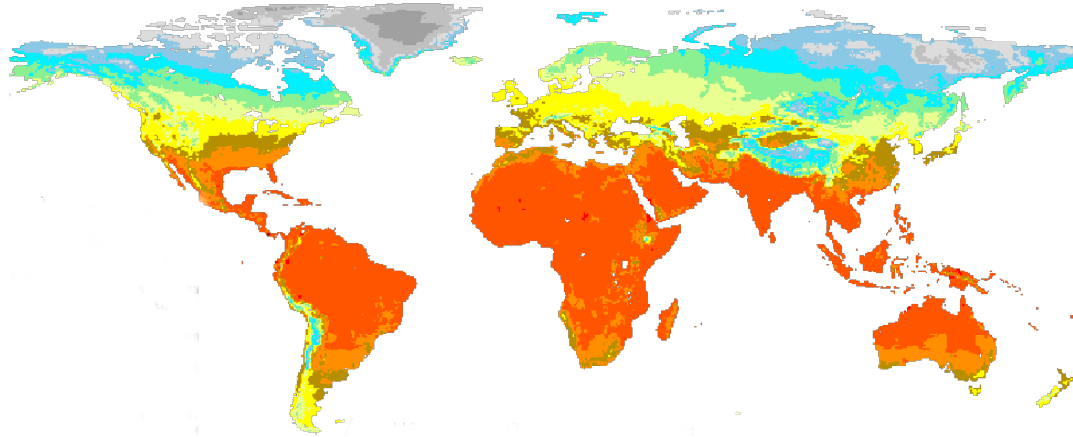# Exploring Weather Trends

Data Analysis Nano-Degree - Udacity

Aviad Giat

Feb 9, 2019

Outline

I chose to do this project on the biggest city next to me, which is San Francisco, California. I added 3 more cities to this analysis to compare the average weekly temperatures in SF with them. To accomplish the visualization of this project I used a few SQL lines to download the global average temperatures and the average temperatures in: San Francisco; Rio De Janeiro; London; and Helsinki. I downloaded the raw temperature data from Udacity's server to my machine; I used R to write the code and Rstudio to produce the .pdf file. With R I created a new column for each table (xls file in this stage) with the weekly average from the 7th record until the last one. This left 6 empty rows (first lines that do not have 7 days prior to its date). Before plotting the data I merged the datasets mentioned above to one dataframe called 'weather'. To plot the data I used the package ggplot2, which I worked with before and is a great tool for fast and easy plotting.

The SQL line I used on Udacity's server to retrieve the data I wanted:

select * from city_data where city = 'San Francisco'; select * from global_data; select * from city_data where city = 'Helsinki' select * from city_data where city = 'Rio De Janeiro' select * from city_data where city = 'London' select * from city_data where city = 'San Francisco'

A quick examination of the dataset

The first lines of the merged dataset

```
##   year   city         country weekly_avg_london weekly_avg_global
## 1 1749 London United Kingdom              7.34                NA
##   weekly_avg_sf weekly_avg_helsinki weekly_avg_rio
## 1            NA                  NA             NA
```

Basic statistics and structure of the different variables

```
##       year          city              country          weekly_avg_london
##  Min.   :1749   Length:865         Length:865         Min.   : 7.340
##  1st Qu.:1850   Class :character   Class :character   1st Qu.: 9.180
```

1

```
##  Median :1905   Mode  :character   Mode  :character   Median : 9.400
##  Mean   :1900                                         Mean   : 9.431
##  3rd Qu.:1959                                         3rd Qu.: 9.610
##  Max.   :2013                                         Max.   :10.780
##                                                       NA's   :600
##  weekly_avg_global weekly_avg_sf   weekly_avg_helsinki weekly_avg_rio
##  Min.   :7.190     Min.   :13.85   Min.   :0.640       Min.   :22.80
##  1st Qu.:8.090     1st Qu.:14.18   1st Qu.:3.890       1st Qu.:23.48
##  Median :8.330     Median :14.41   Median :4.160       Median :23.73
##  Mean   :8.414     Mean   :14.44   Mean   :4.229       Mean   :23.77
##  3rd Qu.:8.650     3rd Qu.:14.64   3rd Qu.:4.530       3rd Qu.:24.05
##  Max.   :9.590     Max.   :15.18   Max.   :5.850       Max.   :24.78
##  NA's   :14        NA's   :706     NA's   :600         NA's   :690

## 'data.frame':    865 obs. of  8 variables:
##  $ year               : int  1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 ...
##  $ city               : chr  "London" "London" "London" "London" ...
##  $ country            : chr  "United Kingdom" "United Kingdom" "United Kingdom" "United Kingdom" ...
##  $ weekly_avg_london  : num  7.34 8.24 8.12 8.93 9.05 9.08 9.06 9.11 8.98 8.82 ...
##  $ weekly_avg_global  : num  NA NA NA NA NA NA NA 8.08 8.12 7.94 ...
##  $ weekly_avg_sf      : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ weekly_avg_helsinki: num  NA NA NA NA NA NA NA NA NA NA ...
##  $ weekly_avg_rio     : num  NA NA NA NA NA NA NA NA NA NA ...
```

