Analyze Weather Trends Project

Udacity Data Analyst Nanodegree

Submitted by: Justin Brown (jb669n@att.com)

Outline of Steps Taken

1) Extracted Global Trends using SQL and downloaded to global_trends.csv
    a. SQL Used
        i. Select * from global_data
2) Extracted US Based Cities using SQL and downloaded to US_Cities.csv
    a. SQL Used
        i. Select * from city_list where country = 'United States'
3) Identified Dallas as the closest city by reviewing US_Cities.csv in Excel.
4) Extracted Dallas Trends using SQL and downloaded to Dallas_trends.csv
    a. SQL Used
        i. Select * from city_data where city = 'Dallas'
5) Imported all .csv files into Analyze_Weather_Trends.xlsx
6) Created merged_trends tab in Excel by placing Global and Dallas data side by side.
    a. Aligned years, noting that Dallas data is only available from 1820 forward.
    b. Noted that Global data is available for 2 additional recent years (2014, 2015) to exclude from visualizations.
7) Created Error_Check column to compare years in global dataset and Dallas dataset to quickly ensure no anomalies existed within the years captured from 1820-2013. (None noted.)
8) Created columns to convert Celsius data to Fahrenheit (for users not familiar with Celsius) and calculated the 10-year and 50 year moving averages of both values for the Global and Dallas Data.
9) Created merged_trends tabs to group like data in preparation for charting. Deleted rows where data does not exist for both datasets.
10) Created comparative charts for Global vs. Dallas trends, first visualizing the raw averages, then visualizing the 10-year and 50-year moving averages (for both Celsius and Fahrenheit).
11) Upon review, the Celsius 10-year moving average data displays the best balance for analysis (less volatility than yearly, more years than 50-year moving average to see more trending information).
12) Created additional chart with a shifted secondary axis to overlay the two datasets.
13) Calculated Actual Difference and % Difference for each dataset using both the Min/Max of each as well as the earliest and latest year for which data is available.

