Udacity - Data Analyst NanoDegree Program

Project #1: Weather Trends Exploration

By Samuel Jiménez Sanabria

Description

In this project I've compared measurements of world global temperatures to temperature measurements from the city of Stockholm in Sweden from the year 1750 to 2015. The steps to accomplish the analysis have been the following:

1. Explore the tables presented on the Udacity workspaces and extract the appropriate information from the tables.
2. Download the .csv files and import them into a Jupyter Notebook for exploration, plots and some computation.
3. Plots.
4. Other cities compared to Stockholm.
5. Analysis and interpretation.

Imports

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from IPython.display import Image

from IPython.core.display import Image, display
```

2. Read the data and explore the tables presented on the Udacity workspaces.

First I check if the City ‘Stockholm’ is present on the table ‘city_list’ with:

```
SELECT * FROM city_list WHERE city = 'Stockholm';
```

It results in one entry:

Output 1 results Download CSV

city	country
Stockholm	Sweden

Then I extract all the temperatures for Stockholm in ‘city_data’ with:

```
SELECT * FROM city_data WHERE city = 'Stockholm';
```

It results in 271 entries:

Output 271 results Download CSV

year	city	country	avg_temp
1743	Stockholm	Sweden	6.02
1744	Stockholm	Sweden	7.67
1745	Stockholm	Sweden	-2.33
1746	Stockholm	Sweden	
1747	Stockholm	Sweden	
1748	Stockholm	Sweden	

Data Exploration

```
global_data = pd.read_csv('data/global_data.csv')1
global_data.shape
```

```
(266, 2)
```

The global temperature average (mean) and other basic statistics can be obtained with '.describe'

```
global_data["avg_temp"].describe()
```

```
count    266.000000
mean       8.369474
std        0.584747
min        5.780000
25%        8.082500
50%        8.375000
75%        8.707500
max        9.830000
Name: avg_temp, dtype: float64
```

```
global_data.head()
```

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47

```
stk_data = pd.read_csv('data/stk_data.csv')

stk_data = stk_data.sort_values(by=['year'])

stk_data.head(10)
```

	year	city	country	avg_temp
65	1743	Stockholm	Sweden	6.02
194	1744	Stockholm	Sweden	7.67
92	1745	Stockholm	Sweden	-2.33
94	1746	Stockholm	Sweden	NaN
196	1747	Stockholm	Sweden	NaN
195	1748	Stockholm	Sweden	NaN
64	1749	Stockholm	Sweden	NaN
93	1750	Stockholm	Sweden	7.35
87	1751	Stockholm	Sweden	6.46
88	1752	Stockholm	Sweden	3.08

Stockholm's temperature average (mean) and other basic statistics can be obtained with '.describe'

