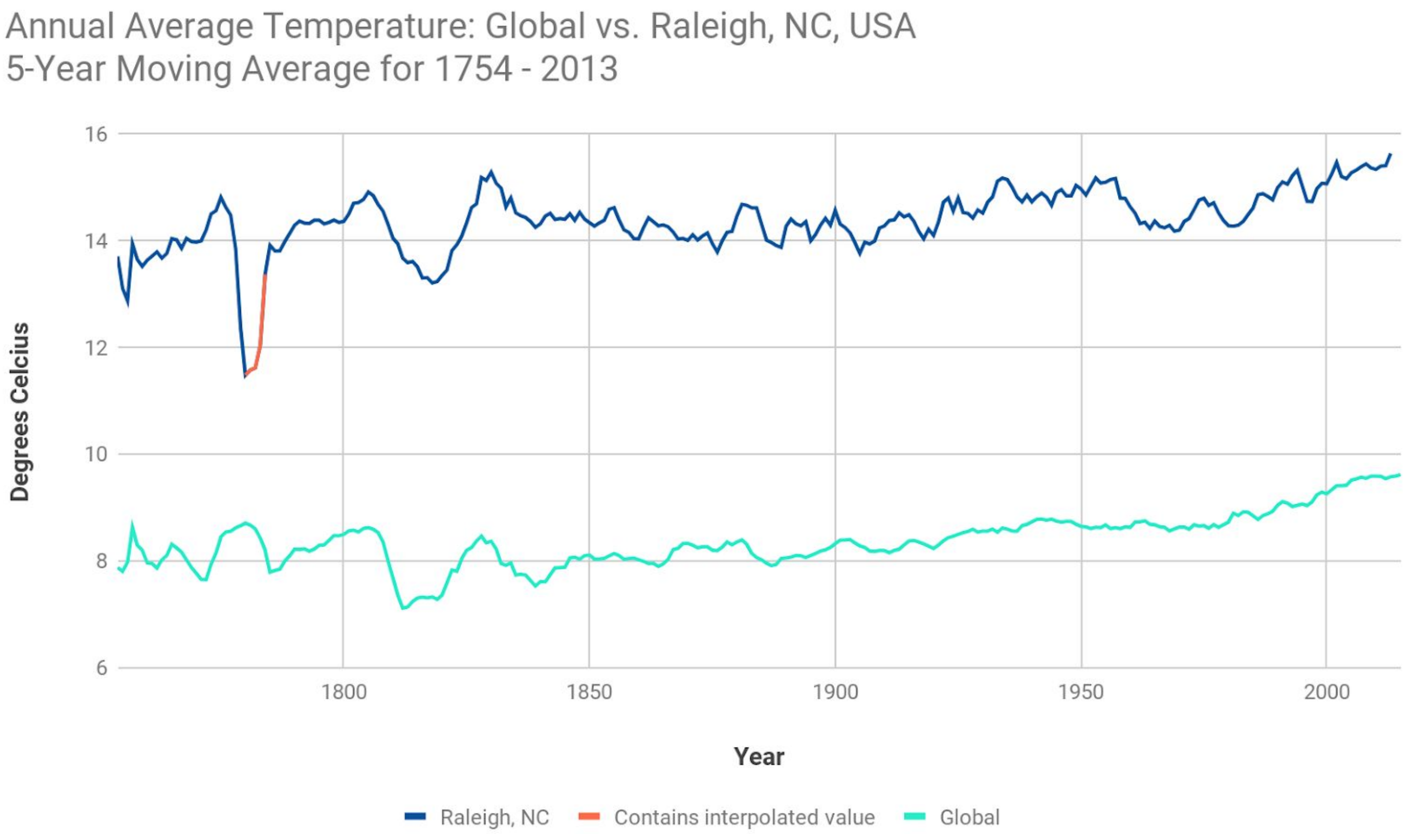
Project | **Explore Weather Trends**



Observations

1. The 5-year moving average of Raleigh, NC's annual temperature was warmer than the global moving average in every year from 1754-2013. On average, Raleigh was 6.01°C warmer than the global average from 1754-2013.

