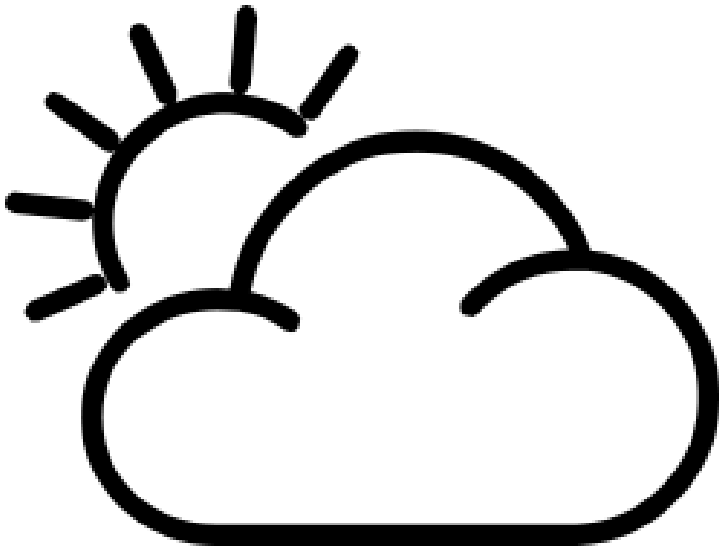


Exploring Weather Trends

06.05.2021

—

Prepared by: Yasemin Topchu
Udacity, Data Analyst Nanodegree
Moscow, Russia

Overview

This paper is the result of working with local and global temperature data.

Goals

1. Extract the data from the database.
2. Make analysis about global temperature trends and temperature trends in Moscow, and compare them.

To understand how to visualize the results of data analysis, let's think about what questions we want to answer. First, we would be interested to know how the different values of the years that will be included in the calculation of the moving average affect the behavior of the graphs - for this, we will build a series of graphs for different values of the years. For example, let's consider a moving average for 10, 20, 30, 40 and 50 years. Secondly, we need to compare global temperature changes with city ones - for this we need to build comparative graphs. The next interesting question is whether we get similar temperature dependences for the city we are studying and for the whole world - in order to do this, it is necessary to conduct an analysis related to the correlation of the values we obtained.

Tools

1. SQL
2. Python

Technical Details

Moving Average

From chapter “Moving averages” , Data Analyst Nanodegree, Udacity:

“Moving averages are used to smooth out data to make it easier to observe long term trends and not get lost in daily fluctuations”.

We use next formula to find moving averages for our data:

$$z_{MA} = (x_1 + x_2 + ... + x_{k-1}) / k$$

Key Stages

I. Extracting the data from the database

First of all we need to export the temperature data with workspace in one of the preparing sections for this task. This workspace is connected to a database, and there are three tables: **city_list**, which contains a list of cities and countries to help choose city nearest to us, **city_data**, which contains the average temperatures for each city by year, and **global_data**, which contains the average global temperatures by year. To interact with the database, we need to write SQL queries:

```
SELECT *
        FROM global_data;
SELECT *
        FROM city_data
                WHERE city = 'Moscow';
```

II. Opening up the CSV

Then we need to open up CSV files with Moscow and global temperature averages using Python.

	year	city	country	avg_temp
0	1743	Moscow	Russia	-2.57
1	1744	Moscow	Russia	6.64
2	1745	Moscow	Russia	-6.13
3	1746	Moscow	Russia	NaN
4	1747	Moscow	Russia	NaN
...	...	...	...	...
266	2009	Moscow	Russia	5.69
267	2010	Moscow	Russia	5.91
268	2011	Moscow	Russia	6.01
269	2012	Moscow	Russia	5.20
270	2013	Moscow	Russia	6.80

271 rows × 4 columns

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47
...	...	...
261	2011	9.52
262	2012	9.51
263	2013	9.61
264	2014	9.57
265	2015	9.83