We can see above the years and weekly averages for San Francisco, Helsinki, London, Rio De Janeiro and for the entire Globe's temperatures statistics. The minimum weekly average of SF was 14 degrees and the max was 15. The global weekly average temperature was 7 at the minimum and 9.5 at the maximum. We can say that San Francisco is on the warm side of the planet's temperature distribution. Let's examine the correlation coefficient of San Francisco and the rest of the cities.

The correlation coefficient between the years variable and the weekly average in San Francisco

San Francisco

$R = 0.67$, $p < 2.2e-16$

Temperature (y-axis)

Years and centuries (San Francisco) (x-axis)
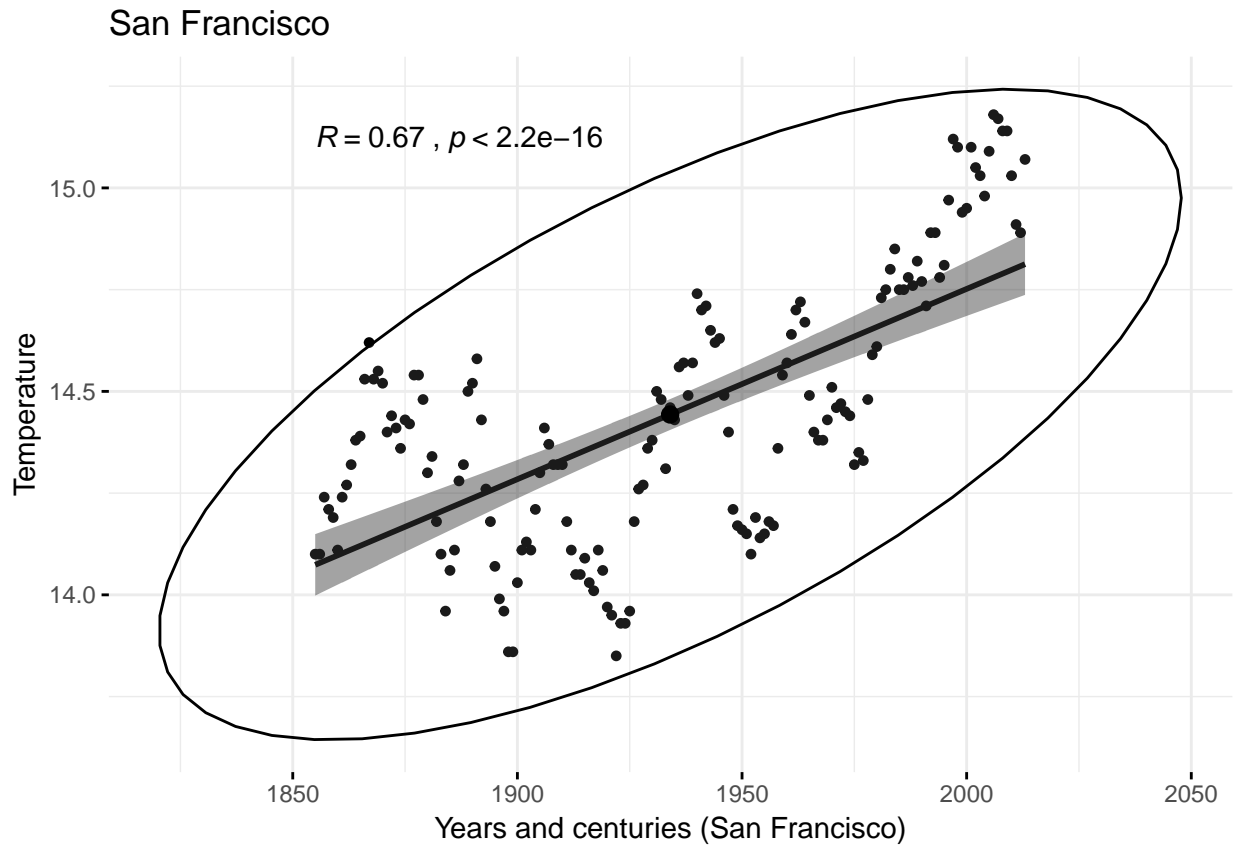
The correlation coefficient between the average temperatures in SF and the year has the value of 0.67. (1 is perfect correlation and 0 is none) When looking at the P level, we can see that it is much smaller than 0.5 (2.2e-16 - is the smallest number of system can show), which means that we can reject the null hypothesis and say that there is a very strong correlation between the years advancement and the rise in temperature in San Francisco.

