Weather Trends

By Svajune Klimasauskaite

17 March 2019

Goal

This project analyses local and global temperature data and compare the temperature trends where you live to overall global temperature trends.

Source data

The Data provided by Udacity was downloaded to CSV file and then uploaded into PostgreSQL.

The same tables as in Udacity were created in PostgreSQL and the following data checks were performed:

- to compare year values within different tables.
- to compare cities and countries within different tables.

EXCERPT operator was used for this.

```
select distinct(year) from city_data
except
select year from global_data
```

For the further analysis City and Global data were joined with OUTER JOIN operator to keep all the year values that were found in one table but not in another.

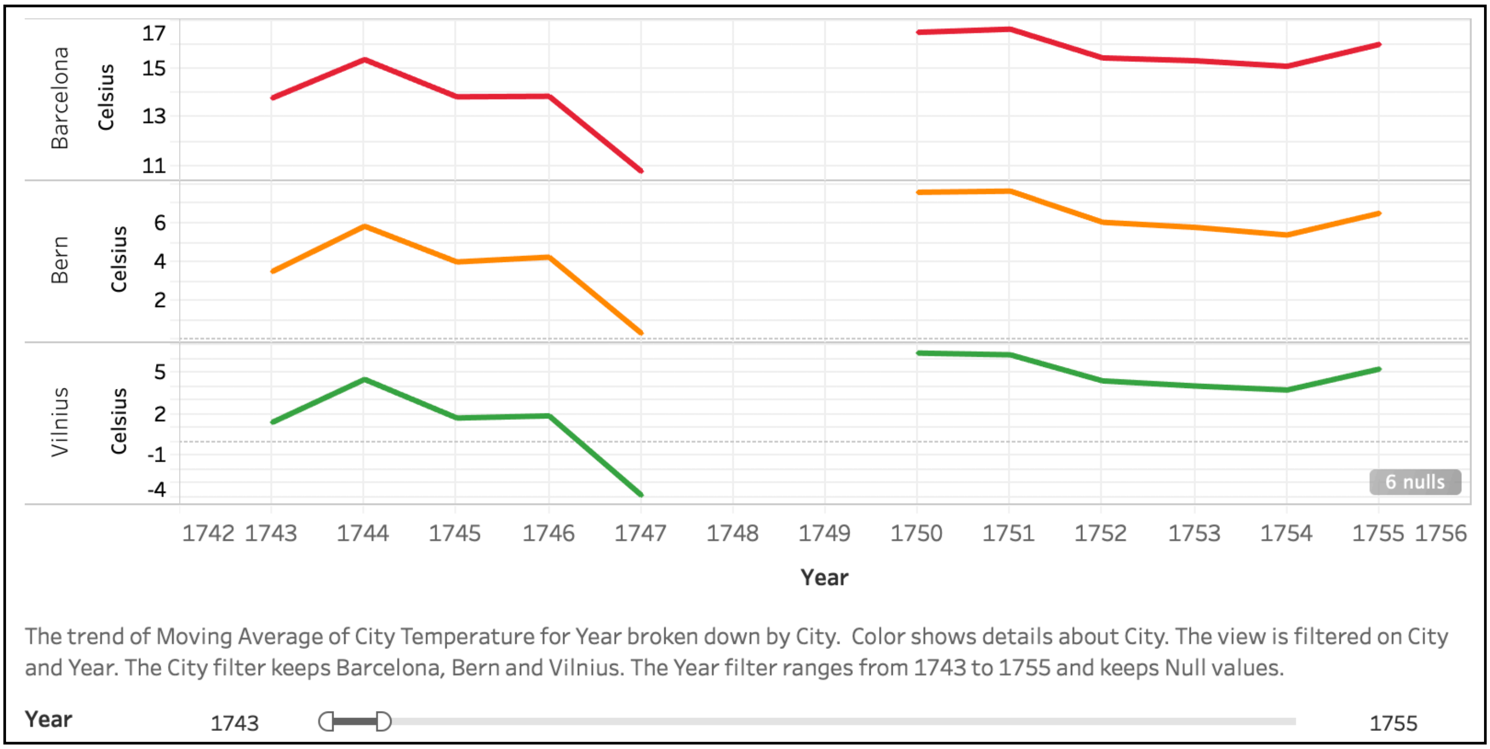
```
select
        case
                when aa.year = bb.year then aa.year
                when aa.year is null then bb.year
                else aa.year
        end as year,
        aa.city,
        aa.country,
        aa.city_tmp,
        bb.global_tmp
from city_data aa
full outer join global_data bb
on aa.year=bb.year
```

Visualisation with Tableau

The cities where I'm a local at, was chosen for the further analyses.

