

## Project #1: Exploring Weather Trends

Analyst: Holly Bedwell

Tools: SQL, Excel

### Procedures/Methods:

1. I extracted the city level data using a SQL query and exported the data to a csv file.

```
SELECT d.year, d.city, d.country, d.avg_temp  
FROM city_data AS d  
WHERE city = 'New York'
```

2. I extracted the global data using a SQL query and exported the data to a csv file.

```
SELECT g.year, g.avg_temp  
FROM global_data AS g
```

UPDATE: My original submission included the queries above because the instructions explicitly stated pulling two separate queries:

*Write a SQL query to extract the city level data. Export to CSV.*

*Write a SQL query to extract the global data. Export to CSV.*

As I was not sure how much flexibility I had, I did not join the queries. With this opportunity to resubmit, however, I have implemented the suggested update and have provided the joined query below which yielded 271 results, including the results from global\_data where avg\_temp was blank.

```
SELECT d.year, d.city, d.country, d.avg_temp, g.year, g.avg_temp  
FROM city_data d  
LEFT JOIN global_data g  
ON d.year = g.year  
WHERE d.city = 'New York'
```

3. I opened both of the csv files in Excel and merged the two data sets by performing a vlookup and matching on the year field.
4. Next, I examined the data to determine that the city level data yielded 271 records for New York, with 5 missing records and 4 outliers/low values.
5. Next, I examined the global data which yielded 266 records for New York, with no missing values or obvious outliers.
6. Next, I cleaned the data up by removing all records prior to 1750 because all of the global data was missing for those years and 4 of the city level avg\_temps were missing.
7. Next, I calculated the 7-day moving average for both New York and globally by calculating the average temperatures from 1750 to 1756 in the 1756 row and then dragging the formula down

for the rest of the years (=AVERAGE(M2:M8)). I did this for both the city level data and the global data.

8. To visualize the data, I removed miscellaneous columns as well as the rows without a moving average and created a line chart to visualize my results in Excel.

9. Next, I calculated the correlation coefficient (=CORREL(T2:T259, U2:U259)) for the 7-day moving averages for New York and Globally using the CORREL function to see how the moving averages were correlated between 1756 and 2013. This calculation yielded a correlation coefficient of 0.73 which shows that the 7-day moving average temperatures are highly correlated and mostly fluctuate together, i.e., when New York's 7-day moving average increases, so does the 7-day global average temperature. This can easily be seen in the visualization.

10. Lastly, I formatted the chart, made sure the axis were labeled accurately, and that the chart had a title.

### **Observations:**

1. New York's moving average temperature is warmer than the global moving average temperature by 1.1 on average between 1756 and 2013.

2. There appears to be a dip in New York's average temperature around 1779/1780, however, further investigations shows that the avg\_temp data was extremely low for New York in 1779 and missing for 1780, creating a false dip in New York's average temperature during that time.

3. The overall trend appears that the weather is getting warmer in both New York and globally. This appears to be the case with the exception of 1809, where both NY and global temperatures dip for about a decade and then begin to rise again until 1832. This pattern continues in both New York and globally until 1843 when both averages begin a more gradual climb.

4. Since 1756, on average, the top 10 warmest years on record for New York occur between 2002 and 2013 and globally between 2004 and 2013.