```
stk_data["avg_temp"].describe()
```

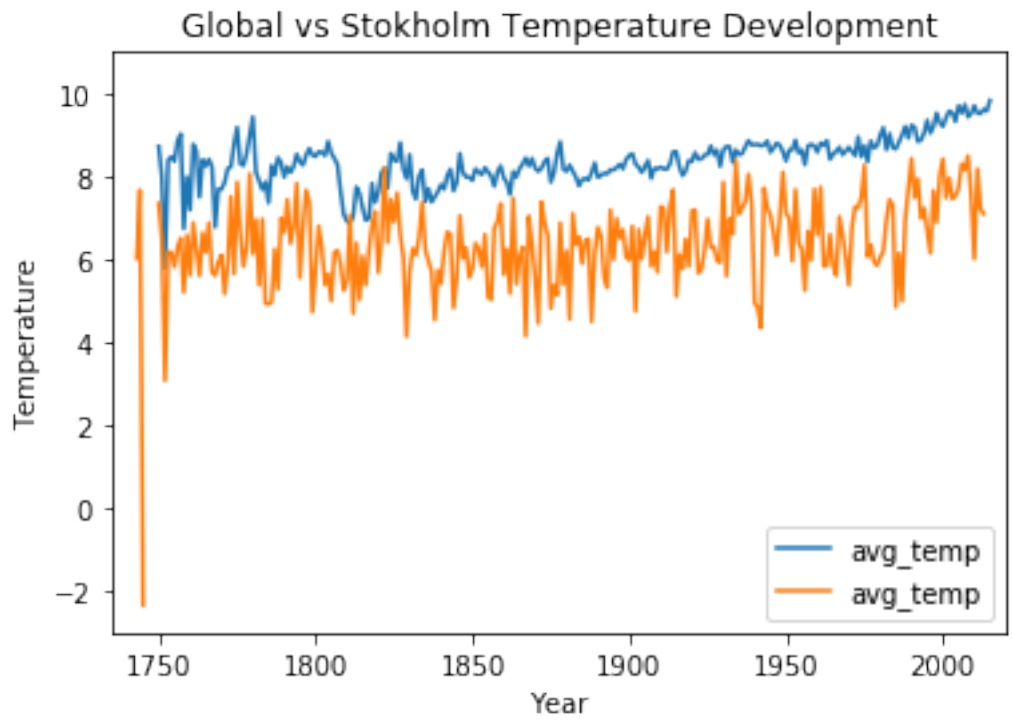
```
count    265.000000
mean       6.362302
std        1.080072
min       -2.330000
25%        5.780000
50%        6.310000
75%        7.050000
max        8.490000
Name: avg_temp, dtype: float64
```

stk_data.head()

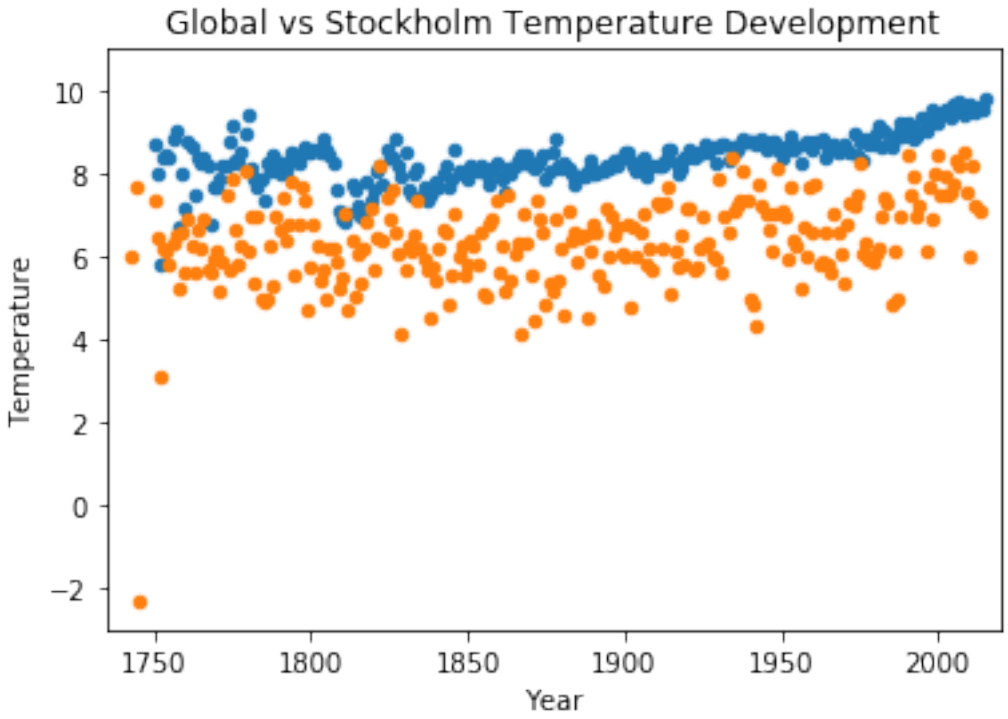
	year	city	country	avg_temp
65	1743	Stockholm	Sweden	6.02
194	1744	Stockholm	Sweden	7.67
92	1745	Stockholm	Sweden	-2.33
94	1746	Stockholm	Sweden	NaN
196	1747	Stockholm	Sweden	NaN

3. PLOTS

