# Exploring Weather Trends

Data Analyst Udacity Nanodegree Program

Rodrigo Brust Santos

# 1. Softwares

To develop this 1st Project, I used the following softwares:

- **SQL**: to extract data from the DB. I chose 4 cities: Rio de Janeiro, Shanghai, Sydney and Ohama.
- **Excel**: to treat data, to calculate the moving averages and the linear correlation coefficent.
- **Python**: to create the graphics and challenge myself. Making a graphic in Excel would be easy, and since I'm in this program because I want to learn learn more, I think making this way would be better for me.

# 1.1- Extracting data from DB with SQL

a) First, I made a quick analysis of all data in the DB;
b) After that, I chose my city's data and global data;
c) Then, I chose others cities, so the comparison could be improved;
d) The following step was the download;

Extracting Rio de Janeiro Data:

```
SELECT *
FROM city_data
WHERE city = 'Rio De Janeiro'
AND year > 1831
AND year < 2014
```

# 1.2- Treating data and calculus with Excel

a) First, I google on how to calculate the moving average

b) After that, I created a new column with the moving averages (MA) using Excel formula. I used a 3 years MA. I made na analysis from 2 to 5 years MA, and in my opinion, 3 years MA fits the best.

c) With Excel I also calculated the linear correlation coefficent for the Global, Rio de Janeiro, Shanghai, Sydney and Omaha average temperature. I will talk about the results ahead.

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

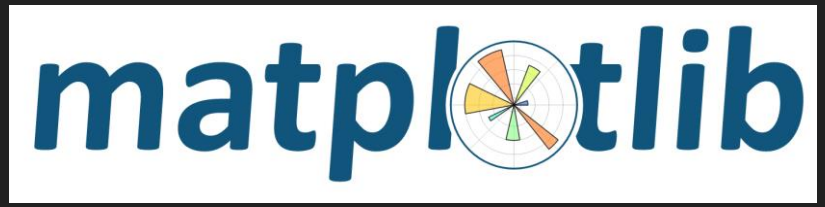
# 1.3- Creating graphics with Python

a) Since I already had the data, I started reading Matplotlib documentation;

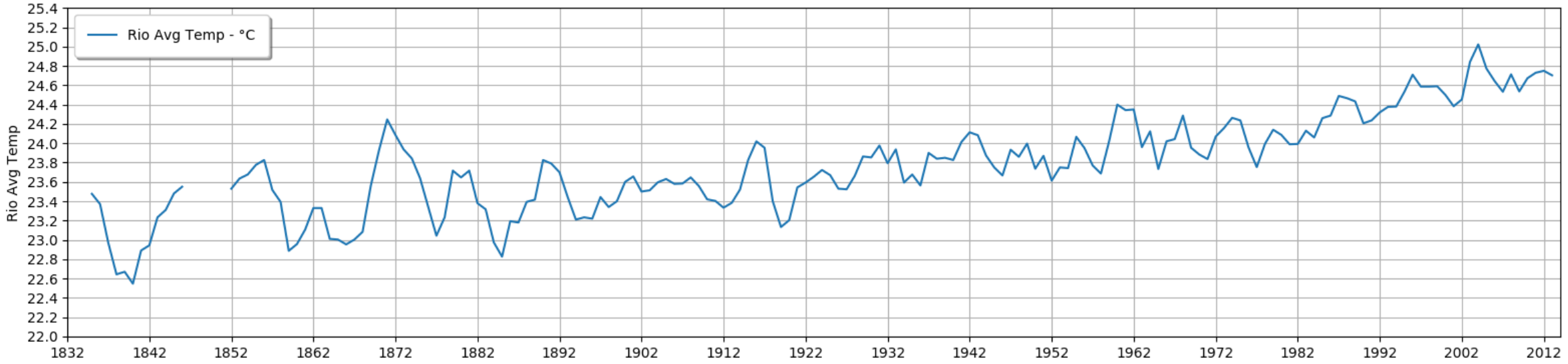
b) Then, I started making some graphic tests, trying to find the best fit.

