Exploring Weather Trends

By Vera Sticker

Goal

To analyze local and global trends in average yearly temperature, using a moving/rolling average and line chart visualization.

Steps Taken

1) Extract the data from the database using SQL. After examining the schema, I realized I would only need the year, temperature for my city (New York), and global temperature. I used the following query to extract the data:

```
1   SELECT c.year, c.avg_temp ny_temp, g.avg_temp
    global_temp
2   FROM city_data c
3   JOIN global_data g
4   ON c.year = g.year
5   WHERE c.city = 'New York'
6   ORDER BY 1;
```

   I then downloaded the resulting table to a CSV file named weather_data.csv.

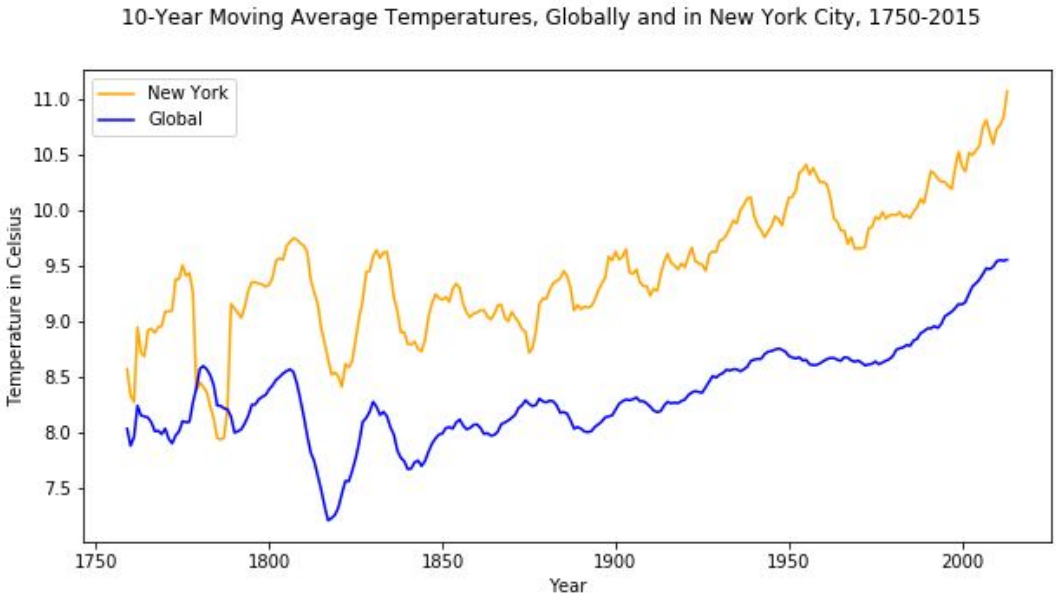
2) Analyze the data. I used Python for all analysis. My code can be found here:

   https://github.com/verascity/learning/tree/master/Udacity%20Data%20Analyst%20Nanodegree/Exploring%20Weather%20Trends

   First, I used the Pandas library in Python to read in and analyze weather_data.csv. I used Pandas' native bfill() function to backfill one missing year in the New York data,

then used the rolling().mean() function to calculate the 10-year moving averages for both the New York and global data.

3) I used the Python library Matplotlib to plot the two moving averages, resulting in the following line chart:



Observations

1) At first, I was surprised to realize that New York City, which I do not think of as particularly warm except on certain days in summer, is clearly warmer (except for one brief period in the late 1700s) than the global average. This is likely due to colder countries pulling the global average down.
2) Starting around the year 1800, the two lines move in very similar ways. While New York's temperature appears somewhat more volatile (the line is "bumpier" and has more "peaks"), they both tend to have peaks at the same times.
3) The New York and global temperatures both show a clear upward trend starting around the year 1900 and escalating sharply in the late 20th Century.
4) Both temperatures also show some interesting dips, including a small cooler spell just before the sharp rise around 1975, as well as a minor ice age in the early 1800s.