The correlation coefficient between the years and the weekly average in Helsinki
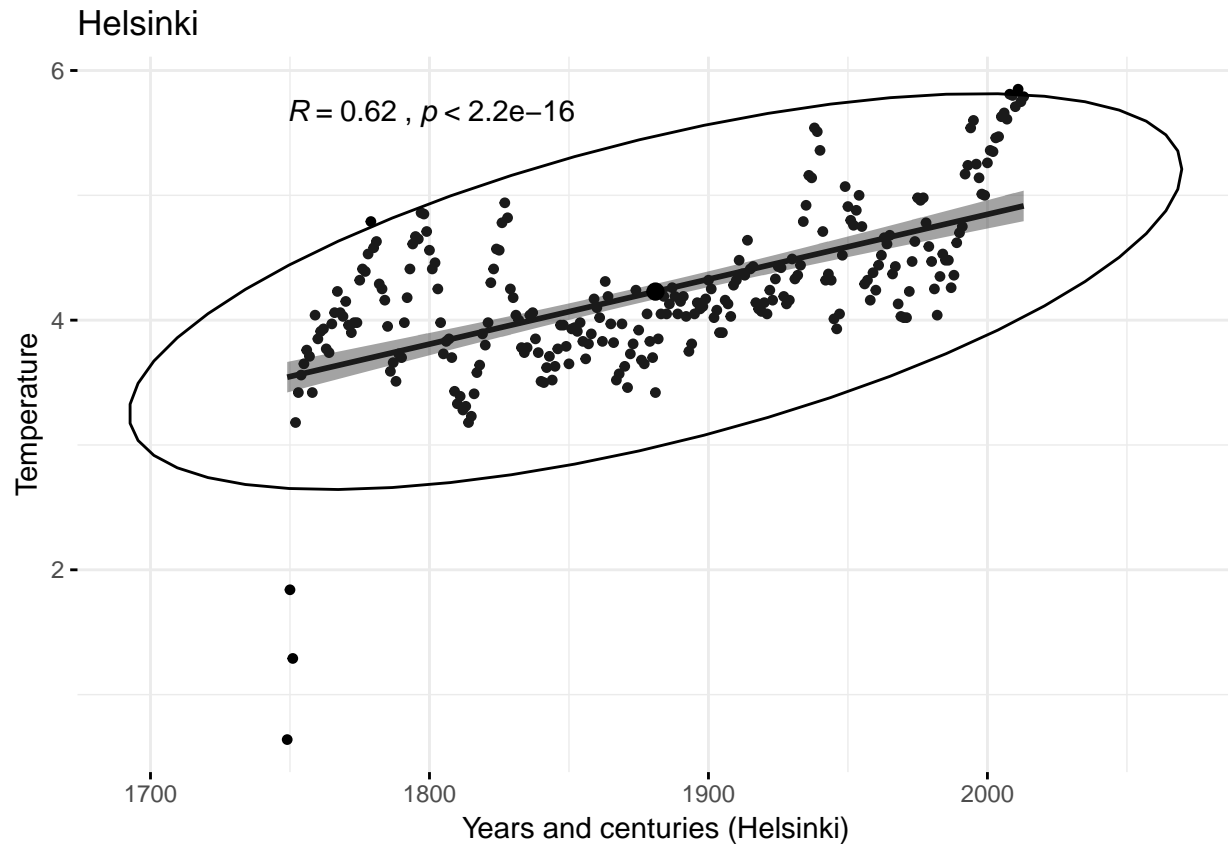

The change in the average weekly temperature in Helsinki from 1749 until 2013 was positive 5.2 degrees Celsius. It rose from a weekly average of 0.6 in 1749 to 5.9 degrees in 2013. Helsinki was taken as a Northern country to compare to San Francisco. Helsinki has a very similar pattern to the San Francisco regression line. There is a correlation between the years and the temperatures in this city and as the years go by the temperature increases exponentially. The regression line is not as steep as SF or Rio, but the p-value is practically 0 (2.2e-16), which tells us that the probability that the next year temperatures will rise in Helsinki is 99.999%.

