Explore Weather Trends

Project 1: Udacity Data Analyst Nanodegree

Patricia Putnam

August 21, 2022

Purpose

The goal of this project is to compare and analyze trends in temperature data between data between Seattle, WA, USA, and overall global averages. The intent is to have a better understanding of how Seattle's local trends relate the overall global trends.

Process

This section describes the process and tools used to examine the data and generate comparisons.

Tools

SQL: used to query and extract the data needed from the provided dataset

Excel: used to import data, generate descriptive analysis of data, and generate visualizations

Data extraction

The first step was to understand the available data and extract what was needed from the database.

1. Confirm the availability of data for Seattle by running the following SQL query against the city_list table:

```
SELECT *
FROM city_list
WHERE city_list.city = 'Seattle'
```

2. Examine the schema for city_data and global_data tables. Observe that both have columns named 'avg_temp'. This will need to be adjusted to provide clear indication of the scope of the data as part of the data extraction, which was handled with a single SQL query:

```
ALTER TABLE global_data RENAME COLUMN avg_temp to
Global_avg_temp;
ALTER TABLE city_data RENAME COLUMN avg_temp to Seattle_avg_temp;

SELECT
    city_data.year,
    city_data.city_avg_temp,
    global_data.global_avg_temp
FROM city_data
INNER JOIN global_data
ON city_data.year = global_data.year
WHERE city_data.city = 'Seattle'
```

3. Save results to file Avg_Temp_Compare.csv

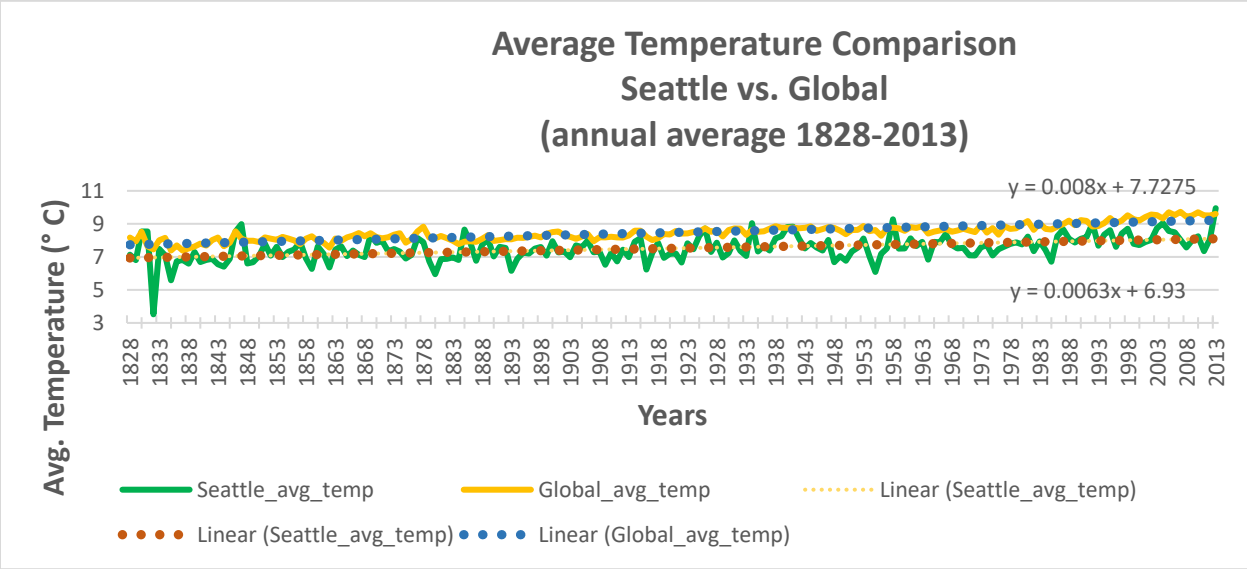
Import and Clean Data

To generate meaningful observations, the data must first be cleaned.

4. Import Avg_Temp_Compare.csv into Excel.
5. Observe that there are missing data points in the Seattle_avg_temp column for years 1830, 1831 and 1846. Rather than simply removing those rows from the analysis, I opted to average the entire column and use the average (8.54) to replace the missing data.

Generate Initial Visualization

6. Using Excel to generate a visual representation of the data resulted in the following line chart with trend lines included:



Generate Descriptive Statistics

7. Excel offers functions to generate descriptive statistics. The following functions were used:
    a. AVERAGE
    b. MIN
    c. MAX
    d. STDEV.P
    e. CORREL

f.