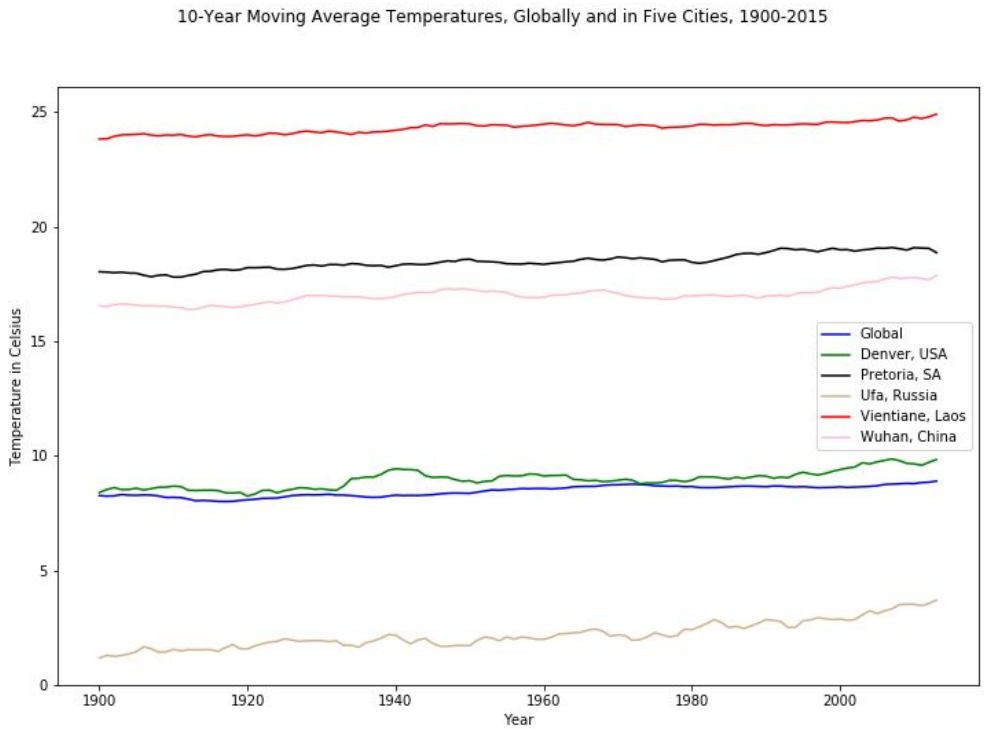
Follow-Up

After this, I was interested in other global trends. Would other cities show similar patterns? I returned to the SQL workspace and ran a new query to pull the data of five random cities:

```
SELECT c.year, c.city, c.country, c.avg_temp city_temp, g.avg_temp global_temp
FROM city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city IN (SELECT city
FROM city_list
ORDER BY random()
LIMIT 5);
```

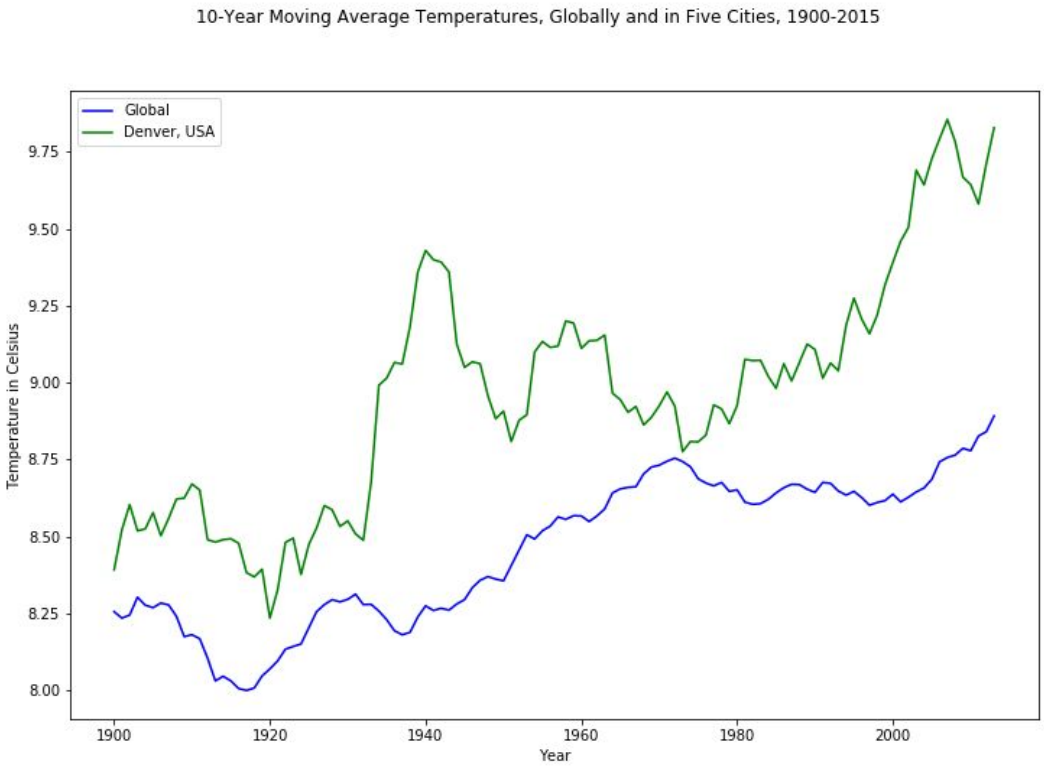
Using this query, I pulled Denver, United States; Pretoria, South Africa; Ufa, Russia; Vientiane, Laos; and Wuhan, China. I used Pandas' interpolate() method to fill in missing data and re-ran the moving average and plotting process.

Originally, I began with all years, but then realized that only the global and Denver data went that far back. I decided to focus on the 20th and 21st Centuries, since that was where the clear upward trend began in the previous data. I wanted to see if all the cities would follow the same trend. However, the graph I ended up with was not quite what I expected:



What happened to the sharp upswing I saw in my previous graph? Of course, I quickly realized that the wide range of temperatures had “flattened” my lines out. Looking at the previous graph, the global rise is actually only about one degree in total. It looks much more dramatic compared to a city with a similar, if higher, temperature, than cities such as Vientiane or Ufa with significantly higher or lower temperatures, respectively.

Rerunning the graph to show only the global and Denver lines plays this out:



So indeed, the upswings are still there, and while Denver does not hew as neatly to the global data as New York did, it still does follow the overall upward trend. So do the other four cities, to greater or lesser extents. One interesting thing I found was that the colder cities such as Denver and Ufa appear to have warmed more than the cities that were already hot, such as Vientiane and Wuhan. That relationship might be worth further investigation.