266 rows × 2 columns

We see that the files have some NaN values, so let's get rid of them:

```
df_city.dropna(subset = ['avg_temp'], inplace=True)
df_global.dropna(subset = ['avg_temp'], inplace=True)
```

	year	city	country	avg_temp
0	1743	Moscow	Russia	-2.57
1	1744	Moscow	Russia	6.64
2	1745	Moscow	Russia	-6.13
7	1750	Moscow	Russia	4.84
8	1751	Moscow	Russia	5.07
...	...	...	...	...
266	2009	Moscow	Russia	5.69
267	2010	Moscow	Russia	5.91
268	2011	Moscow	Russia	6.01
269	2012	Moscow	Russia	5.20
270	2013	Moscow	Russia	6.80

267 rows × 4 columns

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47
...	...	...
261	2011	9.52
262	2012	9.51
263	2013	9.61
264	2014	9.57
265	2015	9.83

266 rows × 2 columns

We get 267 rows instead of 271 for the first file.

III. Counting moving average using Python

```
for i in range(0, df_city.shape[0] - (k - 1)):
    sum_k = sum([df_city.iloc[i + k_i, 3] for k_i in range(k)]) / k
    df_city.loc[df_city.index[i + k - 1], str(f'{k}-years Moving Average')] = sum_k
```

So we get one new column “k-years Moving Average”. For example, for Moscow moving average and k=10:

	year	city	country	avg_temp	10-years Moving Average
0	1743	Moscow	Russia	-2.57	NaN
1	1744	Moscow	Russia	6.64	NaN
2	1745	Moscow	Russia	-6.13	NaN
7	1750	Moscow	Russia	4.84	NaN
8	1751	Moscow	Russia	5.07	NaN
...	...	...	...	...	...
266	2009	Moscow	Russia	5.69	5.655
267	2010	Moscow	Russia	5.91	5.650
268	2011	Moscow	Russia	6.01	5.719
269	2012	Moscow	Russia	5.20	5.676
270	2013	Moscow	Russia	6.80	5.857

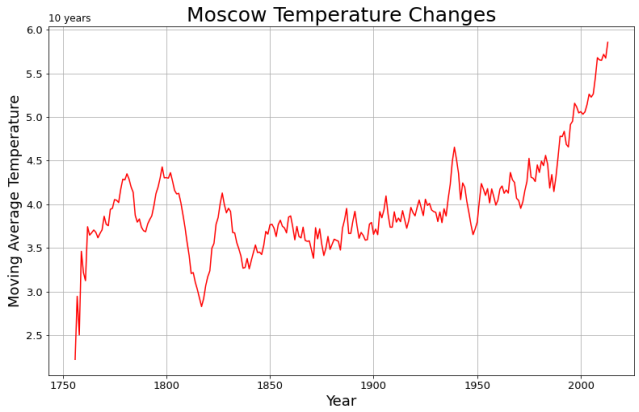
267 rows × 5 columns

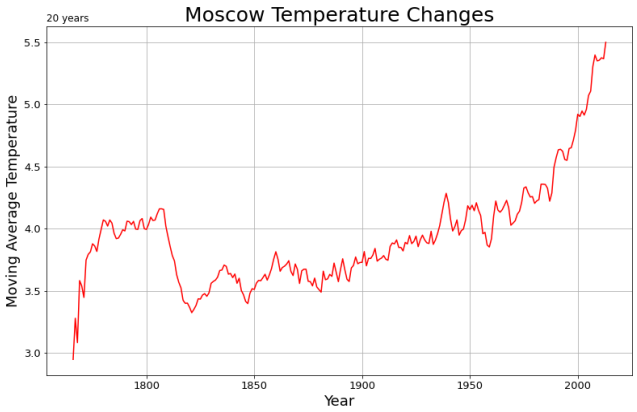
After that there are some NaN values, and we can drop them:

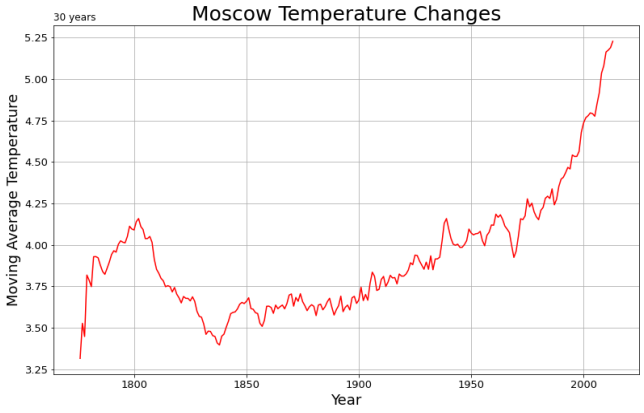
```
df_city.dropna(subset = [str(f'{k}-years Moving Average')], inplace=True)
df_global.dropna(subset = [str(f'{k}-years Moving Average')], inplace=True)
```

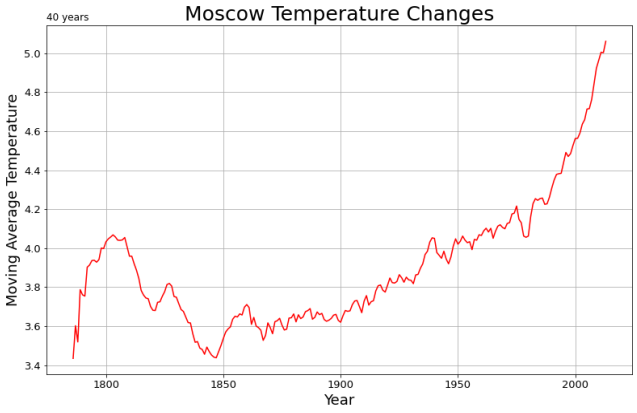
IV. Creating plots for moving average using Python

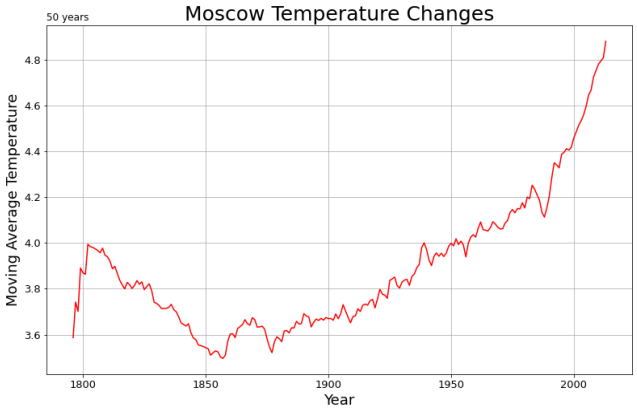
Let's get graphs of Moscow and global moving temperature average values for 10, 20, 30, 40 and 50 years:

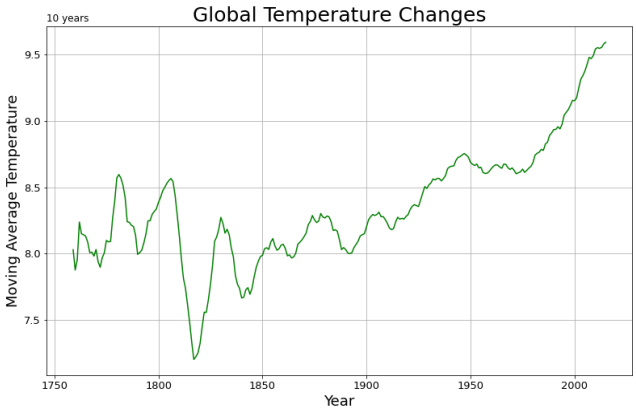


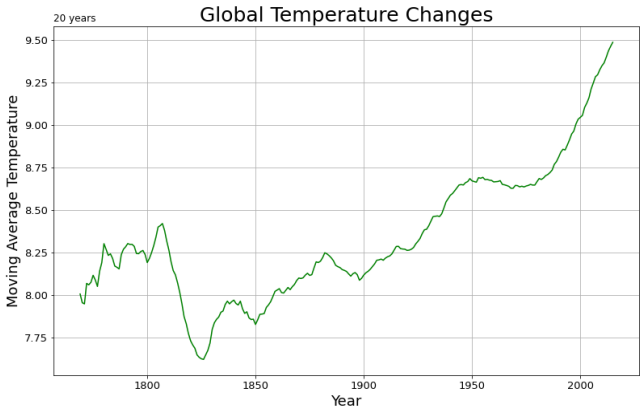


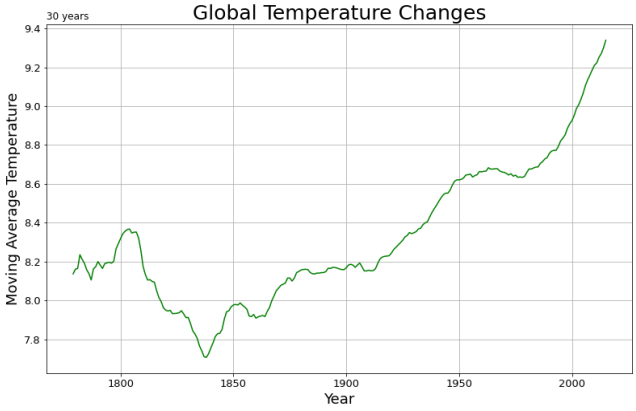


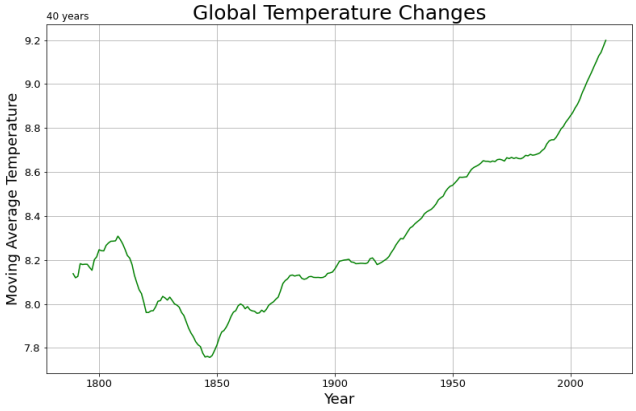


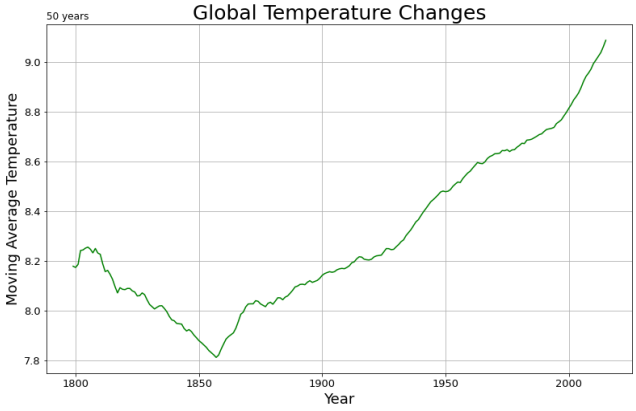



