Project 1 - Exploring Weather Trends

1 Introduction

In this project report, the closest big city and global weather data are extracted from a database. This report walks through the steps taken to extract, analyze and visualize the data. The trends are then compared and summarized.

2 Data Preparation

2.1 Data Extraction

In order to extract the weather data for a big city closest to me, I ran the below query to extract the names of the cities in US from the list.

```
SELECT * FROM city_list
WHERE country='United States'
```

I chose Raleigh from the list of cities and executed the below query to get more information about the weather data for Raleigh city from city_data table. I repeated the process to get the data for other cities like Paris, Shanghai, and Moscow.

```
SELECT year,avg_temp FROM city_data
WHERE city='Raleigh'
```

I was able to extract global weather data from global_data table using the command below.

```
SELECT * FROM global_data
```

2.2 Data Manipulation

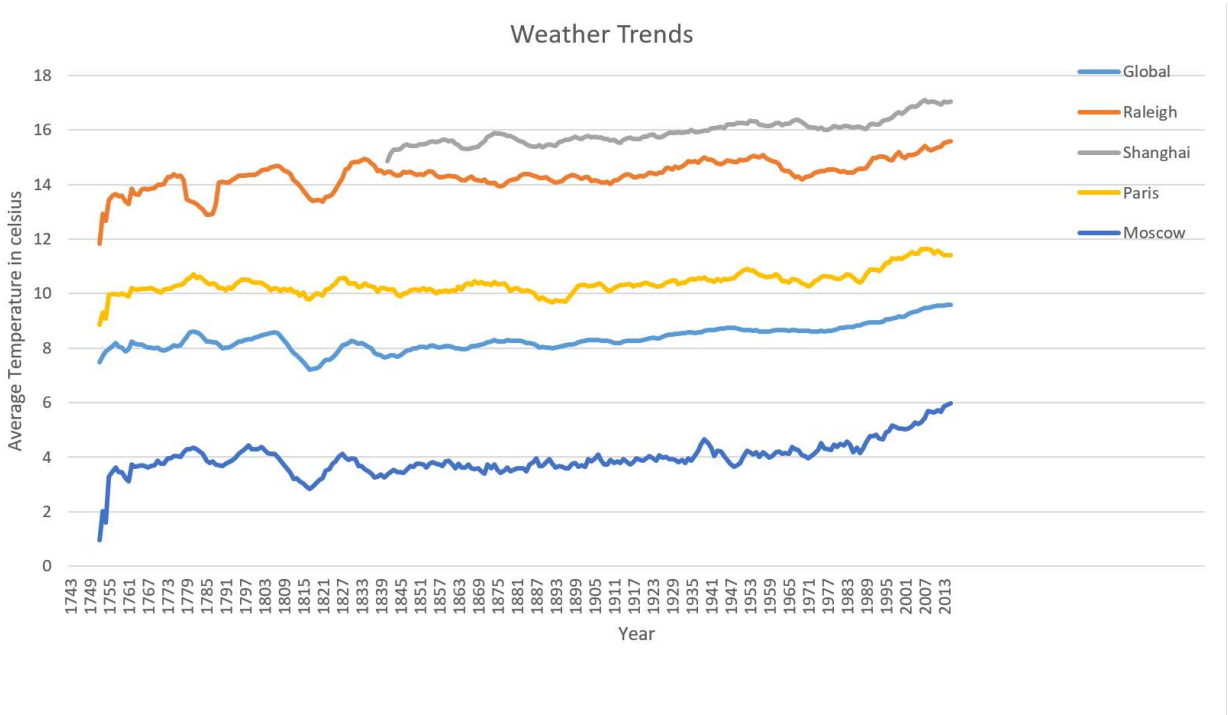
I decided to use Excel for analyzing and visualizing data due to the ease of use. I downloaded the csv files for each of the result set above and combined them in to a single sheet to get a comprehensive year and average temperature data set with columns shown below.

