dfMean(windowRolling, df_in):
    df_out = df_in.rolling(window = windowRolling, center=False, on = "year").mean().dropna()
    return df_out
```

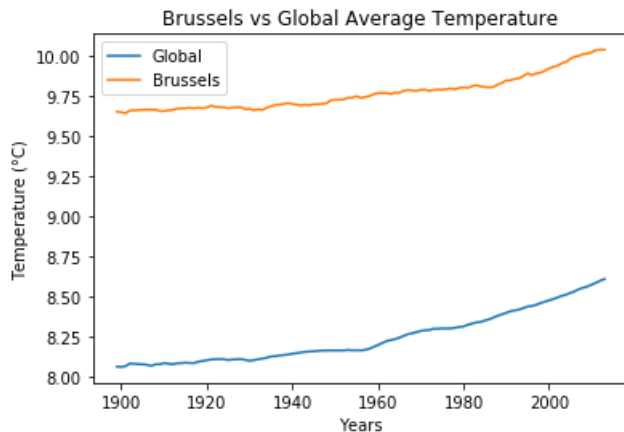
In [27]:

```
III [2/].
```

```
df1 = 150  
df_movingAverage = dfMean(df1, df)
```

```
In [28]:
```

```
# Drawing graph in matplotlib library  
  
plt.plot(df_movingAverage['year'], df_movingAverage['global_avg_temp'], label='Global')  
plt.plot(df_movingAverage['year'], df_movingAverage['city_avg_temp'], label='Brussels')  
plt.legend()  
plt.xlabel("Years")  
plt.ylabel("Temperature (°C)")  
plt.title("Brussels vs Global Average Temperature".format(df1))  
plt.show()
```



Observations

I checked in general the dataset and look the tables, the head and the tail and its description, it seems that always Brussels was in higher average temperature than the global. If we also look inside the last 10 years say from 2004 to 2013 the global average temperature was not less than 9.32 as for Brussels was always higher than the global and not less than 9.74.

References

I used <https://miktex.org/howto/install-miktex> with pandoc to convert the ipynb jupyter notebook to PDF.