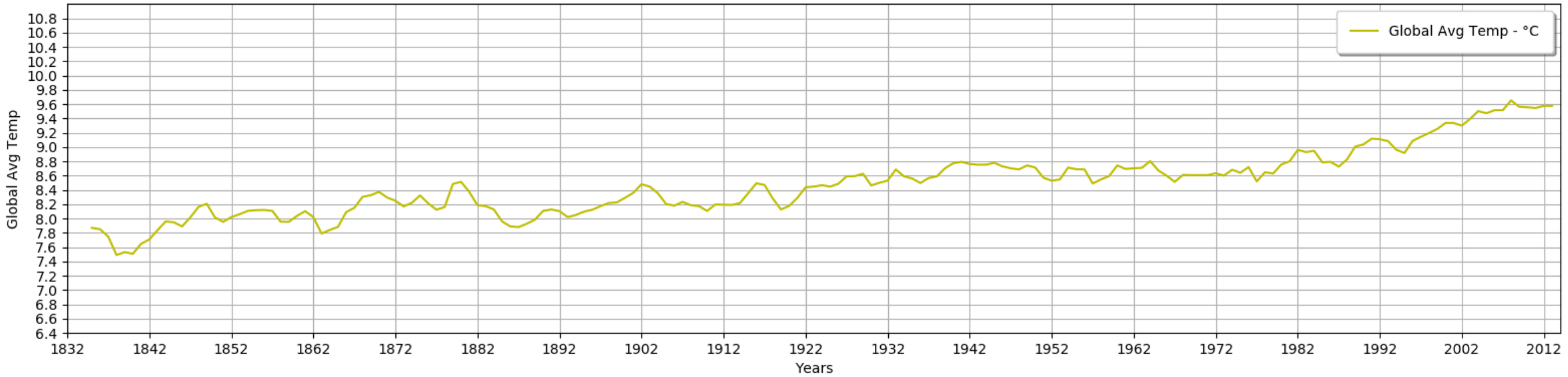
c) After one day working on the script, trying to understand erros, I finally got results.

d) The key consideration on how to visualize the trends is that they must be near each other, so we can see wether both trends follow up each other or show different results.