10 years
Global Temperature Changes
Moving Average Temperature
Year


20 years
Global Temperature Changes
Moving Average Temperature
Year

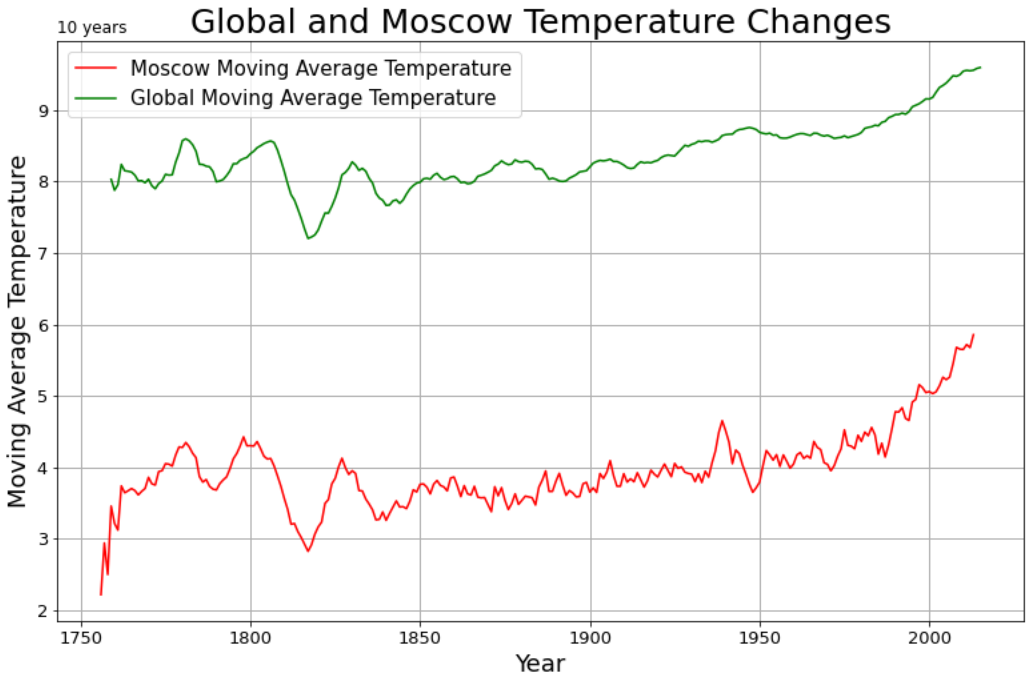

30 years
Global Temperature Changes
