

# Project 1:Exploring Weather Trends

Jared Thacker

September 1, 2020

# Chapter 1

## Report

### 1.1 Outline

In this project PostgreSQL was used to obtain both the global and local Atlanta temperature data from a database kept by Udacity. The SQL statements to extract the Atlanta temperature data is:

```
SELECT *  
FROM city_data  
WHERE country = 'United States' AND city = 'Atlanta'  
LIMIT 100
```

And the SQL statements to extract the global temperature data is:

```
SELECT *  
FROM global_data
```

Python, with the pandas and matplotlib libraries, was then used to load and visualize the data. The rolling mean method from pandas was used to calculate the moving average with a window of 5. This means that the previous 5 data points (including the current data point) were used to calculate the average at the current point in time. A window of  $n = 20$  was initially chosen but tended to smooth out the curve too much, so a window of  $n = 5$  was settled upon and maintained the noise in the graph, but also captured the trend in the data well.

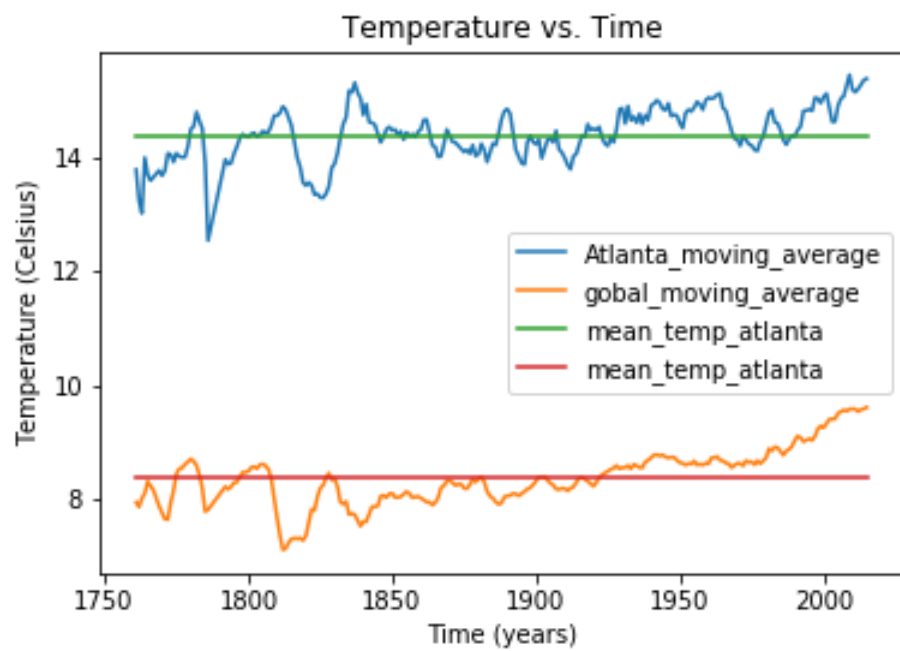


Figure 1.1: The Atlanta temperature data and moving average to show the trend in the data.