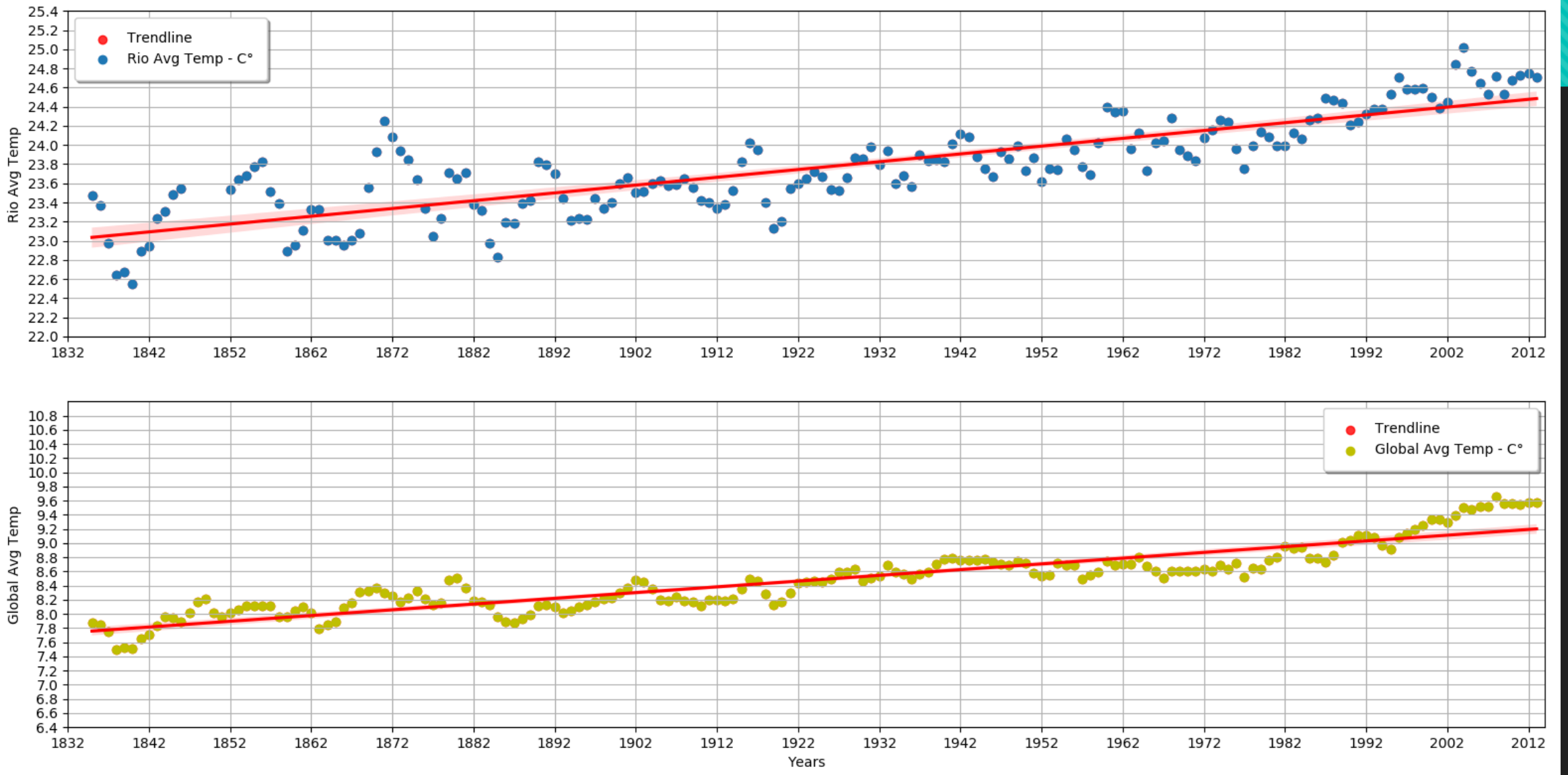


# 2.1 Data – Rio de Janeiro vs Global

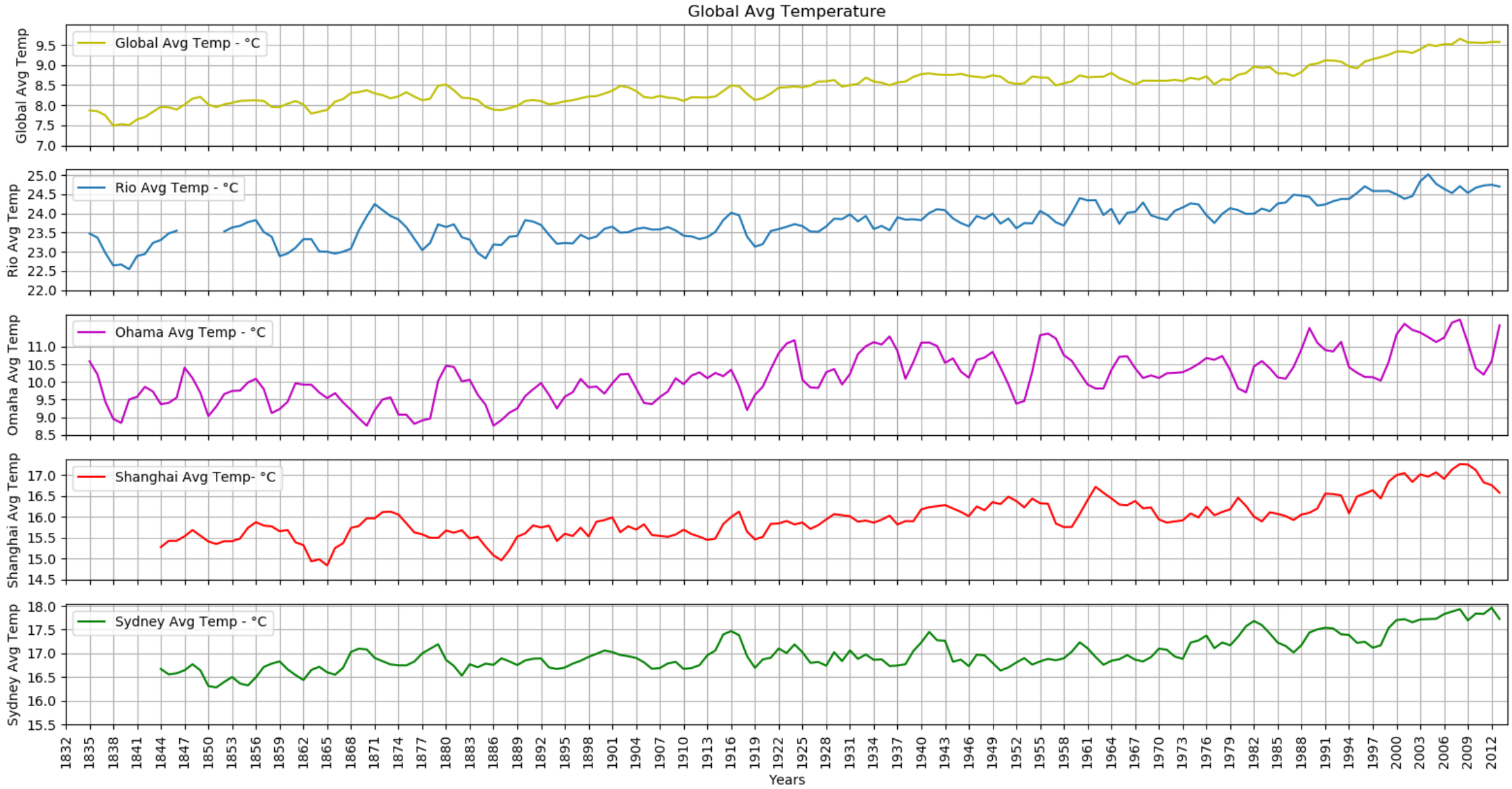