Correlation coefficient between the years and the weekly average in Rio De Janeiro

Rio De Janeiro

$R = 0.9$, $p < 2.2e{-}16$
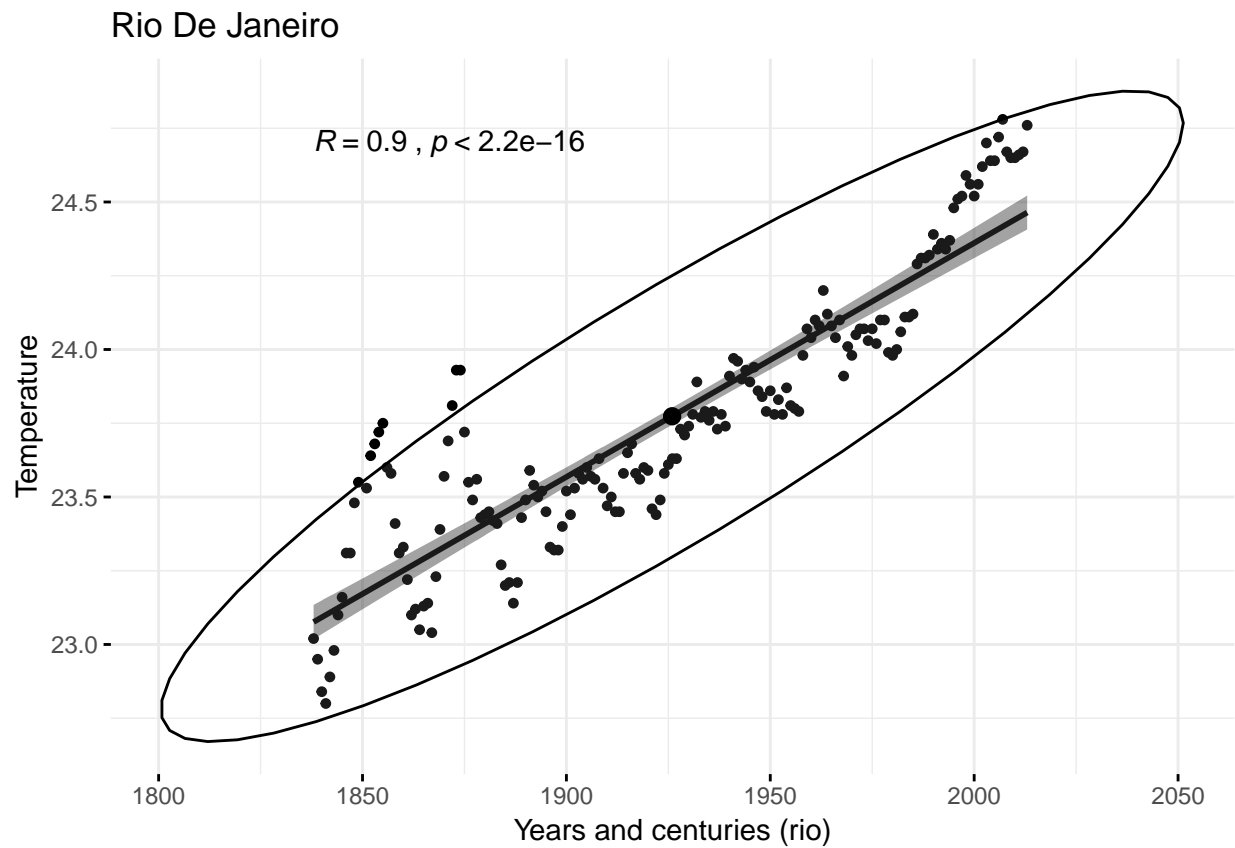
Rio De Janeiro was taken as a city from the Southern hemisphere. It has an almost perfect correlation ($R =$ 0.9) between the years and temperatures. The P value here is also almost 0, so we reject the null hypothesis and say that there is a very strong correlation here as well.