Moving Average Temperature
Year

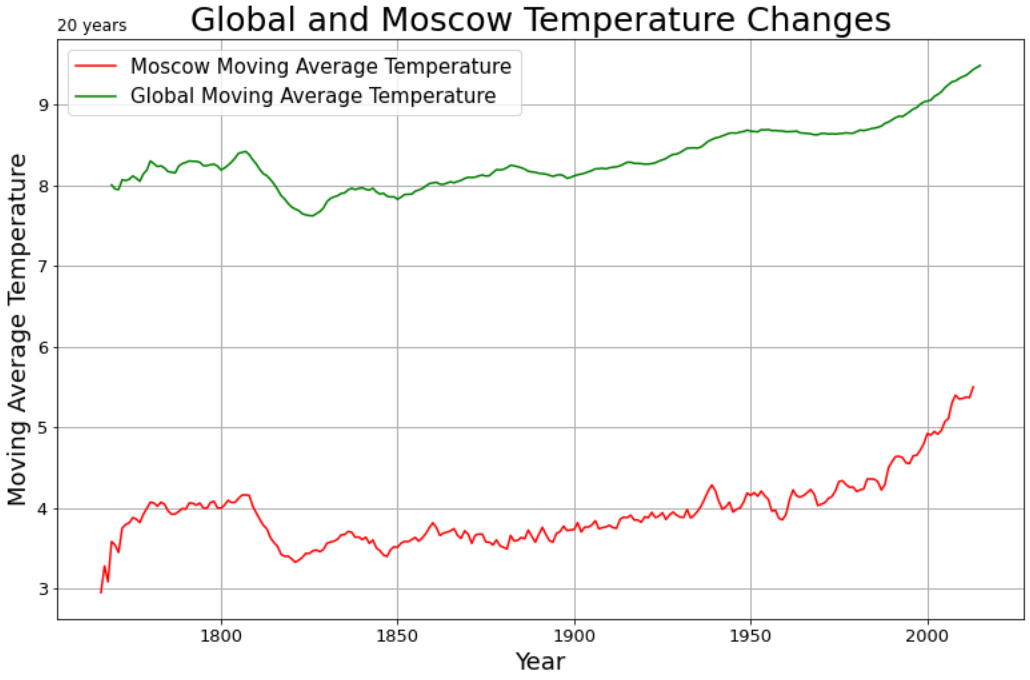

40 years
Global Temperature Changes
Moving Average Temperature
Year

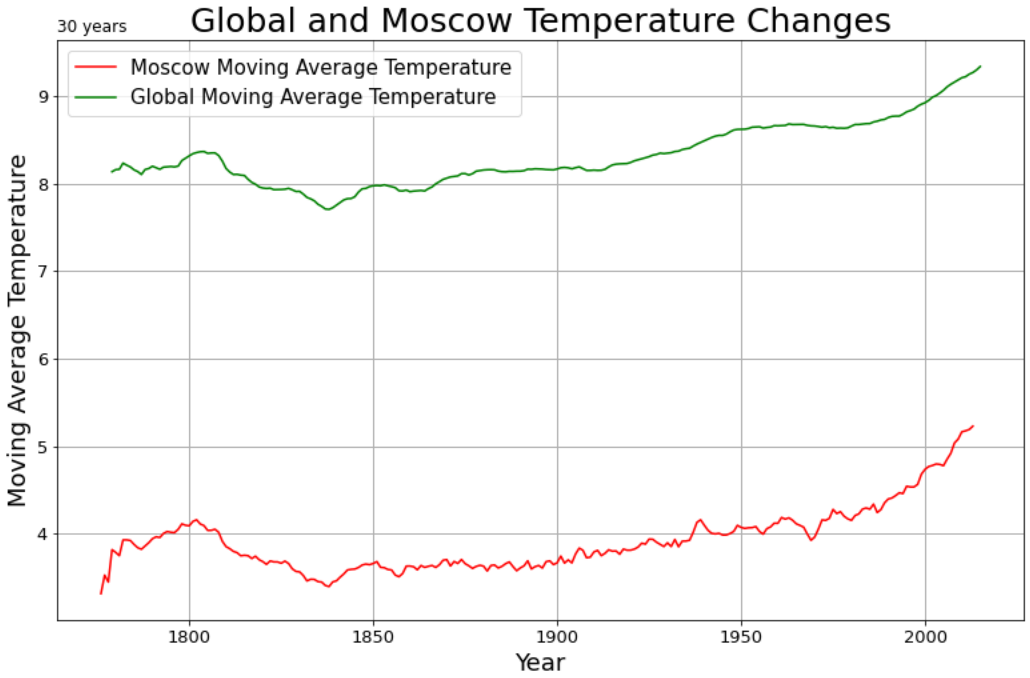

50 years