Great Frost in Europe around 1700-1750

Since all my selected cities were in Europe, the temperature around 1743 and 1747 shows the ending period of Great Europe's Frost. The next visualisations will filter out this anomaly in order not to fall into the scaling problem.
The years from 1743 to 1750 will be filtered out from visualisation, however still considered in moving average calculation and in the trend line.



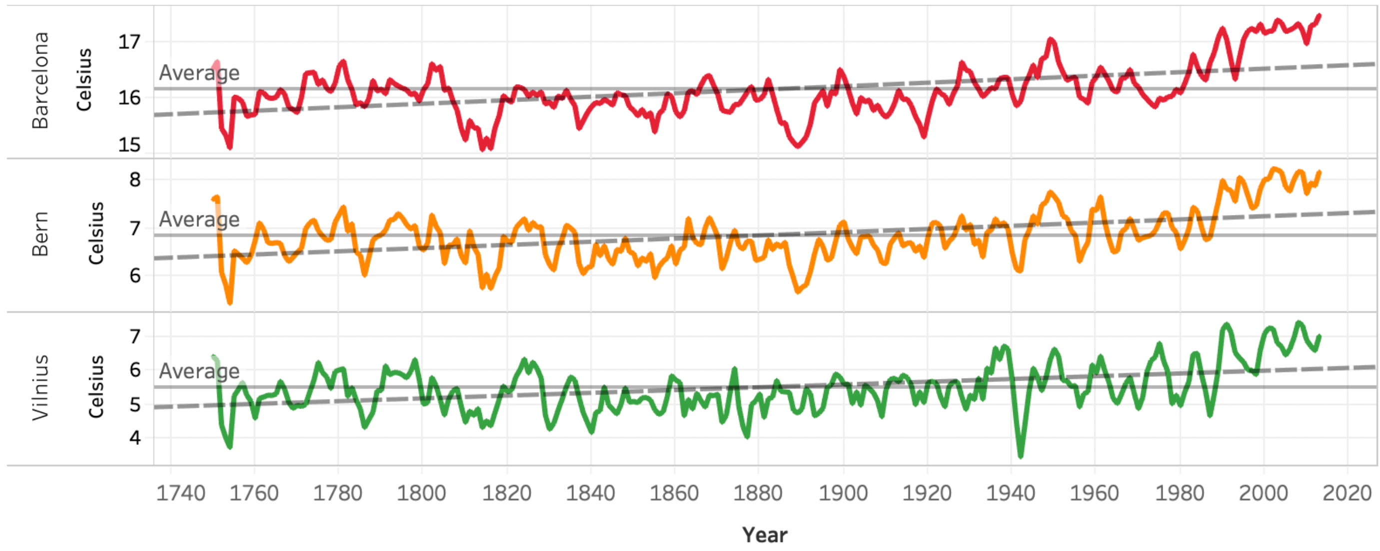
The trend of Moving Average of City Temperature for Year broken down by City. Color shows details about City. The view is filtered on City and Year. The City filter keeps Barcelona, Bern and Vilnius. The Year filter ranges from 1743 to 1755 and keeps Null values.

Average City Temperature

All 3 selected cities shares the same temperature characteristics:

- in the last ~50 years the temperatures shows warming tendencies.
- in the last ~30 years the different between the higher and lower temperature is shrinking. The higher and lower temperatures are getting closer. But the temperature continue increasing, it appears that the earth cannot cool down as before; what is called greenhouse effect.
- around 1980 and 1990 the temperature has crossed the average line and from then it hasn't fell down below the average.

The scale for each city is the same, but due to different temperatures, the Y-axis are not the same.



The trend of Moving Average of City Temperature for Year broken down by City. Color shows details about City. The view is filtered on City and Year. The City filter keeps Barcelona, Bern and Vilnius. The Year filter ranges from 1750 to 2013 and keeps Null values.