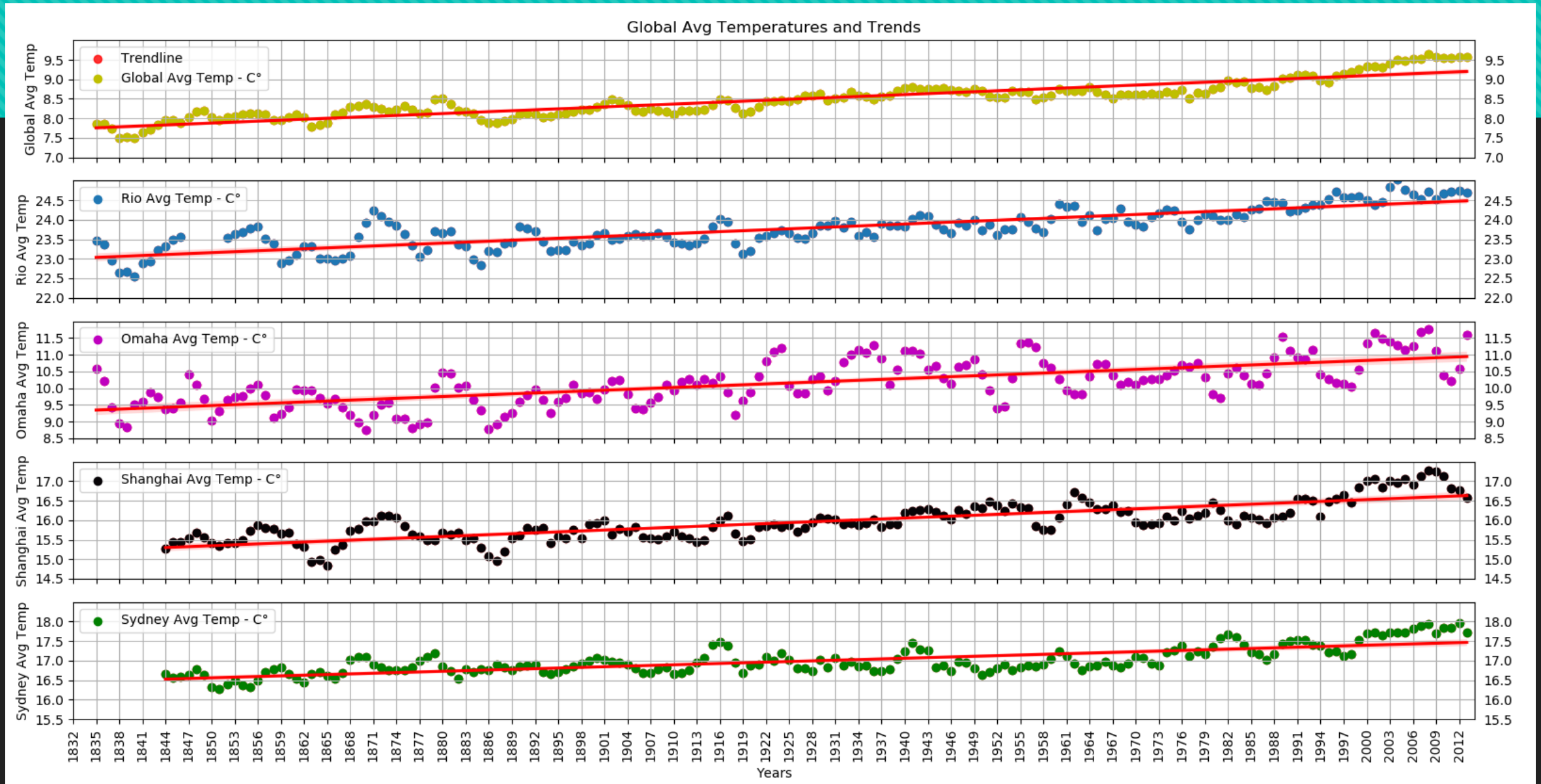
# 2.1 Data – Rio de Janeiro vs Global Trend