Global Temperature Changes
Moving Average Temperature
Year

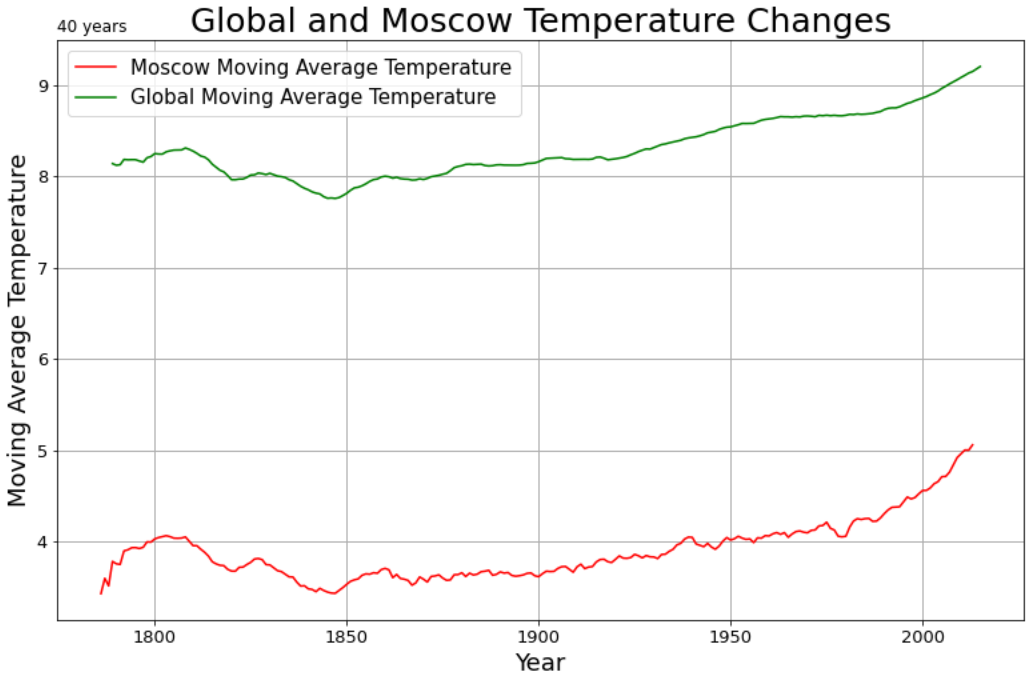
V. Comparing results

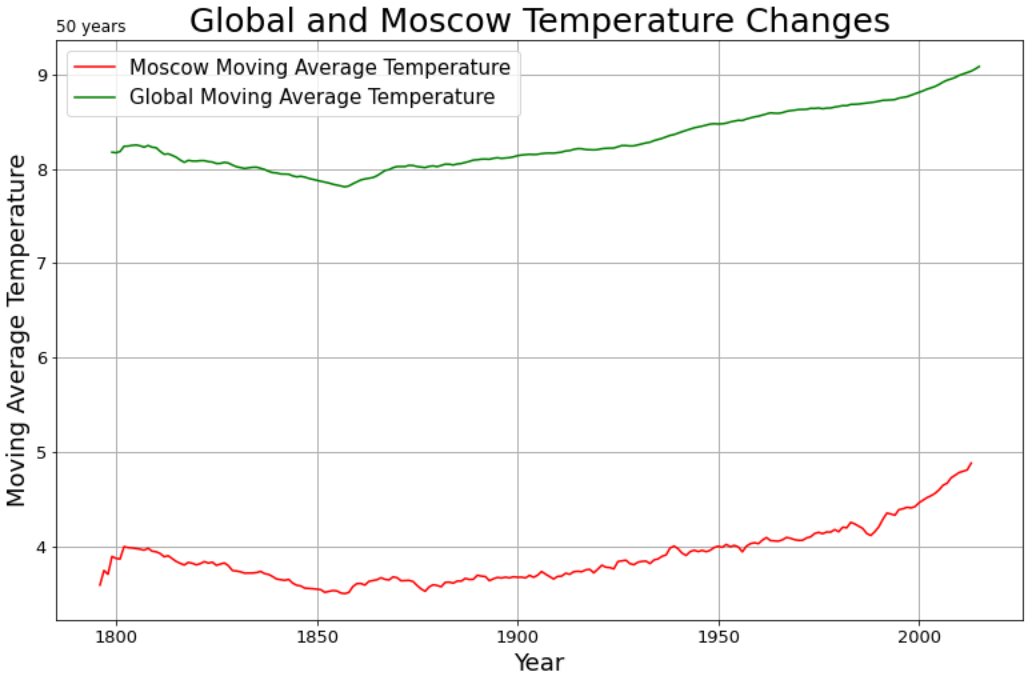
And now let's combine our graphs for visual comparison.