Correlation coefficient between the years and the weekly average in London

London

$R = 0.66$, $p < 2.2e{-}16$

Temperature

Years and centuries (london)
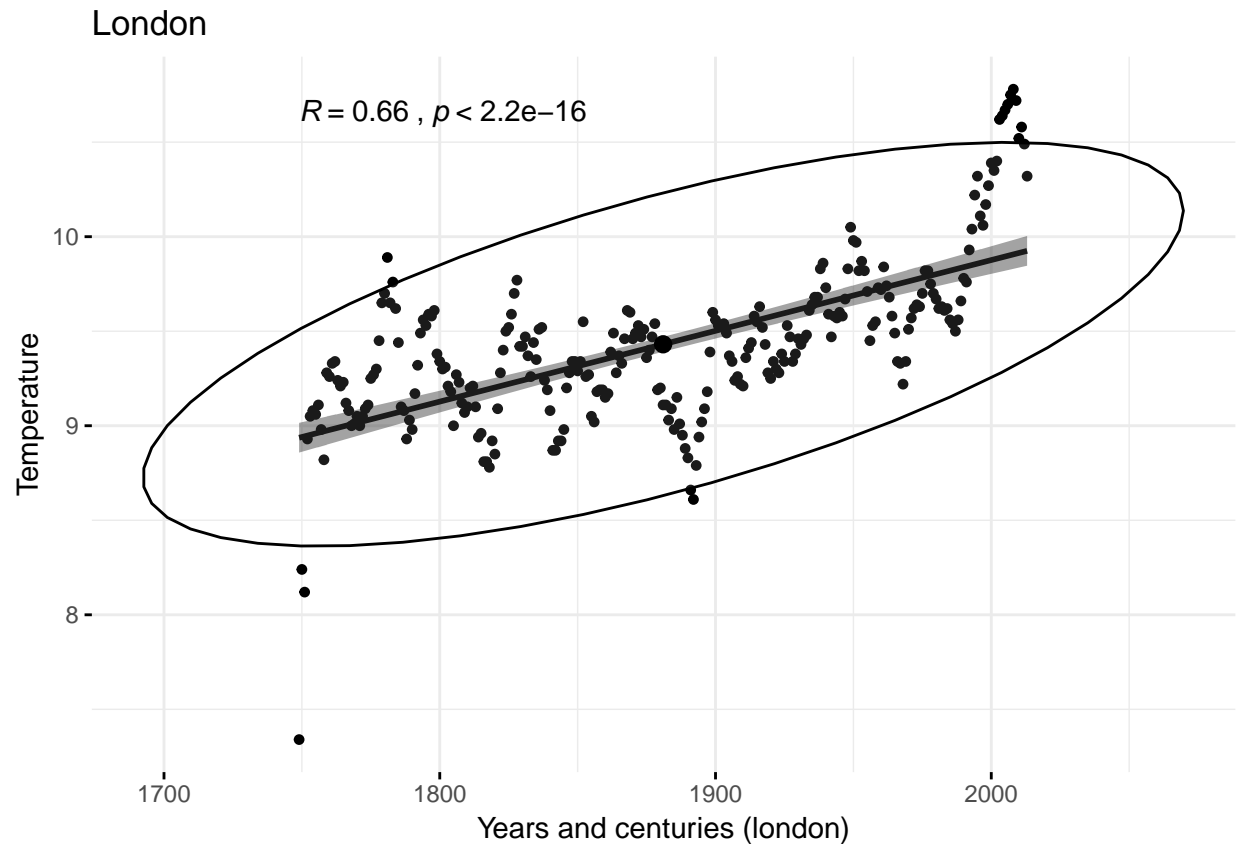
We can see above that there is a strong correlation between the years and the temperature in London, as with the previous cities. London was taken for its part in being the epicenter of the Industrial Revolution, which started in the 18th century. In the UK the Industrial Revolution during the 18th and 19th centuries was based on the use of coal. Industries were often located in towns and cities, and together with the burning of coal in homes for domestic heat, urban air pollution levels often reached very high levels. Scientists found that there is a strong correlation between air pollution and rising air temperature. So, coal pollution might have been the first reason for rising temperatures in London, as can be seen in the following chart.