8. Results of the functions:

Average of Seattle_avg_temp	Average of Global_avg_temp
7.52	8.48
Min of Seattle_avg_temp	**Min of Global_avg_temp**
3.52	7.38
Max of Seattle_avg_temp	**Max of Global_avg_temp**
9.95	9.73
StdDev of Seattle_avg_temp	**StdDev of Global_avg_temp**
0.74	0.50

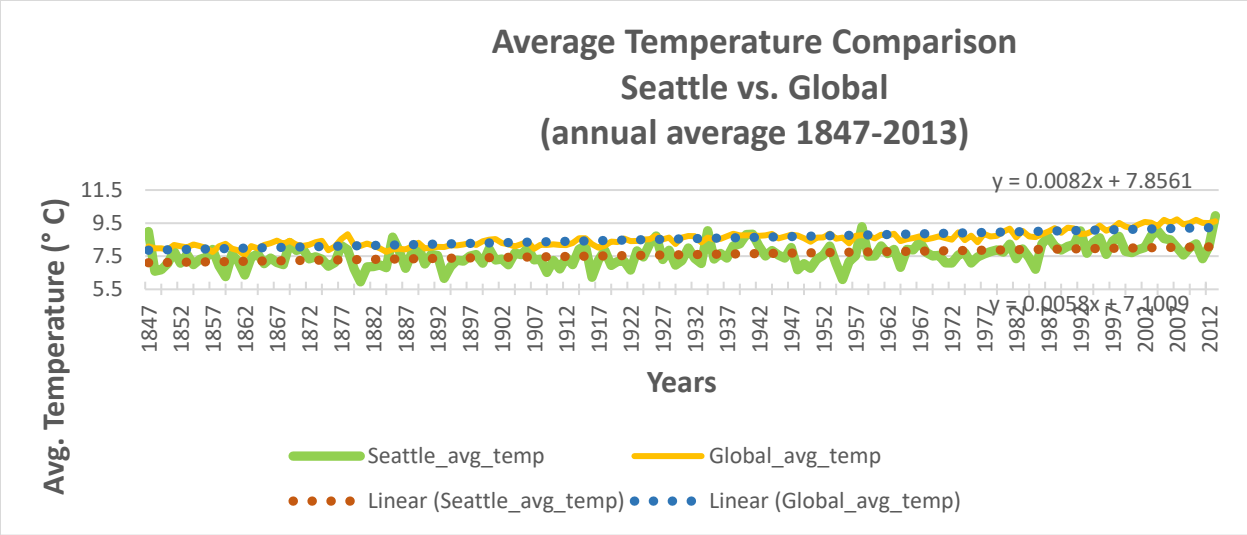
Correlation Coefficient: Seattle and Global
0.592565991