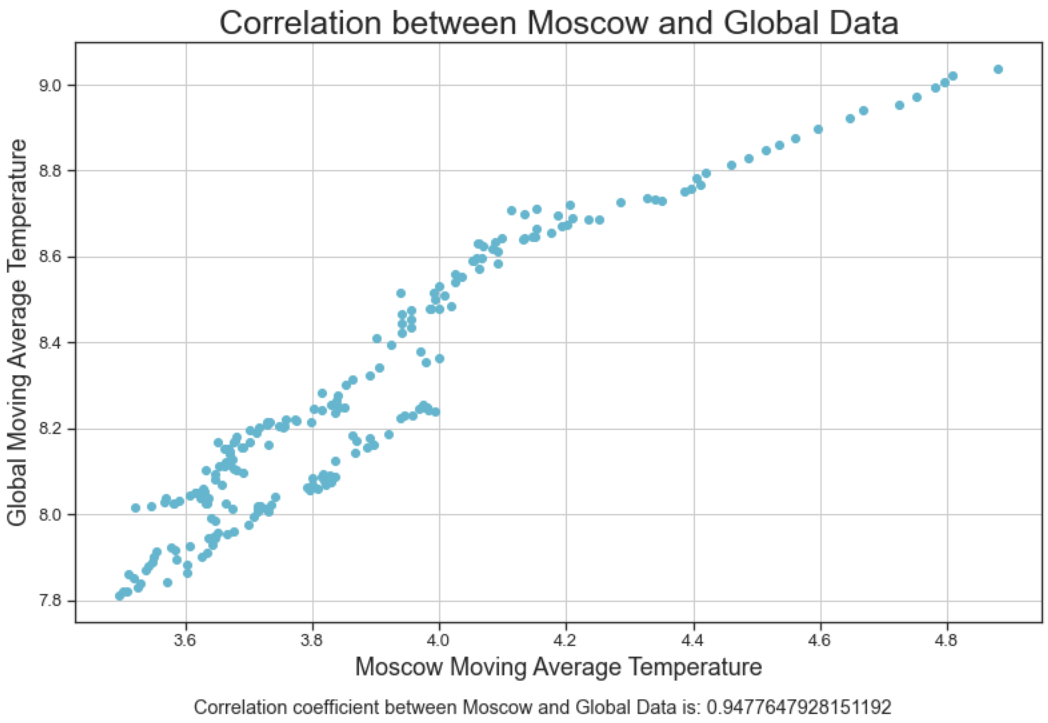

30 years
Global and Moscow Temperature Changes
Moscow Moving Average Temperature
Global Moving Average Temperature
Moving Average Temperature
Year
1800
1850
1900
1950
2000
4
5
6
7
8
9


40 years
Global and Moscow Temperature Changes
Moscow Moving Average Temperature
Global Moving Average Temperature
Moving Average Temperature
Year
1800
1850
1900
1950
2000
4
5
6
7
8
9


Global and Moscow Temperature Changes
50 years
Moscow Moving Average Temperature
Global Moving Average Temperature
Moving Average Temperature
Year
9
8
7
6
5
4
1800
1850
1900
1950
2000

VI. Correlation

Let's also get a scatterplot for changes of our two moving averages: Moscow Moving Average and Global Moving Average:



VII. Conclusions

First, let's think about the technical side of the issue. It is noticeable that in the period from 1750 to 1850 there are large peaks and troughs in the graphs. This can be explained by the fact that due to the absence of some temperature values in the table for this time period, the corresponding rows were discarded from consideration and graphing.

As we can see, within the range of values we have chosen for the moving average, the graph becomes smoother as the number of years increases. This may be due to the fact that fluctuations of the average over a certain number of years become smaller, the more the sample size increases and the more it begins to behave like a general population.

Now it is necessary to draw conclusions about the trends in the average temperature in the world and in Moscow. We know that human activity has been

demonstrating the fruits of its labor in the form of global warming for half a century. Looking at the resulting graphs, we see that the temperature is steadily increasing. It is known that the last 10 years were the hottest on record in the world, which we see on our graphs. The graphs also show that the average temperature in Moscow has a similar trend. The correlation between the temperature values in Moscow and in the world is strong and positive, as evidenced by a positive correlation coefficient close to one, and which we can clearly see by paying attention to the scatter plot in Section V. It follows that the trends in the mean annual temperature in the world and in Moscow have a similar character.