Project: Export Weather Trends

Step 1: Extract Weather Data

To view the available cities, I executed the following query:

```
SELECT *
FROM city_list
WHERE country = 'United States'
```

One of the results was Columbus, which is the city where I live. I then executed a query to extract the temperature of Columbus from the `city_data` table (then proceeded to download the CSV):

```
SELECT *
FROM city_data
WHERE city = 'Columbus'
```

Finally, I executed the query to extract the global temperature trends from the `global_data` table (then proceeded to download the CSV):

```
SELECT *
FROM global_data
```

Step 2: Add Moving Average Columns in Data Sets

Using Apple Numbers spreadsheet, I added two columns to both global_data_sets.csv and columbus_data_sets.csv - the 7-year moving average and 14-year moving average, with respective columns `moving_avg_7` and `moving_avg_14`.

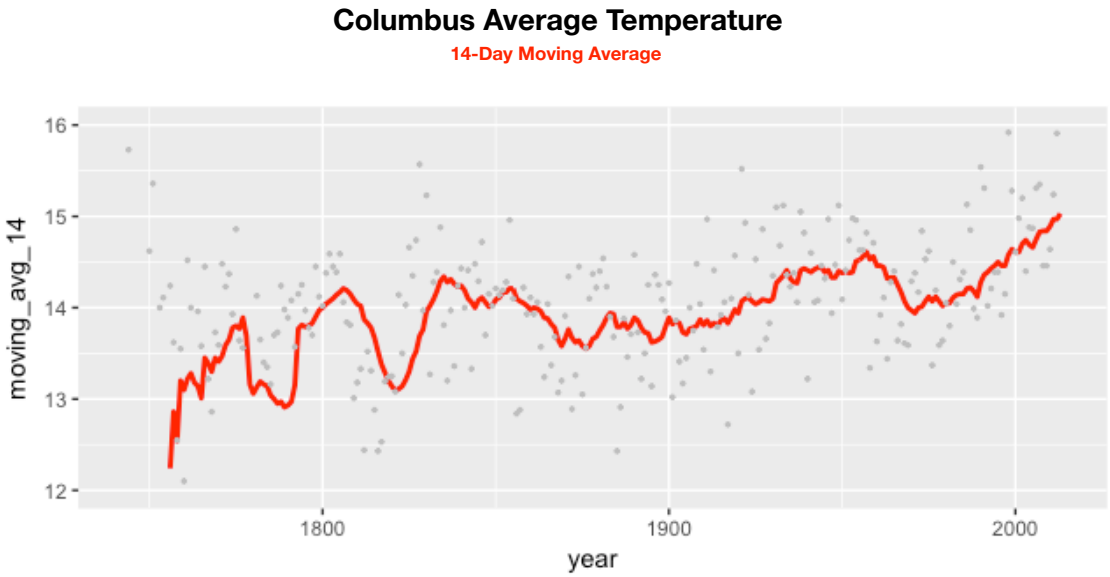
The 7-year moving average was calculated by using the `average` function, with the range of the previous 7 values from the `avg_temp` column. Similarly, the 14-year moving average was calculated using the previous 14 values. I then copied this formula down through the spreadsheet, resulting in a new data table with a list of moving averages.

I then exported these files back to .csv format in order to be easily read and imported into RStudio for analytics.