# 2.2 Data – Global vs Rio vs Om vs Sh vs Syd

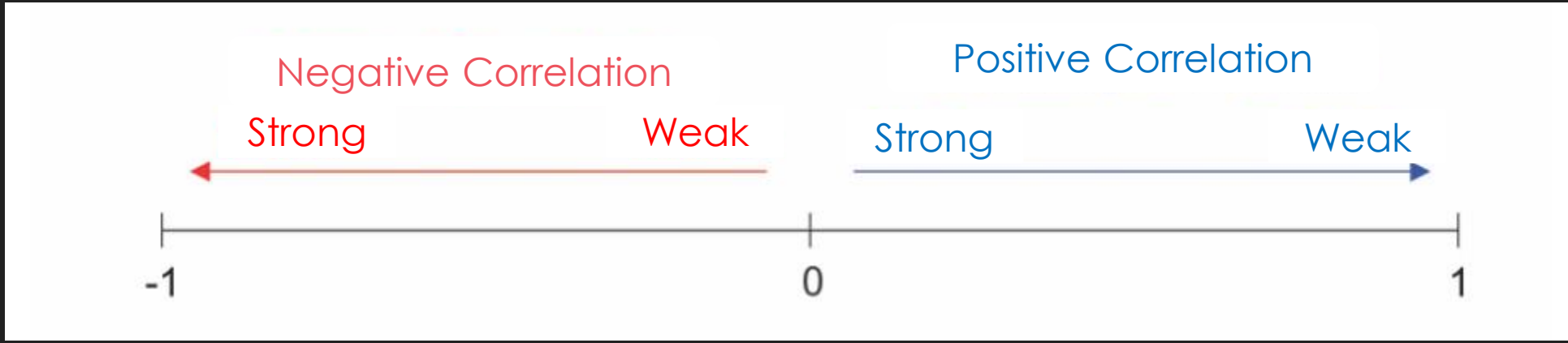


# 2.2 Data – Global vs Rio vs Om vs Sh vs Syd



# 2.3 Data – Linear Correlation Coefficent

Looking for a correlation between the average temperature in each city and the passing years, we can see, in the next slide, a Positive Weak Correlation for all analysis. This represente that the tendency of the temperature is to get higher, but sometimes it might not occur.



# 2.3 Data – Linear Correlation Coefficent

| Global | Rio de Janeiro | Omaha | Shanghai | Sydney |
| --- | --- | --- | --- | --- |
| R = 0.549281 | R = 0.400976 | R = 0.503665 | R = 0.395205 | R = 0.340718 |

# 3. Observations

- Global Average Temperature has been raising since 1832. Since then, the temperature increased between 1.5°C and 2°C.
- Rio de Janeiro and Sydney have shown similarties variation in temperature averages. There are at least 5 episodes that their curves responded almost the same (1868-1880, 1913-1919, 1940-1943, 1958-1961, 1979-1985, 2000-2012).
- Shanghai sometimes responds the same way those two cities.
- Omaha has the most variable average through the decades.

# 3. Observations

From 1832 to 2012:

- Global Temperature have risen between +1.5°C and +2°C;
- Rio de Janeiro is about  +2°C to +2.5°C hotter;
- Omaha is thereabout +2.5°C hotter;
- Sydney shows the smaller temperature increase between +1.25°C and +1.5°C;
- Shanghai temperature has risen about +2°C.

# 4. References

Where I looked for help:


https://pt.wikipedia.org/wiki/Coeficiente_de_correlação_de_Pearson

https://operdata.com.br/blog/coeficientes-de-correlacao/

https://www.w3schools.com/python

https://matplotlib.org/contents.html

https://stackoverflow.com/questions/54308172/adding-a-trend-line-to-a-matplotlib-line-plot-python