```
ax = plt.gca()

global_data.plot(kind='line',x='year',y='avg_temp', color='C0', ax=ax)

stk_data.plot(kind='line',x='year',y='avg_temp', color='C1', ax=ax)

plt.xlabel("Year")

plt.ylabel("Temperature")

plt.title("Global vs Stokholm Temperature Development")

plt.ylim((-3,11))

plt.xlim((1735,2020))

plt.show()
```




```
ax = plt.gca()

global_data.plot(kind='scatter',x='year',y='avg_temp', color='C0', ax=ax)

stk_data.plot(kind='scatter',x='year',y='avg_temp', color='C1', ax=ax)

plt.xlabel("Year")

plt.ylabel("Temperature")

plt.title("Global vs Stockholm Temperature Development")

plt.ylim((-3,11))

plt.xlim((1735,2020))

plt.show()
```



Join Datasets for Analysis

```
data_joined =  pd.merge(left = global_data, right = stk_data,
                        how = "outer", left_on = "year", right_on = "year")
data_joined.head()
```

	year	avg_temp_x	city	country	avg_temp_y
0	1750	8.72	Stockholm	Sweden	7.35
1	1751	7.98	Stockholm	Sweden	6.46
2	1752	5.78	Stockholm	Sweden	3.08
3	1753	8.39	Stockholm	Sweden	6.18
4	1754	8.47	Stockholm	Sweden	6.16

```
ax = plt.gca()

data_joined.plot(kind='line',x='year',y='avg_temp', color='C0', ax=ax)

stk_data.plot(kind='line',x='year',y='avg_temp', color='C1', ax=ax)

plt.xlabel("Year")

plt.ylabel("Temperature")

plt.title("Global vs Stokholm Temperature Development")

plt.ylim((-3,11))

plt.xlim((1735,2020))

plt.show()
```

Temperature correlation

```
data_joined['avg_temp_x'].corr(data_joined['avg_temp_y'])
```

0.49168629092726324

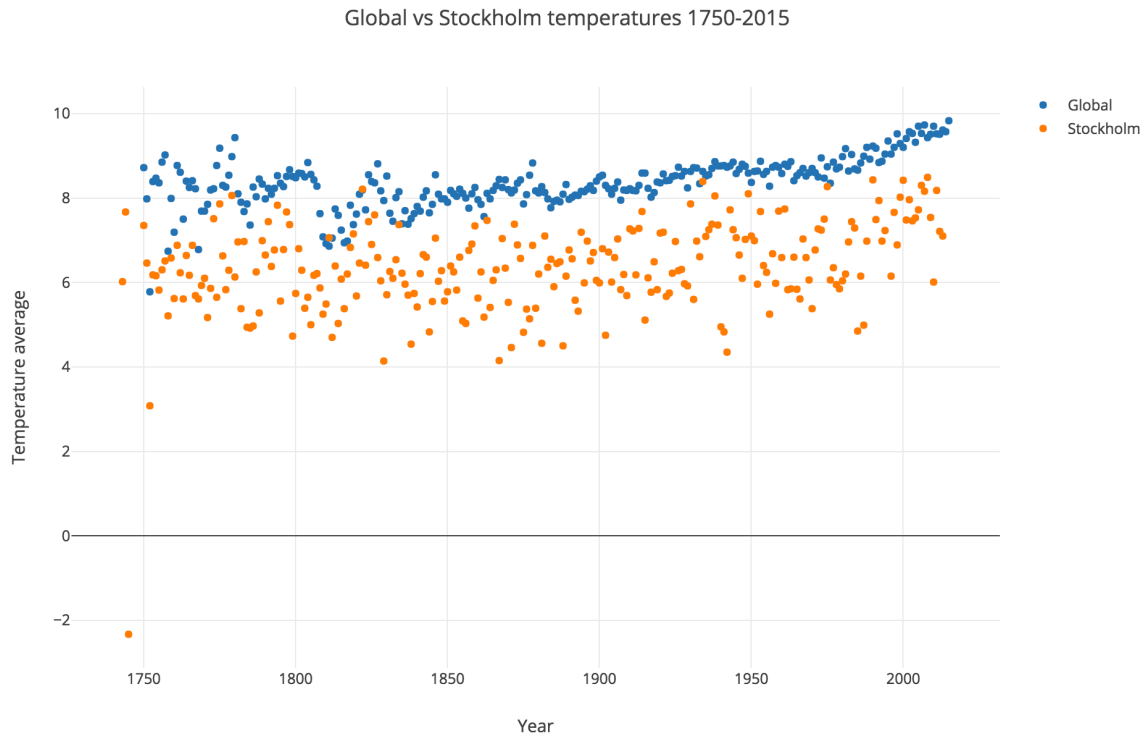
Moving Average

```
# Create a moving average (5 years) column for the global data

data_joined['global_ma'] =
data_joined['avg_temp_x'].rolling(window=5,center=False).mean()