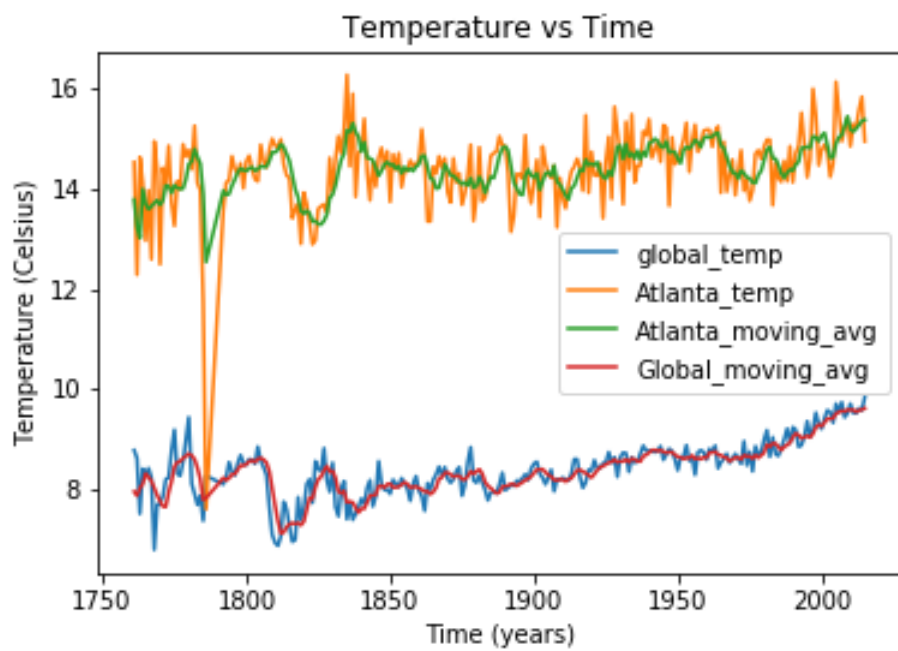


Figure 1.2: The global temperature data and moving average to show the trend in the data.

## 1.2 Observations

The moving averages definitely show that the average global temperature increases as time has passed, however, it is rather unclear with the Atlanta data. In addition, the Atlanta moving average tends to dip and peak in the same place as does the global moving average; this is most easily seen in at approximately year 1825 by comparing the moving averages between the two plots. Third, the temperature in Atlanta is definitely warmer than the global average temperature; the Atlanta temperature is only lower than the global average at one point in time. Lastly, while both temperature trends increase as time passes, the increase in global temperature is more dramatic than Atlanta's as can be seen when comparing the moving averages.

In figure (1.1), the moving average and mean temperatures for the globe and Atlanta are plotted in the same figure and in figure (1.2) the raw time-series and moving averages are plotted both globally and for Atlanta. In the paragraphs that follow, visual inspection as well as the Dickey-Fuller test is used to come to conclusions about the data.

If we visually inspect the raw data (orange curve) for Atlanta in figure (1.2) there does not seem to be an upward trend (values get steadily larger) or a downward trend (values get steadily smaller). In addition, there does not seem to be significantly large differences in each value in the time-series with the exception in the late 1700s which implies a constant variance in time in this time-series. Therefore, upon visual inspection, it looks like the Atlanta data is stationary and has no trend. In figure (1.3), the results of the Dickey-Fuller test performed on the Atlanta data are given. Since the test statistic is smaller any of the critical values, we can conclude that the data Atlanta is stationary. This means that, at least with the data collected so far, there is no trend in the Atlanta data that shows that temperature is increasing or varying significantly.

If we visually inspect the raw (blue curve) for the global data in figure (1.2) there is definitely an upward trend (values get larger). This can be seen more clearly in figure (1.1) where the values past around 1925 seem to *stay* above the mean (the red horizontal line). In addition, the variance seems to decrease as time goes on since the difference between the values seems to decrease as time goes on. Therefore, upon visual inspection, the global time-series data

```

Dickey-Fuller Test Results:
Test Statistic      -7.570341e+00
p-value             2.852367e-11
Lags Used           1.000000e+00
Number of Observations Used  2.480000e+02
Critical Value (1%)  -3.456996e+00
Critical Value (5%)  -2.873266e+00
Critical Value (10%) -2.573019e+00
dtype: float64

```

Figure 1.3: The results of the Dickey-Fuller test for the Atlanta data.

is non-stationary and exhibits trends, the most important of which that the average global temperature is increasing as time passes (global warming). We can confirm this by referring to the Dickey-Fuller test in figure (1.4) . The test statistic is not smaller than any of the critical values which means that we cannot reject that the time-series is non-stationary. Therefore, by visual comparison and by the Dickey-Fuller test, the global time-series has an upwards (increasing) trend and is non-stationary. Therefore, we can conclude that the average temperature of the earth is increasing.

```
Dickey-Fuller Test Results:
Test Statistic          -0.430562
p-value                  0.904870
Lags Used                12.000000
Number of Observations Used 237.000000
Critical Value (1%)      -3.458247
Critical Value (5%)      -2.873814
Critical Value (10%)     -2.573311
dtype: float64
```

Figure 1.4: The results of the Dickey-Fuller test for the global data.