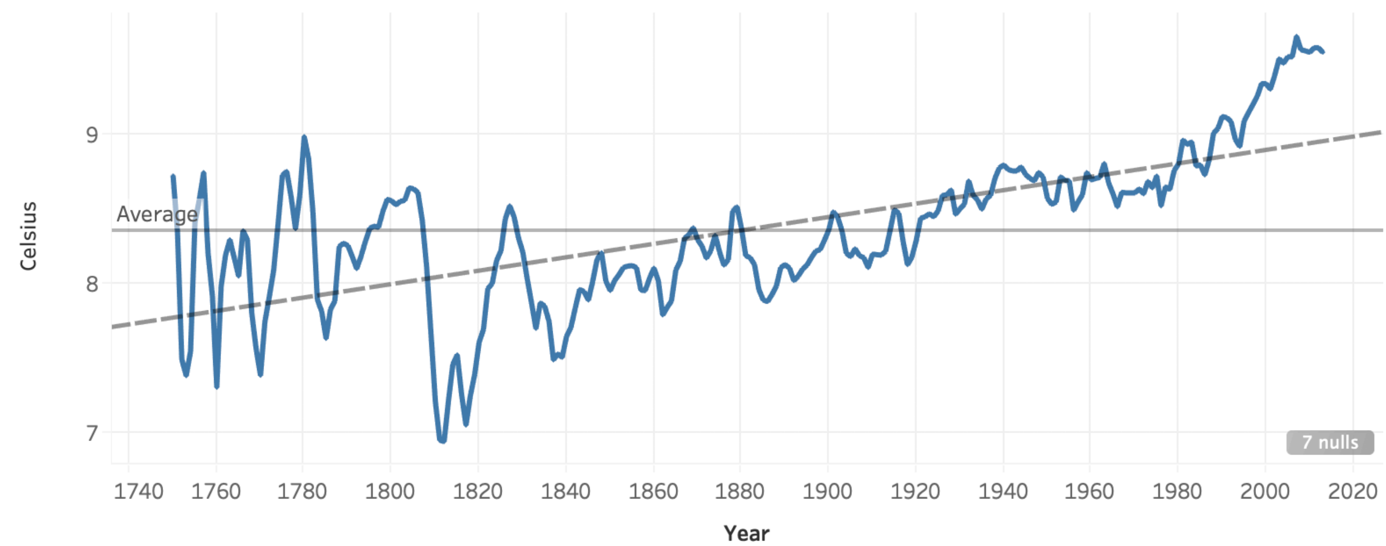
Year 1750 2013

Average Global Temperature

The global temperature shows the same tendency as the city temperatures:

- it shows warming tendencies,
- the difference between the higher and lower temperature is shrinking but the temperature continue increasing. However this appears to be much earlier in the global scale with the comparison of my selected cities.
- around 1920 the temperature has crossed the average line and from then it hasn't fell down below the average.

For the further analysis, it could be interesting to plot when the highest and lowest average temperature points will collide, to know the time point of no return for green house effect.