# Create a moving average (5 years) column for Stockholm data

data_joined['stk_ma'] =
data_joined['avg_temp_y'].rolling(window=5,center=False).mean()

data_joined.head(10)
```

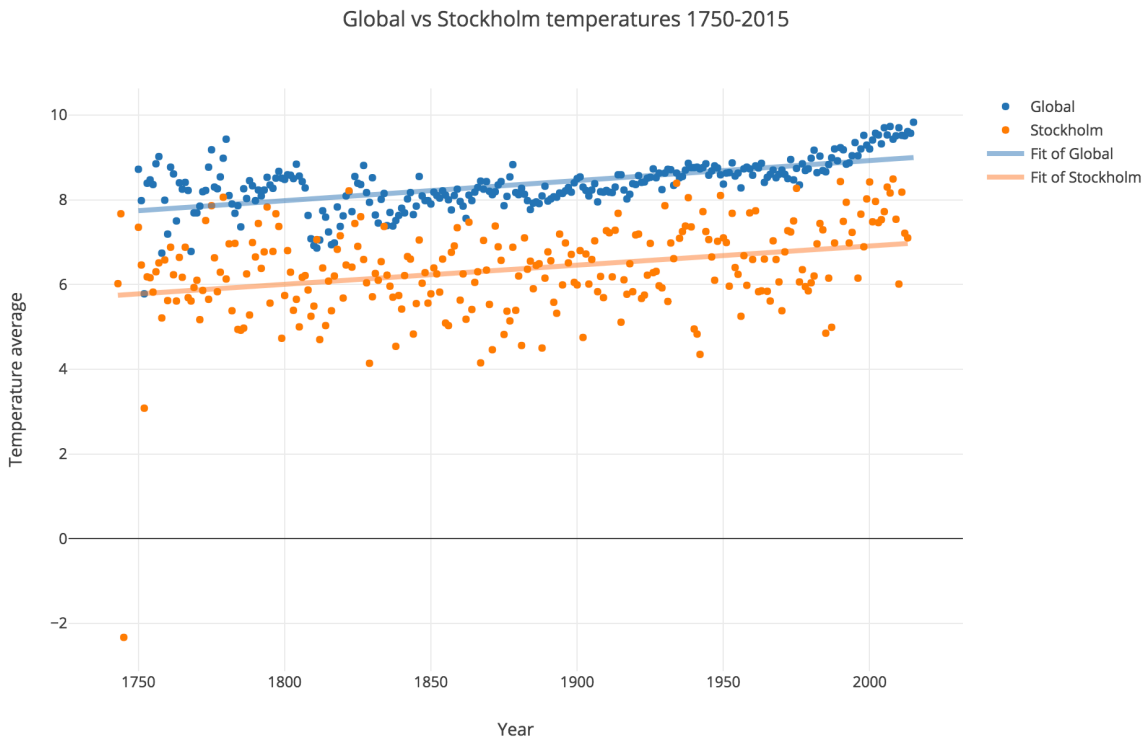
	year	avg_temp_x	city	country	avg_temp_y	global_ma	stk_ma
0	1750	8.72	Stockholm	Sweden	7.35	NaN	NaN
1	1751	7.98	Stockholm	Sweden	6.46	NaN	NaN
2	1752	5.78	Stockholm	Sweden	3.08	NaN	NaN
3	1753	8.39	Stockholm	Sweden	6.18	NaN	NaN
4	1754	8.47	Stockholm	Sweden	6.16	7.868	5.846
5	1755	8.36	Stockholm	Sweden	5.82	7.796	5.540
6	1756	8.85	Stockholm	Sweden	6.30	7.970	5.508
7	1757	9.02	Stockholm	Sweden	6.51	8.618	6.194
8	1758	6.74	Stockholm	Sweden	5.21	8.288	6.000
9	1759	7.99	Stockholm	Sweden	6.58	8.192	6.084

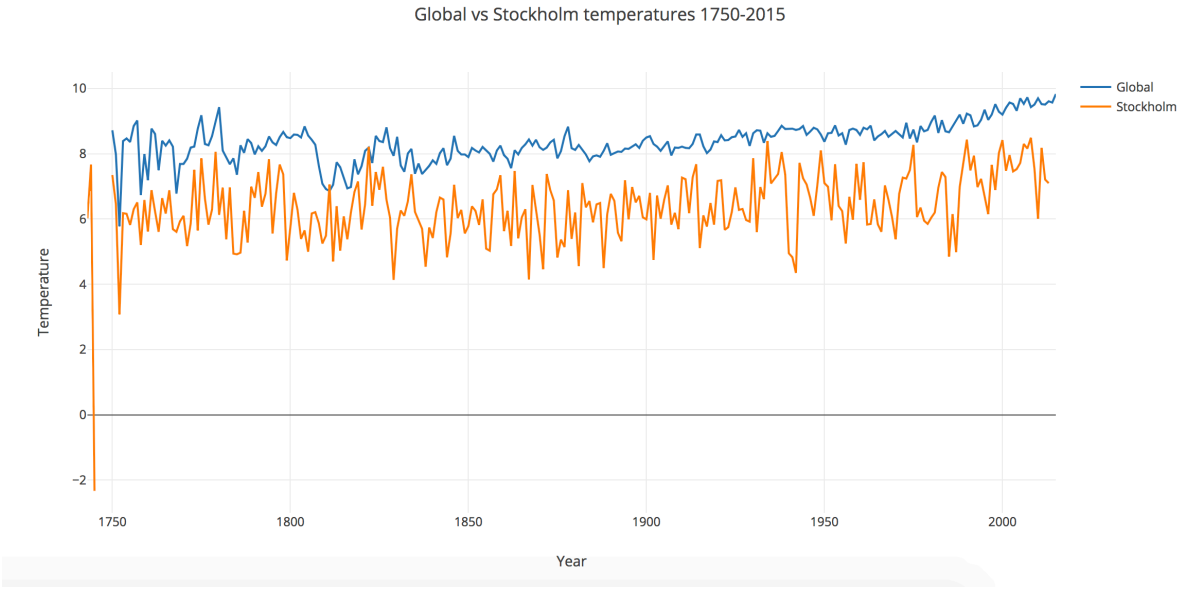
Create additional plots for easier visualisation.

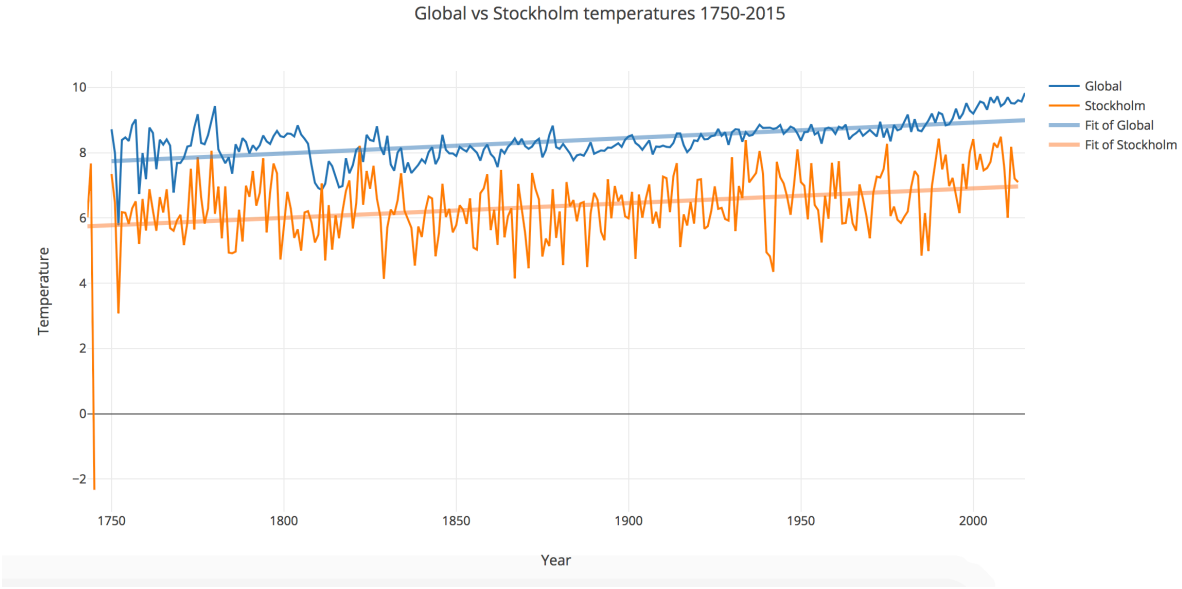
First a ScatterPlot showing both global and Stockholm temperatures



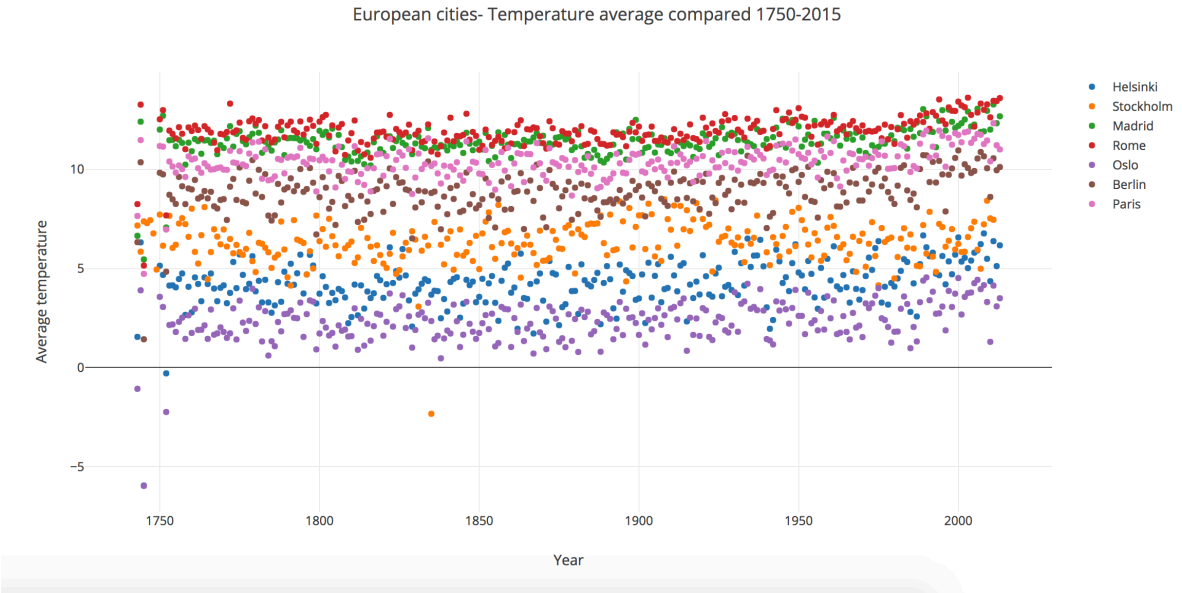
The second plot has a "Curve Fitting" Linear function applied (M*X +B)




Global vs Stockholm temperatures 1750-2015
Global
Stockholm