How are the above look next to each other and compared to the Global temperature change throughout the years and centuries?
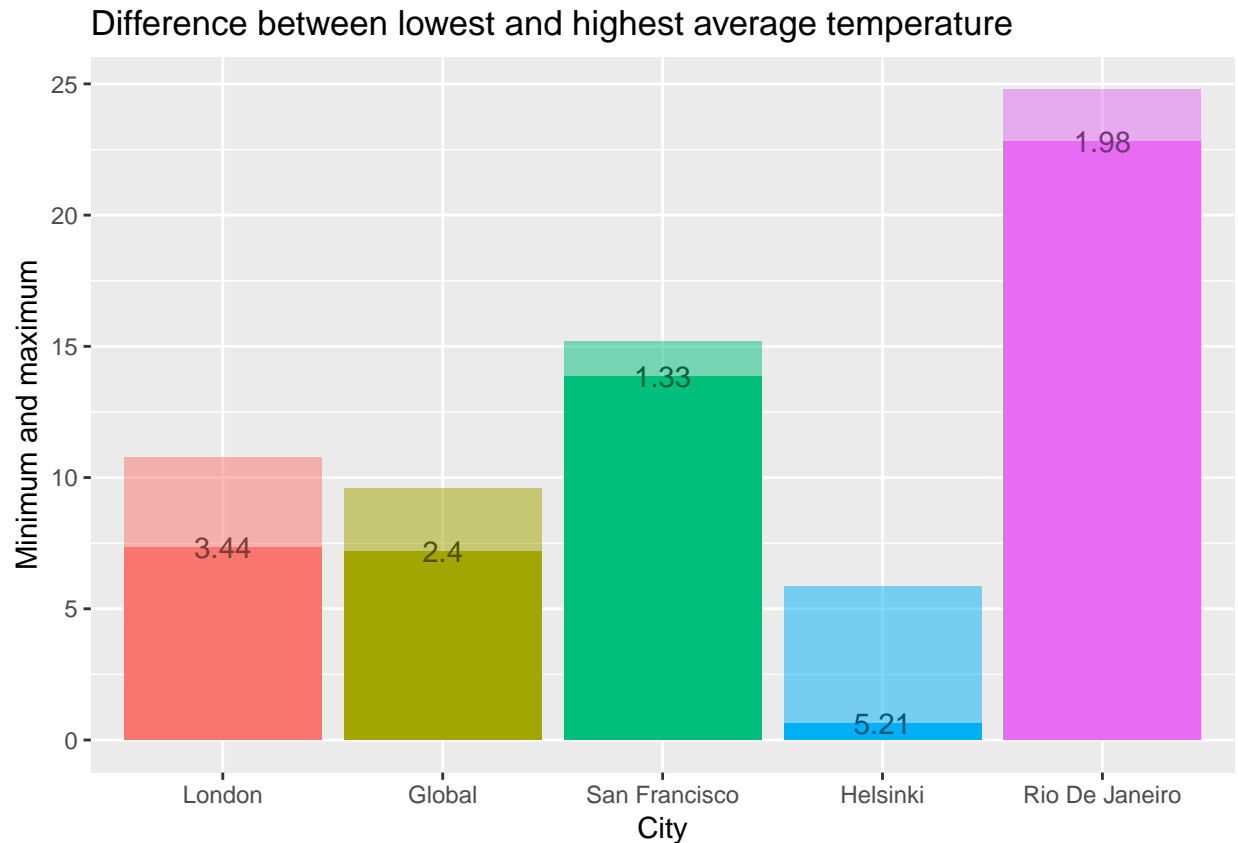
Cities VS Global weekly average temperatures
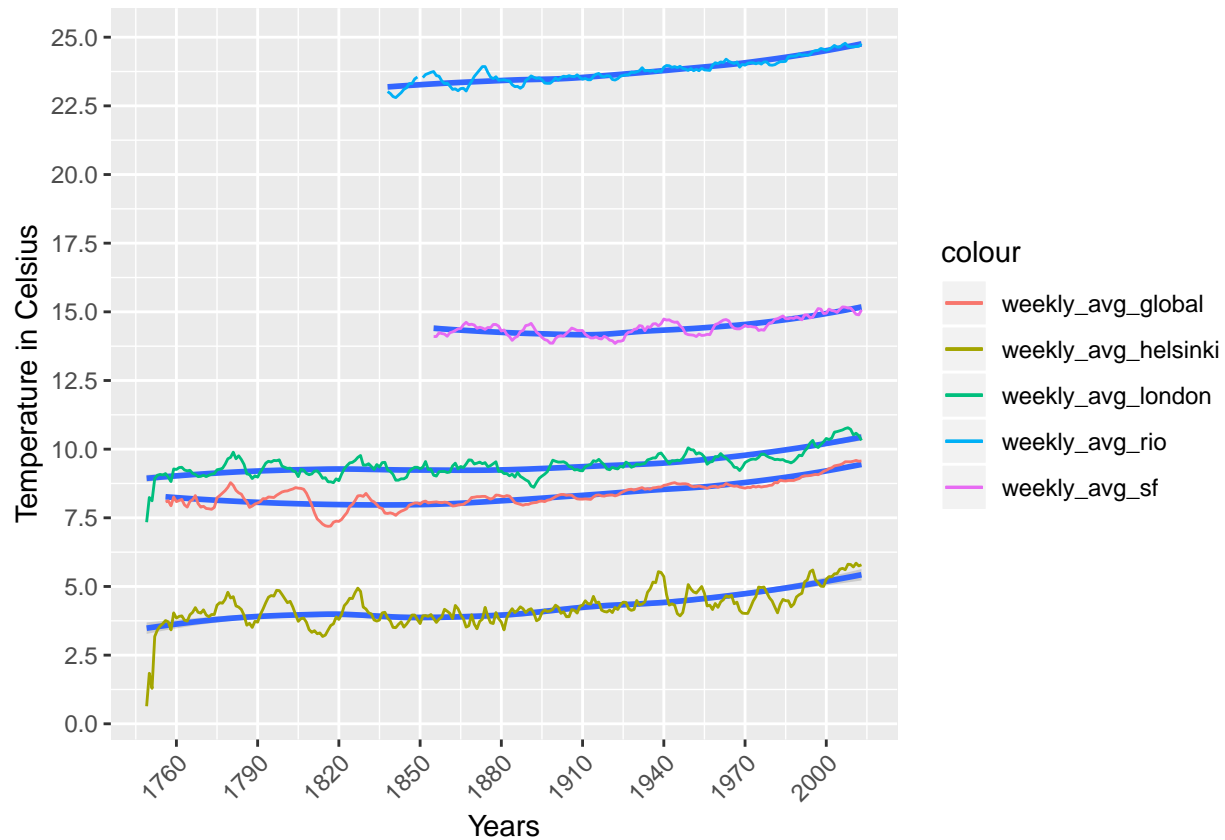
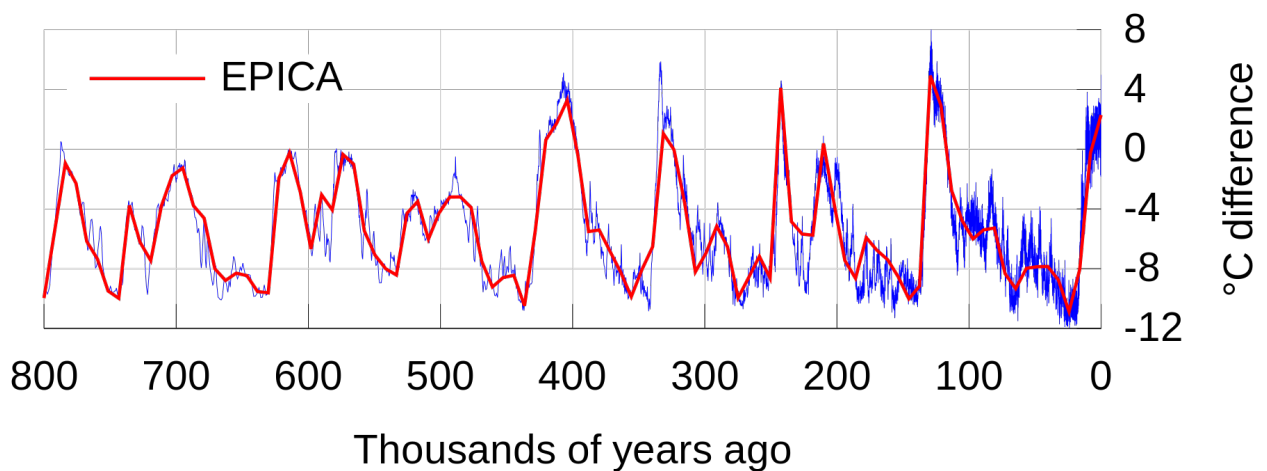
What can we see in the above chart? * San Francisco's average temperatures were higher, on average, 5.9 degrees than the Global average temperature for the same years. This trend seems to be consistent throughout the years and centuries. * The temperatures in SF went down very gradually from the beginning of records (1855) until 1913, when it started to rise. * Both London and Helsinki had a sharp increase in the temperature in the 18th century. * The Global temperatures, by this data, started to rise in the middle of the 18th century and rose since then. * Since 1913, temperatures in SF rose exponentially. * There seems to be a correlation between the years and the average temperatures, in all cases.