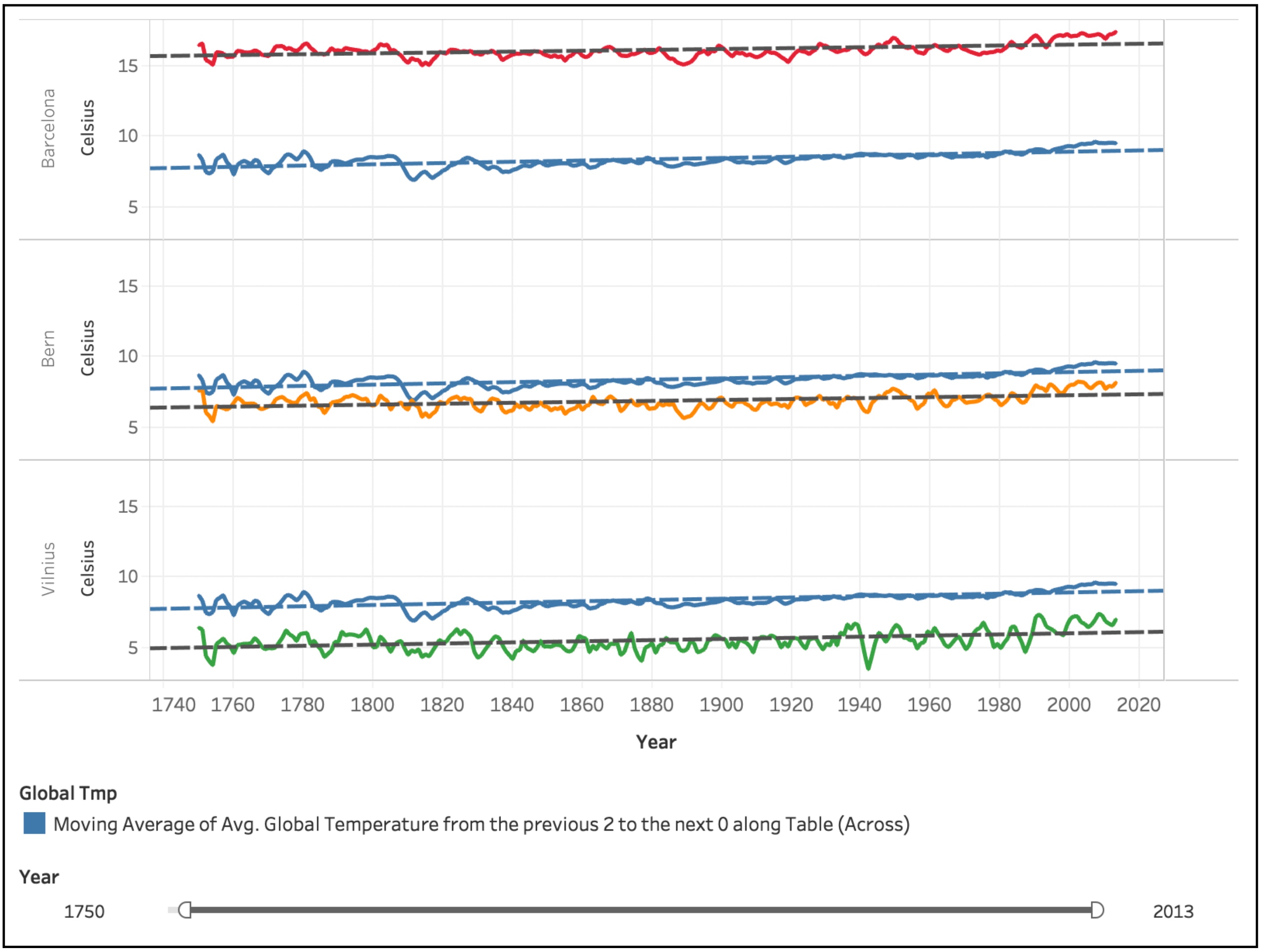


The trend of Moving Average of Avg. Global Temperature for Year. The view is filtered on Year, which ranges from 1743 to 2013 and keeps Null values.

Year 1743 2013

Cities and Global Trend Comparison

Independent wether the city temperature is above or below the global average temperature, the tendencies are very similar.



Correlation Coefficient

There are several types of correlation coefficient formulas. For the following analyses Pearson's correlation coefficient formula was applied.

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

The table on the right represents Correlation coefficient for several cities.
The low coefficient for Barcelona and Valencia forced me to check several hypothesis:

- wether this is related with the sea,
- wether this is related with the Mediterranean sea
- wether this is related with cold weather and heating.
- wether this is related to lack of data from these 2 cities.

However non of that was true. I haven't found any explanation for the low Barcelona and Valencia coefficient.

City temperature	Pearson's correlation coef.
Vilnius	**0.5666176**
Barcelona	**0.2467629**
Bern	**0.564615**
Valencia	0.2522597
Novosibirsk	0.6671488
Madrid	0.6842263
Rome	0.6611924
Athens	0.7132214
Oslo	0.4959659

The following scrips are shared:

- SQL to select city data,
- R script to calculate correlation coefficient.

```
select
		case
			when aa.year =
bb.year then aa.year
			when aa.year is null
then bb.year
			else aa.year
		end as year,
		aa.city_tmp,
		bb.global_tmp
from
city_data aa, global_data bb
where aa.year=bb.year and
city='Novosibirsk' and city_tmp is not null
and global_tmp is not null
```

```
data=read.csv(file.choose())
n=nrow(data)
xy=data[,2]*data[,3]
x2=data[,2]^2
y2=data[,3]^2
sumxy=sum(xy)
sumx=sum(data[,2])
sumx2=sum(x2)
sumy=sum(data[,3])
sumy2=sum(y2)
part1=n*sumxy-(sumx*sumy)
part2=(n*sumx2-sumx^2)*(n*sumy2-
sumy^2)
part3=sqrt(part2)
corr_coef=part1/part3
corr_coef
```

Observations

- Global temperature shows temperature changing characteristics much more earlier than my selected cities.
- Global temperature didn't cross below the average line from 1920, while my selected cities from 1980 and 1990.
- The differences between the highest and lowest points are shrinking with warming tendencies, however for global temperature it started around 1900, while for my selected cities around 2000.
- Correlation coefficient differs per city. Different cities has different correlation coefficient with Global temperature, however the reasons are not obviously clear. It could lead to a further investigation.