Exploring weather trends project

City selected for the analysis: Barcelona, Spain

SQL Query used to download data for Barcelona:

```
SELECT year, city, avg_temp as average_temp
FROM city_data
WHERE city='Barcelona'
AND country='Spain'
AND avg_temp is not null
ORDER BY year ASC
```

After a first attempt selecting only **city='Barcelona'** I was getting duplicates up from a certain year. I went back to the data and realized I was also selecting data for Barcelona(Venezuela), therefoe I included the additional criteria by country to ensure I wasn't getting wrong data

Query used to download global data:

```
SELECT *
FROM global_data
WHERE avg_temp is not null
```

Both data selections coming from SQL already have its null values removed

As a first step I import the python libraries that we will need for this exercise

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
```

After importing the libraries I load the csv's and assign them to two different dataframes

In [3]:
```
barcelona = pd.read_csv('results_barcelona.csv')
barcelona.head()
```

Out[3]:

	year	city	average_temp
0	1743	Barcelona	13.81
1	1744	Barcelona	16.98
2	1745	Barcelona	10.78
3	1750	Barcelona	16.52
4	1751	Barcelona	16.78

In [4]:
```
overall = pd.read_csv('results_overall.csv')
overall.head()
```

Out[4]:

	year	average_temp
0	1750	8.7
1	1751	8.0
2	1752	5.8
3	1753	8.4
4	1754	8.5

After importing both csv's separately I proceed to join them using the merge method in Pandas

In [5]:
```
barcelona_world = barcelona.merge(overall,how='left', on='year', suffixes=['_barcelona','_overall'])
barcelona_world.head(10)
```

Out[5]:

	year	city	average_temp_barcelona	average_temp_overall
0	1743	Barcelona	13.81	NaN
1	1744	Barcelona	16.98	NaN
2	1745	Barcelona	10.78	NaN
3	1750	Barcelona	16.52	8.7
4	1751	Barcelona	16.78	8.0
5	1752	Barcelona	13.09	5.8
6	1753	Barcelona	16.17	8.4
7	1754	Barcelona	16.09	8.5
8	1755	Barcelona	15.80	8.4
9	1756	Barcelona	16.11	8.9

Using the describe function to check for inconsistencies that may appear in the data, such as outlier values

In [6]:
```
barcelona_world.describe()
```

Out[6]:

	year	average_temp_barcelona	average_temp_overall
count	267.000000	267.000000	264.000000
mean	1879.955056	16.116105	8.364015
std	77.298695	0.691412	0.572195
min	1743.000000	10.780000	5.800000
25%	1813.500000	15.830000	8.100000
50%	1880.000000	16.090000	8.400000
75%	1946.500000	16.470000	8.700000
max	2013.000000	17.900000	9.700000

Also, setting up the year as index so it will be easier to plot. Removing not useful columns for the analysis such as the 'city' column

In [7]:
```
barcelona_world.set_index('year',inplace=True)
```

In [8]:
```
barcelona_world.drop('city', axis='columns', inplace=True)
barcelona_world.head()
```

Out[8]:

year	average_temp_barcelona	average_temp_overall
1743	13.81	NaN
1744	16.98	NaN
1745	10.78	NaN
1750	16.52	8.7
1751	16.78	8.0

I have also decided to remove any null values that could be present in any of both columns, so I ensure that I am always comparing year vs year with valid data in both columns

In [9]:
```
barcelona_world.dropna(how='any', inplace=True)
barcelona_world.head()
```

Out[9]:

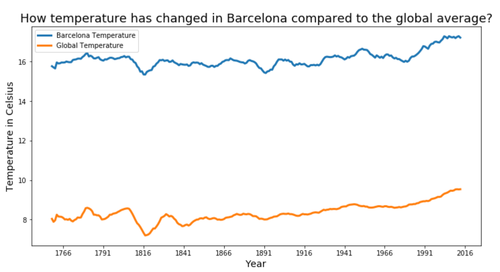
year	average_temp_barcelona	average_temp_overall
1750	16.52	8.7
1751	16.78	8.0
1752	13.09	5.8
1753	16.17	8.4
1754	16.09	8.5

Finally, before plotting I create two new columns consisting on a 10-year moving average using the pandas rolling method. I will end up plotting only the moving averages columns

In [10]:
```
barcelona_world['moving_average_barcelona'] = barcelona_world.rolling(window=10)['average_temp_barcelona'].mean()
barcelona_world['moving_average_overall'] = barcelona_world.rolling(window=10)['average_temp_overall'].mean()
```

In [11]:
```
barcelona_world[['moving_average_barcelona','moving_average_overall']].plot(figsize=(12,6), linewidth=3)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Temperature in Celsius', fontsize=14)
plt.legend(['Barcelona Temperature', 'Global Temperature'], loc='best')
plt.title('How temperature has changed in Barcelona compared to the global average?', fontsize=18)
plt.gca().xaxis.set_major_locator(mdates.DayLocator(interval=25))
;
```

Out[11]: ''



Conclusions extracted from the visualization

Temperature in Barcelona is on average **around 8 degrees celsius higher than the global average**, this difference has been preety much consistent since the beginning of the measurements.

Since these early measurements Barcelona has been facing very similar temperature trends and disruptions compared to the global average, including both main cold periords during the 19th century: the end of the little ice age period in the beginning of the XIXth century and the Krakatoa 1883 eruption in the end of the XIXth century, although this last one has been a bit more accentuated in Barcelona compared to the global trend.

However, in the last century the overall trends shows a global and rapid increse of temperature, the world is getting hotter, this fact is more evident from around 1970 onwards, with an increasing rise trend of the temperature

In [ ]:
```
```