Cities VS Global weekly average temperatures

```
## # A tibble: 5 x 4
## # Groups:   City [5]
##   City              Max   Min  Diff
##   <fct>           <dbl> <dbl> <dbl>
## 1 London           10.8  7.34  3.44
## 2 Global            9.59  7.19  2.40
## 3 San Francisco    15.2  13.8   1.33
## 4 Helsinki          5.85  0.64  5.21
## 5 Rio De Janeiro   24.8  22.8   1.98
```

## Difference between lowest and highest average temperature

We can see above the difference between the minimum and maximum average temperatures in the 4 cities and global average. Helsinki experienced the biggest change in temperatures (5.21 degrees) since the beginning of records, followed by London with 3.44 degrees change since the beginning of the Industrial Revolution. Here are two external charts that show changes in global temperatures for the last thousand and 800 thousand years:

Source: Wikipedia
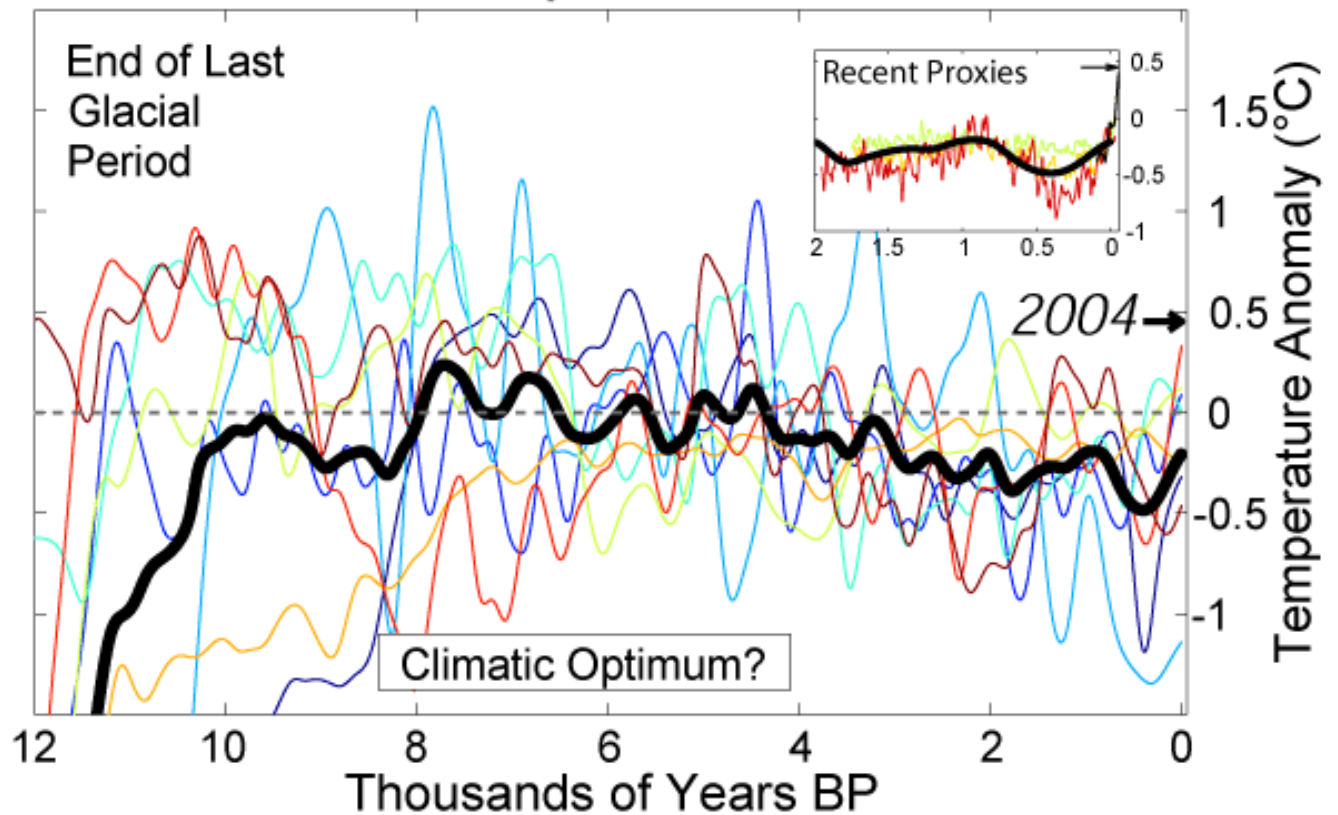
Source: Wikipedia

As can be seen in the above two charts, taken from Wikipedia, the trend that we have in our exploration here might very much fit the chart of the thousand years and of the ten thousand years. From this data, it seems that we are currently on a small heat wave of a couple of hundred years, and we are also on the hundred thosand year pick of a heat wave.

Conclusion

The Earth's atmosphere has been steadily and exponentially heating up in the last few centuries. This was verified with 4 different cities and with the given Global average temperatures in the above dataset. I used the Pearson correlation coefficient to find the strength of relationships between the years and the weekly average temperatures. An interesting point to find out in further research is why Helsinki had such a big increase in temperatures during the last 200 hundred years. Is it also related to the smog produced by coal in the 18th century, as was the case with London? Another interesting avenue to explore is why there was a decline in average temperatures in San Francisco in the late 19th century and the beginning of the 20st century?

Finally, We can expect to have higher temperatures, both locally and globally, if all the conditions that created the above trends remain the same, in the coming years and decades.