

Udacity Data Analysis Term 1 Weather Trend Project

by
Ismail Osman

Introduction:

I extracted data from the provided database, then exported the temperature data for the world as well as the closest big city to where I live. I have used SQL statements to extract the data from the database then exported into a comma-separated value (CSV) format for further analysis.

I calculated the moving average by taking the arithmetic mean of a given set of values. For example, to calculate a seven-year moving average I added up the average yearly weather from the past seven years and then divide the result by seven (7).

I originally plotted the data with the yearly temperature then plotted 7-year moving averages to smooth out data to make it easier to observe long-term trends rather than yearly trend and create easily interpreted visual aid.

Data Preprocessing:

When I extracted the data for the Tucson and Global, I noticed that Tucson did not have data earlier than 1835 when the Global data goes back as far as 1750. To conduct better analysis and since the data already in average values I decided to ignore the global average temperature from 1750 to 1835. Usually I would have replaced the missing values for Tucson with the average temperature but in this case, we already working with average data.

SQL Scripts Used:

Tucson

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

Global

```
SELECT year, avg_temp  
FROM global_data
```

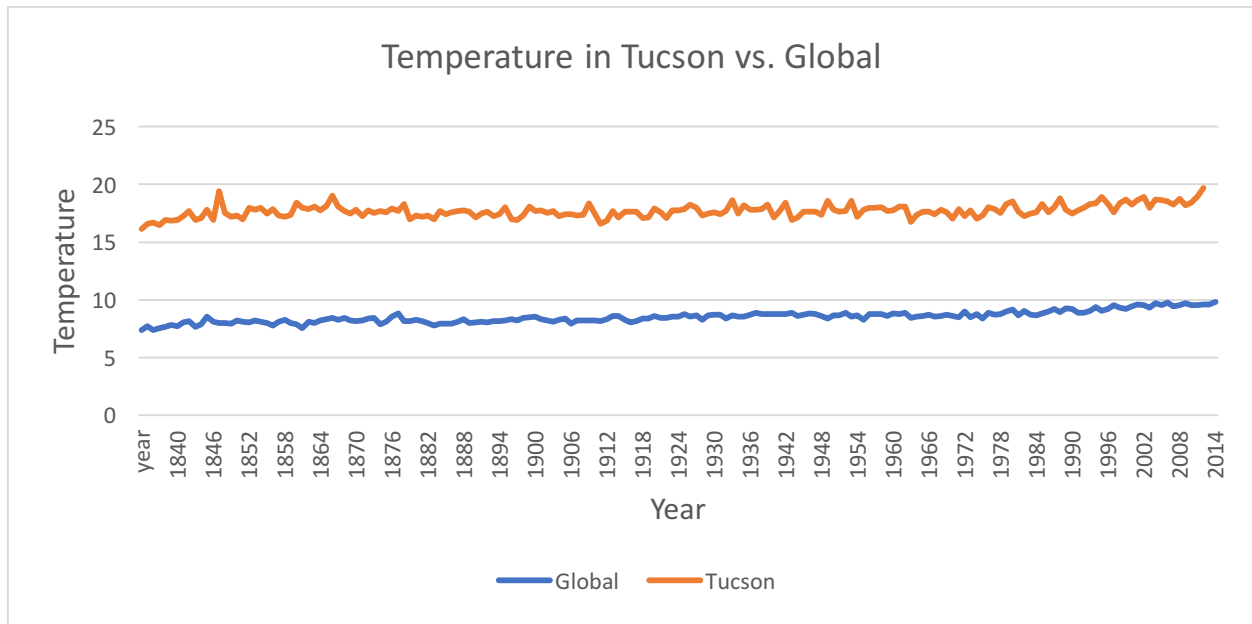


Figure 1: Average Temperature in years before transformation

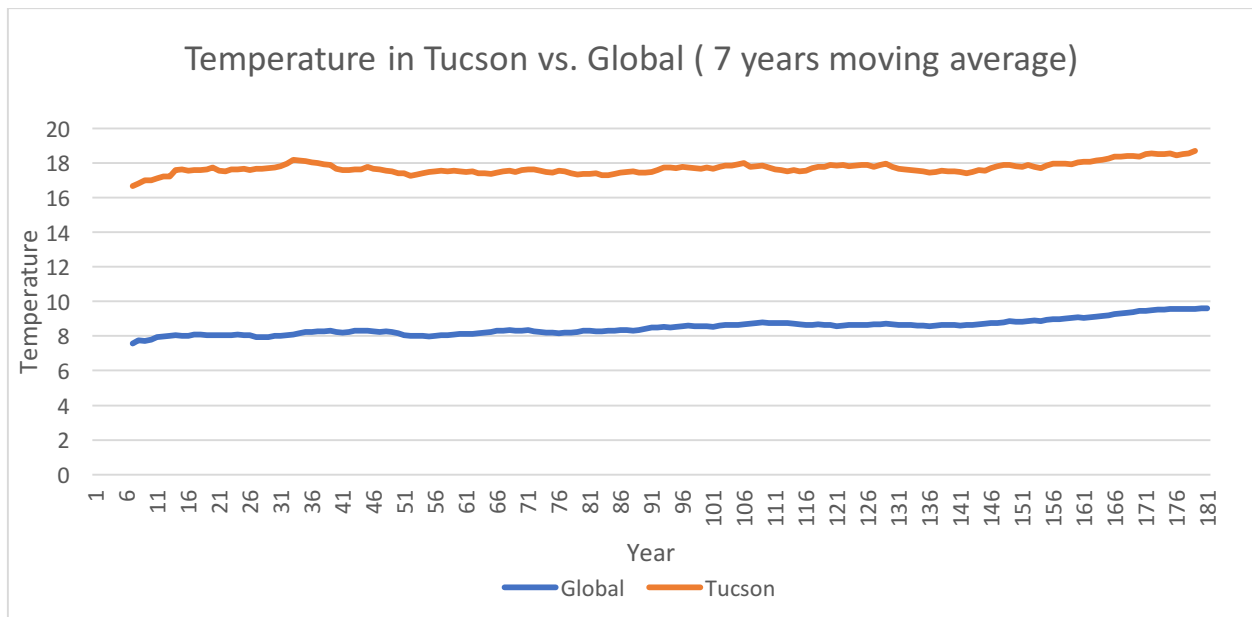


Figure 2: Seven year moving average of the temperature in Tucson and Global

Results and Discussions:

The plots show that the average temperature for Tucson has higher fluctuation than Global and both Global & Tucson are slowly on the upward trend.

The average temperature in Tucson has wider variance than the global average over time. This could have to do with larger coverage and various areas for global which means that it is difficult to change the global average values quickly.

Another key trend I noticed is that Tucson in general, has higher (warmer) average temperature than the average global which means that the global average tends to have more cold places relative than warm ones comparing to Tucson. These wider colder zones drove the average temperature to be lower than the average temperature in the City of Tucson.

When the data was transformed to a 7-year moving average, from figure 2. The trends become more noticeable because they are smoother to visualize and become more consistent as they become less prone to outliers.

Conclusion:

In summary, the evidence from the analysis above suggests that the temperature of the world has been rising over the past century (1920) exponentially, wherein Tucson the temperature is also rising exponentially but only in the two decades. With this small analysis, it is encouraging to do more research to determine what caused the change in temperature and what to do about.