2. Both Raleigh, NC and the Earth were warmer in 2013 than in 1754. Raleigh's 5-year moving average annual temperature increased by 1.93°C from 1754 to 2013. The global 5-year moving average increased by 1.74°C over the same time period.

3. The overall trend for Raleigh, NC has varied more than the global trend, as one would expect. The chart shows the trend reversing direction several times for Raleigh, while the global trend more consistently rises slowly. The 5-year moving average temperature for Raleigh increased from the previous moving average in 135 of 258 years (not including 1754, which is the base year). Globally, the moving average increased in 145 of 258 years.

4. With one exception around the 1780s, Raleigh's moving-average appears to loosely coincide with the global trend. Investigating more closely, the correlation coefficient for Raleigh and the global average is 0.55, which shows there is indeed a moderately strong positive relationship between the two moving averages.

Project Steps

Extract the data

I used the SQL Workspace to extract the data for Raleigh, NC and the Earth. I extracted the data for Raleigh using the SQL query:

```
SELECT *
FROM city_data
WHERE city IN ('Raleigh');
```

I extracted the global data using the query:

```
SELECT *
FROM global_data;
```

Then I uploaded both CSV files to Google Sheets.

Calculate the moving average

After inspecting the data, I noticed that the global data were available for 1750-2015 and the Raleigh data were available for 1743-2013. So in order to compare the two variables, I had to limit my observations to the years 1750-2013. I also noticed that Raleigh was missing a temperature observation for the year 1780. Not having a temperature for 1780 would affect the moving average for several other years, so I knew I needed to determine a way to replace the missing value. I decided to replace the missing value with the average of the observations for 1779 and 1781. This made sense to me, since the temperature data appeared to gradually move from one year to the next.

Then, I needed to determine how many years to take a moving average across. I did some research and discovered a NASA publication[1] that displayed 5-year moving averages for "Global Annual Mean Surface Air Temperature Change." So I followed NASA's example and calculated a 5-year moving average. To calculate the moving average, I took the average of the year 1750-1754 to get the moving average for 1754. The I copied the formula for the rest of the years through 2013.

I calculated the Pearson's correlation for the data for Raleigh and the Earth using the CORREL function after consulting the Google Sheet documentation.[2] I consulted a publication[3] from the National Center for Biotechnology Information to refresh myself on interpreting the correlation coefficient.

As mentioned in Observation #3, I calculated the number of years in which the moving average increased from the previous years moving average, both for Raleigh and the Earth. I did this by subtracting the moving average for 1754 from the moving average for 1755 and copying the formula. Then I used the COUNTA function[4] to count the number of years in which the difference between that year and the previous year's moving average temperature was greater than zero.

---

[1] https://data.giss.nasa.gov/gistemp/graphs_v3/
[2] https://support.google.com/docs/answer/3093990?hl=en
[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3576830/
[4] https://support.google.com/docs/answer/3093991?hl=en

Design the chart

My main considerations when deciding how to visualize the trends were:

1. Ease of comparing the data
2. Depicting the data accurately
3. Highlighting the values that include the interpolated observation for Raleigh in 1780.
4. Design for color blindness

To achieve #1 and #2, I created a stacked line chart that displays the data for Raleigh directly above the global data. This allows the viewer to quickly identify where the trends match and differ. The stacked line chart intuitively displays the interrelated nature of the data, since Raleigh is a subset of the Earth. I chose to include vertical gridlines in 50 year increments to aid the viewer in indexing the dataset without making the graph visually busy. I included horizontal gridlines in 2° increments to maximize the display size of the dataset and help the viewer index the data.

To achieve #3, I researched how to achieve this in Google Sheets and found a guide from Ben Collins.[5]

To achieve #4, I consulted the material, 'Designing for color blindness' in the Udacity Data Foundations Nanodegree Program. I researched color blindness-friendly color palettes and found a helpful resource[6] from Martin Krzywinski. I chose a color palette[7] optimized for viewers with deuteranopia, the most prevalent form of color blindness.

---

[5] https://www.benlcollins.com/formula-examples/data-labels/
[6] http://mkweb.bcgsc.ca/colorblind/
[7] http://mkweb.bcgsc.ca/colorblind/img/colorblindness.palettes.png