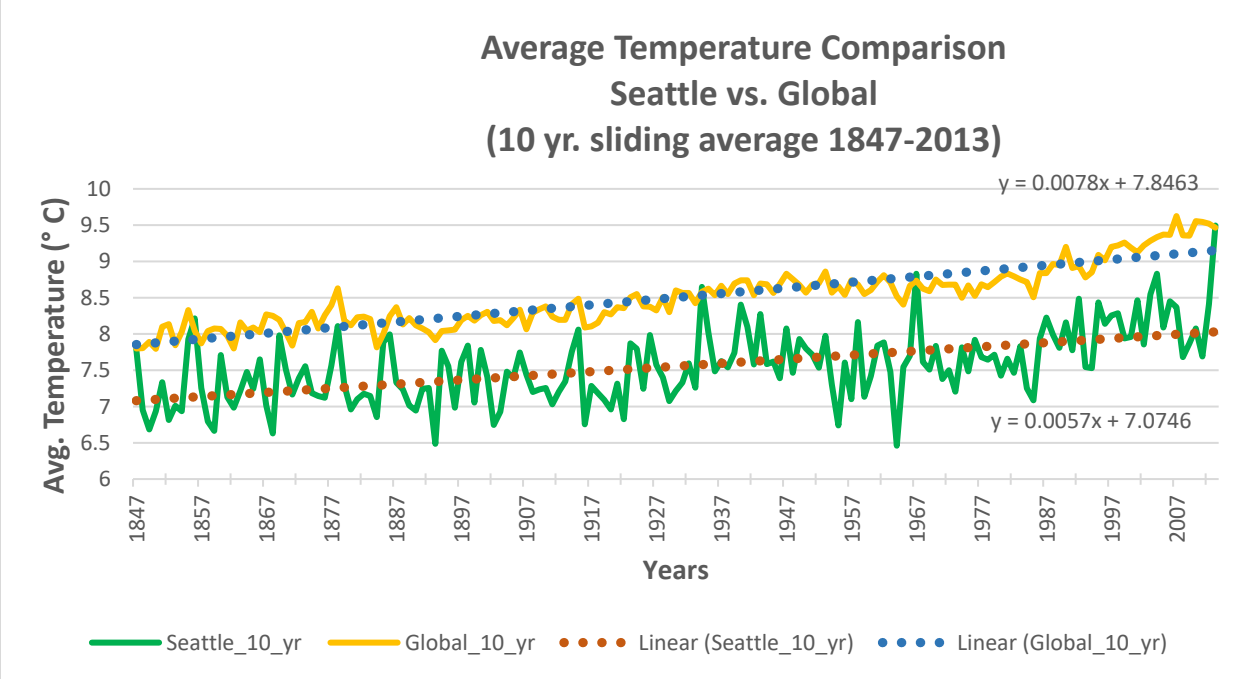
Observations

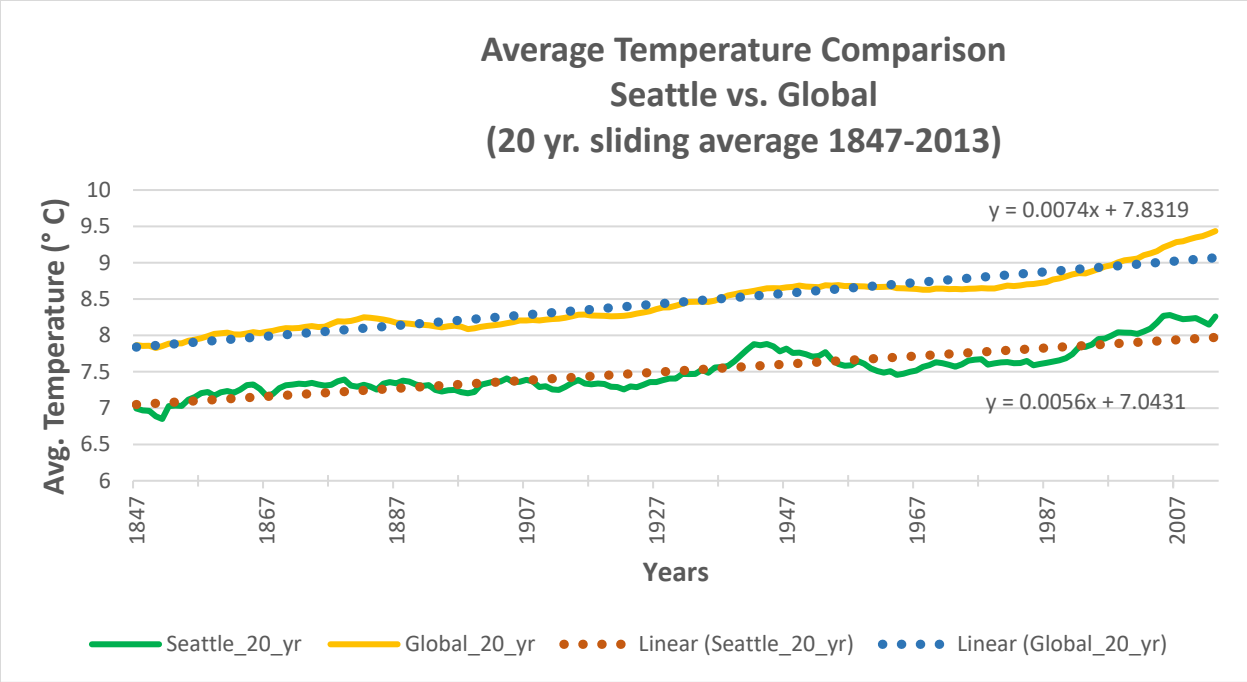
1. The overall comparison of average temperatures indicates that Seattle averaged nearly 1 degree lower temperatures than the global average.
2. Seattle has some noticeable minimum and maximum temperature spikes compared to the global average. Taking those into account, the range from minimum to maximum for Seattle is 6.43 degrees, vs Global range from minimum to maximum in the same time period is 2.35 degrees.
3. Examining the trend lines slope for Seattle and Global averages, the Global average temperature is rising more quickly than Seattle's average over this time range.
4. Looking at the correlation coefficient, there is a strong, but not perfect correlation between Global average temperatures and Seattle average temperatures. This aligns with the general positive increase of the trend lines for both.

Additional Visualizations of Moving Average data

After running the initial analysis, I decided I wanted to do a comparative visual representation using moving averages of over 10 years and 20 years to smooth the data and resulting lines. I truncated the visualizations to start at 1847 to account for the 20 year moving averages. Here are the visualizations for that effort:


Average Temperature Comparison
Seattle vs. Global
(annual average 1847-2013)
y = 0.0082x + 7.8561
y = 0.0058x + 7.1009
Avg. Temperature (° C)
11.5
9.5
7.5
5.5
1847 1852 1857 1862 1867 1872 1877 1882 1887 1892 1897 1902 1907 1912 1917 1922 1927 1932 1937 1942 1947 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007 2012
Years
Seattle_avg_temp
Global_avg_temp
Linear (Seattle_avg_temp)
Linear (Global_avg_temp)


Average Temperature Comparison
Seattle vs. Global
(10 yr. sliding average 1847-2013)
y = 0.0078x + 7.8463
y = 0.0057x + 7.0746
Avg. Temperature (° C)
10
9.5
9
8.5
8
7.5
7
6.5
6
1847 1857 1867 1877 1887 1897 1907 1917 1927 1937 1947 1957 1967 1977 1987 1997 2007
Years
Seattle_10_yr
Global_10_yr
Linear (Seattle_10_yr)
Linear (Global_10_yr)



As can be seen, the smoother lines with the 20 year moving average makes it easier to assess general trends, but some of the interesting outliers are removed in the process.