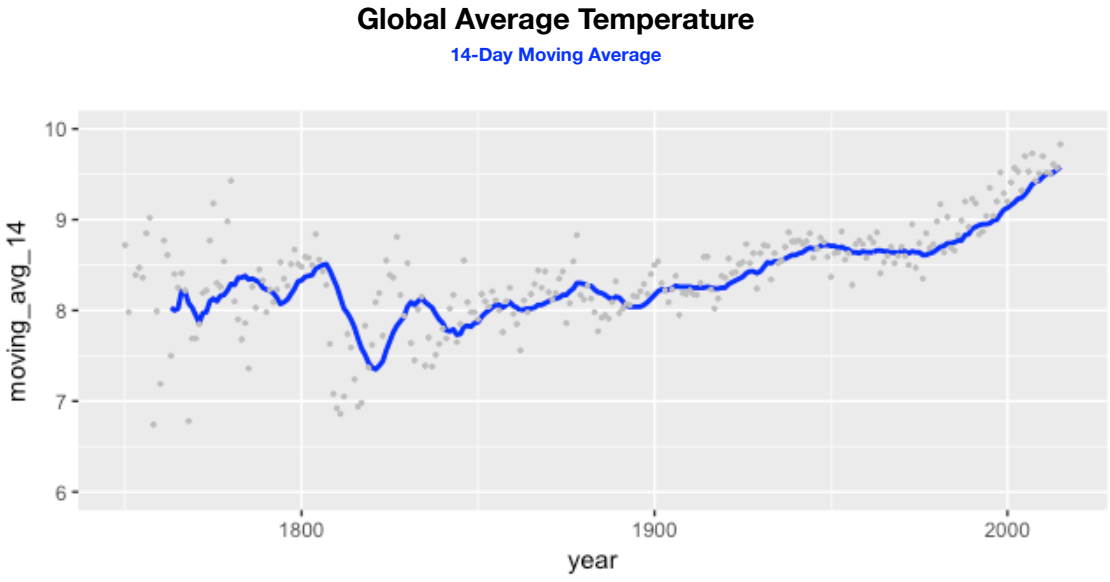
Step 3: Visual Data with RStudio

I imported the two CSV files into RStudio as `columbus_temp_data` and `global_temp_data`, and then used the `ggplot` library to create charts and visualize the data; I used the 14-year moving average. Then to compare apples to apples, I used the same y-scale of 4 degrees.

```
ggplot(columbus_temp_data, aes(x=year)) + ylim(12,16)
      + geom_line(aes(y=moving_avg_14), color="red", size=1)
      + geom_point(aes(y=avg_temp), size=.5, color="gray")
```




```
ggplot(global_temp_data, aes(x=year)) + ylim(6,10)
      + geom_line(aes(y=moving_avg_14), color="blue", size=1)
      + geom_point(aes(y=avg_temp), size=.5, color="gray")
```



Step 4: Analyze the Data and Visualization

Looking at the chart for the global, there was some temperature range variance with sharp decline, subtle increase, and stabilization. There was a sharp turn up in the middle to late 1900s. Overall there is an upward trend since the start of the data set. Data for the Columbus city temperatures show a similar result, with choppy variance and then increase over time.

With some further analytics, we can compare Columbus to Global temperatures by looking at the median of the temperatures and looking at the most recent 14-year moving average. We see a general correlation coefficient of about 1.6, meaning that on median Columbus temperatures are about a 1.6 times factor of the Global temperature. Note: I had to account for the null/NA values.

```
global_avg = median(global_temp_data$avg_temp[global_temp_data$year>1750], na.rm = TRUE)
> 8.37
columbus_avg = median(columbus_temp_data$avg_temp[columbus_temp_data$year>1750],
     na.rm = TRUE)
> 14.06

tail(global_temp_data$moving_avg_14,1)
> 9.58
tail(columbus_temp_data$moving_avg_14,1)
> 15.03
```

Looking at data beginning at 1763 (first year with 14-year moving average for global temperatures) and ending at 2013 (most recent year with Columbus temperatures), we see that in general temperatures have risen each year at .00596°C and .0074°C globally and in Columbus respectively. Temperatures in Columbus have risen at a slightly higher rate than around the world.

```
(global_temp_data$moving_avg_14[264] - global_temp_data$moving_avg_14[14]) / (264-14)
> .00596
(columbus_temp_data$moving_avg_14[271] - columbus_temp_data$moving_avg_14[21]) / (271-21)
> .0074
```