Temperature
Year
10
8
6
4
2
0
−2
1750
1800
1850
1900
1950
2000


Global vs Stockholm temperatures 1750-2015
Global
Stockholm
Fit of Global
Fit of Stockholm
Temperature
Year
10
8
6
4
2
0
−2
1750
1800
1850
1900
1950
2000

4. Other cities compared to Stockholm



5. Analysis and interpretation.

- **Overall trends?** We can observe a clear trend where temperatures have been raising steadily during the whole period. The raise in averages seems more abrupt from the seventies and forward. From the charts we could conclude that the world is indeed getting hotter, and that the raise in temperature is growing at a higher rate since the seventies.

- **Comparing Globals to Stockholm** Stockholm is a much colder city than the world's average. Temperature averages are also more disperse with wider ranges between the lowest points and highest peaks. But this difference in temperature between global and Stockholm's averages is consistent during the whole period except during the last forty-fifty years, where Stokholm seems to have been able to slowdown this raising trend compared to the global averages. Even then, Stockholm has also seen average temperatures increasing steadily during the whole period.

- **Interesting areas** Observing the chart "Other cities compared to Stocholm" we can see that hotter cities have also narrower temperature ranges. We can also see that Cities like Rome, Madrid and Berlin have been suffering fro higher temperature increases during the last half century.

- **Negative values** It's interesting to observe that some cities have had negative average temperatures (minus degrees) sometime around 1750 (Stockholm) and 1830 (Oslo). Is this data real? Were those years so cold? Because these "suspects" are actually cold cities we could conclude that this is actually real, verified data, but it could be studied more in detail.

- **Finally** Stockholm's average temperature numbers have been given in Celsius. The "mean" we obtained in **stk_data["avg_temp"].describe()** is 6.362302. But is this correct? After comparing my numbers to those presented and explained **here** I can confirm that the numbers seem legit.

Image for analysis reference only