The Data

Global vs. Dallas - 10 Year Moving Average Data

Global vs. Dallas - 10 Year Moving Average Data
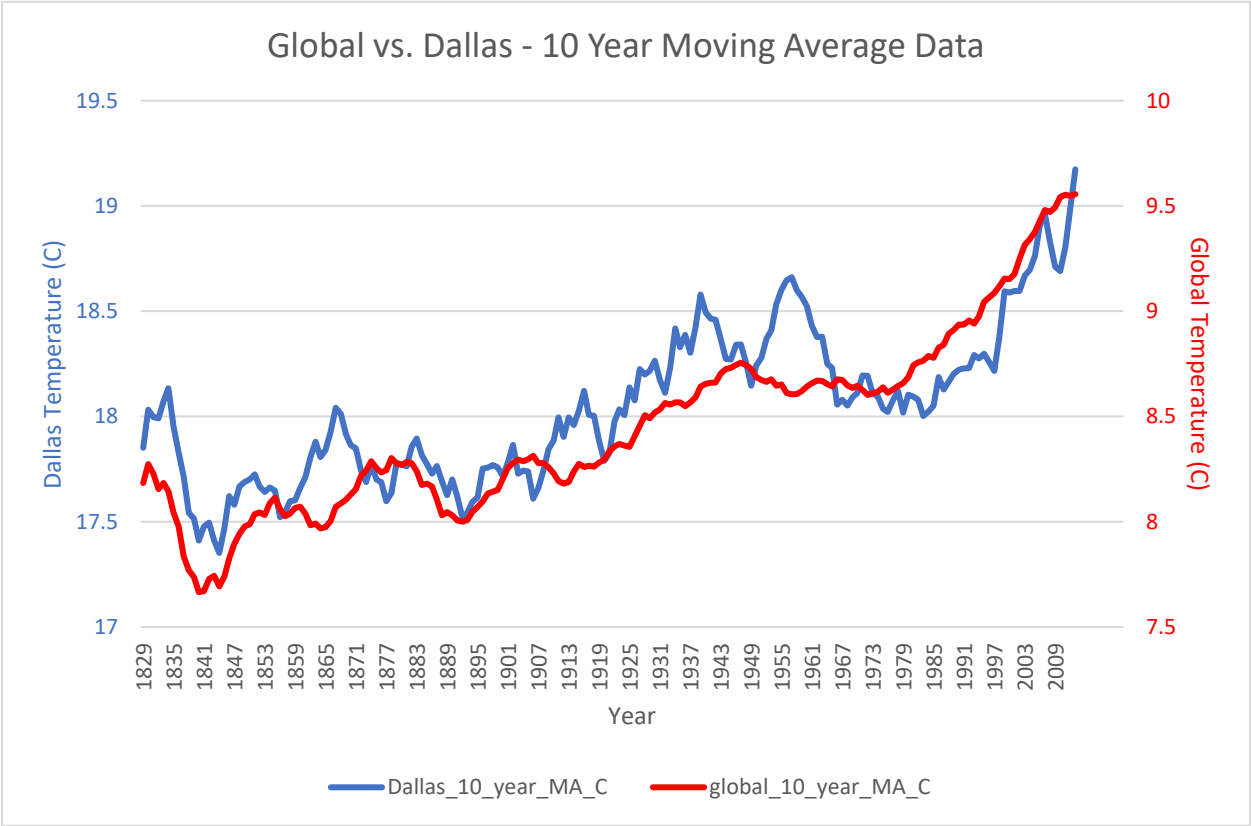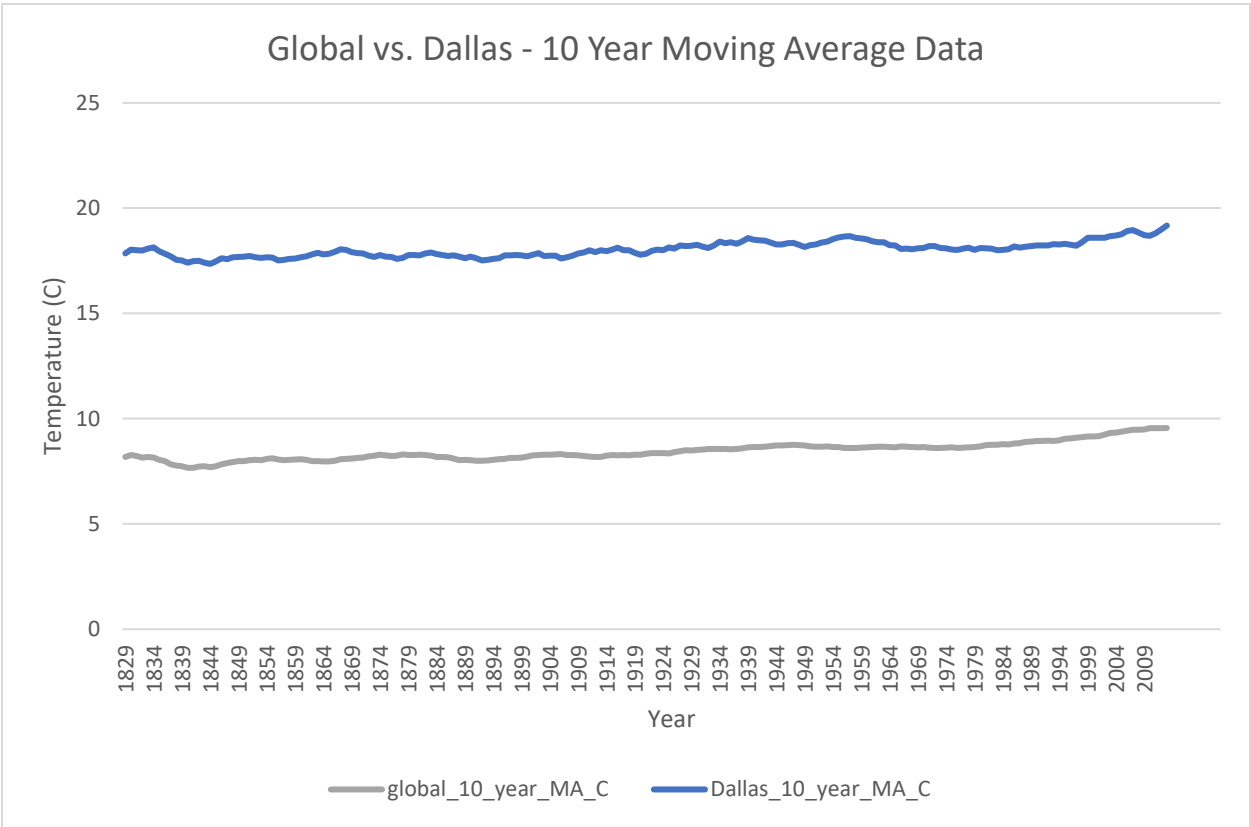
Measured Changes (Min vs. Max)

	Global	Dallas
Minimum	7.666	17.352
Maximum	9.556	19.174
Actual Difference	1.89	1.822
% Difference	24.7%	10.5%

Measured Changes (Earliest vs. Latest)

	Global	Dallas
1829	8.184	17.852
2013	9.556	19.174
Actual Difference	1.372	1.322
% Difference	16.8%	7.4%

Analysis

After reviewing the above data, the following observations can be made:

- The Dallas data, representing only the Dallas area, is reflective of a moderate climate, whereas the global data incorporates information from the globe which contains more extreme climates.
- Since the extremes of the poles are drastically colder than Dallas than are the extremes at the hottest points on the globe warmer than Dallas, the average global data is colder than that of the moderate Dallas climate.
- The actual change in degrees Celsius over the reported period (1829-2013) as well as when comparing the minimum and maximum vales over the reported period changed an almost identical number of degrees.
- The percent change in degrees Celsius is therefore higher for the global data than for the Dallas data.

Conclusions

Based on a 10-year moving average, the Global temperatures have risen between 1.37 and 1.89 degrees Celsius depending on whether comparing min/max or earliest/latest observations. In either case, the global temperatures reflect a warming trend.

The Dallas area has warmed at a nearly identical pace (in actual degrees) to the global data. This can be seen clearly by the second graph above in which the axes have been split and shifted to overlay the two datasets on an identical scale.