Year	Avg_temp_Global	Avg_temp_Raleigh	Avg_temp_Shanghai	Avg_temp_Paris	Avg_temp_Moscow

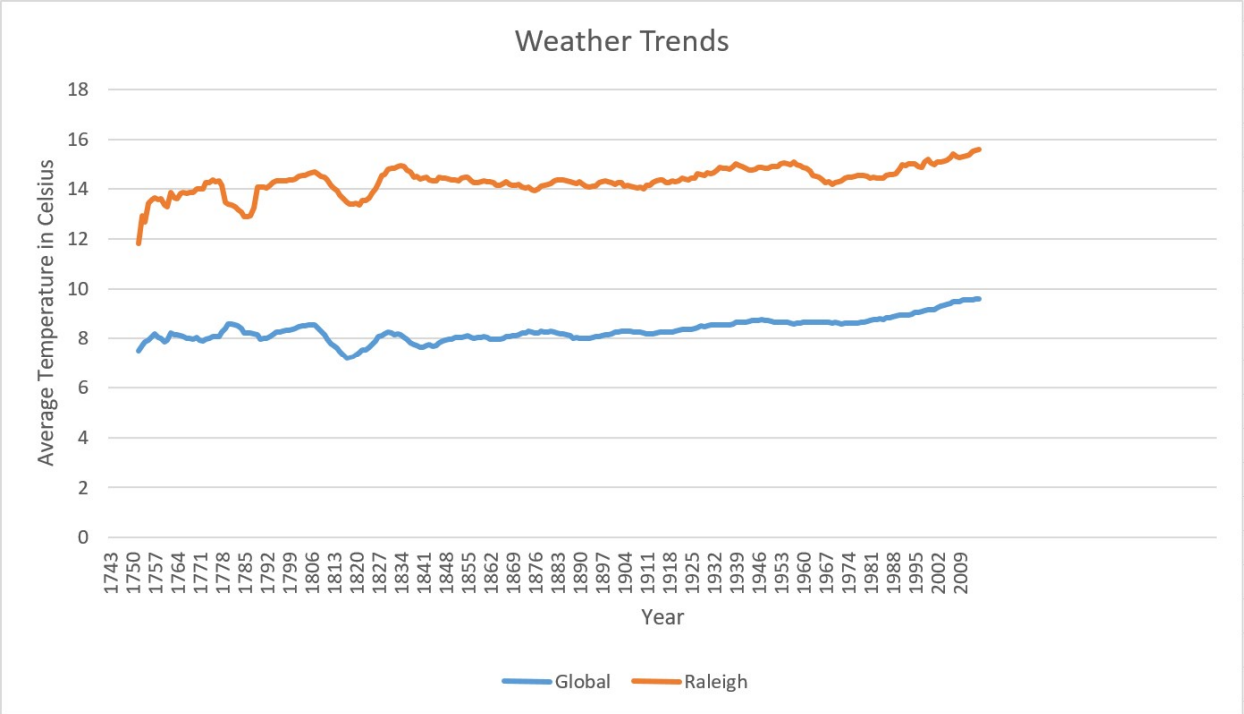
3 Data Analysis and Visualization

To analyze the weather trend over few hundred years, I have computed moving averages for each of the cities over a period of 10 years. The 10 year moving averages allows me to smooth out the noise and make the trends more observable.

I use Excel charts to plot a line chart of the Moving average temperatures of various cities along with the global temperatures.



Below is a line chart of just Raleigh vs Global moving average temperature.



In addition to the visual insights, I have calculated the correlation coefficients between the cities and global weather data to provide insight into how the weather data of various cities correlate with the global data. This helps us know if the city weather data trends are in line with the global trends. I have used the data analysis tool in Excel instead of CORREL function.

Based on the correlation coefficients below, it can be concluded that the city weather data is strongly correlated to global data.

L	M	N	O	P	Q
	Global	***Raleigh***	***Shanghai***	***Paris***	***Moscow***
Global	1				
Raleigh	0.686386	1			
Shanghai	0.969461	0.807036	1		
Paris	0.863587	0.682832	0.907256	1	
Moscow	0.870488	0.695547	0.883534	0.894038	1

4 Observations

The following observations can be made from the line charts in the previous section.

1. Raleigh has higher average temperature than Global temperature.
2. Raleigh follows the same trend as global temperature.
3. The global temperatures have been on a consistent increasing trend over the last several years.
4. 1816 seems to be a year with weather abnormalities with considerable drop in average temperatures across various cities as evident by the dip in the line charts.