

# Exploring Weather Trends

*Emilie Burke*

*10/12/2017*

## Goal

The goal of this project was to create a visualization and write up describing the similarities and differences of temperature trends between Raleigh, North Carolina and the globe.

The structure of this analysis mimics the requirements of this project. It will be presented in four parts: extraction, CSV, visualization, and observations.

## Extraction

### City Level Data

Raleigh, North Carolina, USA is the closest large city to my home in Fayetteville, North Carolina. After confirming that it was in the `city_list` table, I used it as the focus city in my analysis.

I extracted Raleigh-specific data from the `city_data` table using the following SQL query:

```
SELECT year
      ,city as location
      ,avg_temp
      ,(AVG(avg_temp) OVER(ORDER BY year ROWS BETWEEN 7 PRECEDING AND CURRENT ROW))::float
      AS mov_avg_temp
FROM city_data
WHERE city = 'Raleigh'
```

### Global Data

I extracted the global data using the following SQL query:

```
SELECT year
      ,'global' as location
      ,avg_temp
      ,(AVG(avg_temp) OVER(ORDER BY year ROWS BETWEEN 7 PRECEDING AND CURRENT ROW))::float
      AS mov_avg_temp
FROM global_data
```

Of note here is that in both of these queries I calculated the moving average for the last two years in the query, so as to remove that calculation from a future analysis. I used the moving average for the last seven years (the project did not specify), but given the structure of this analysis, this could be changed just in the queries and the rest of the analysis would not have to be changed.

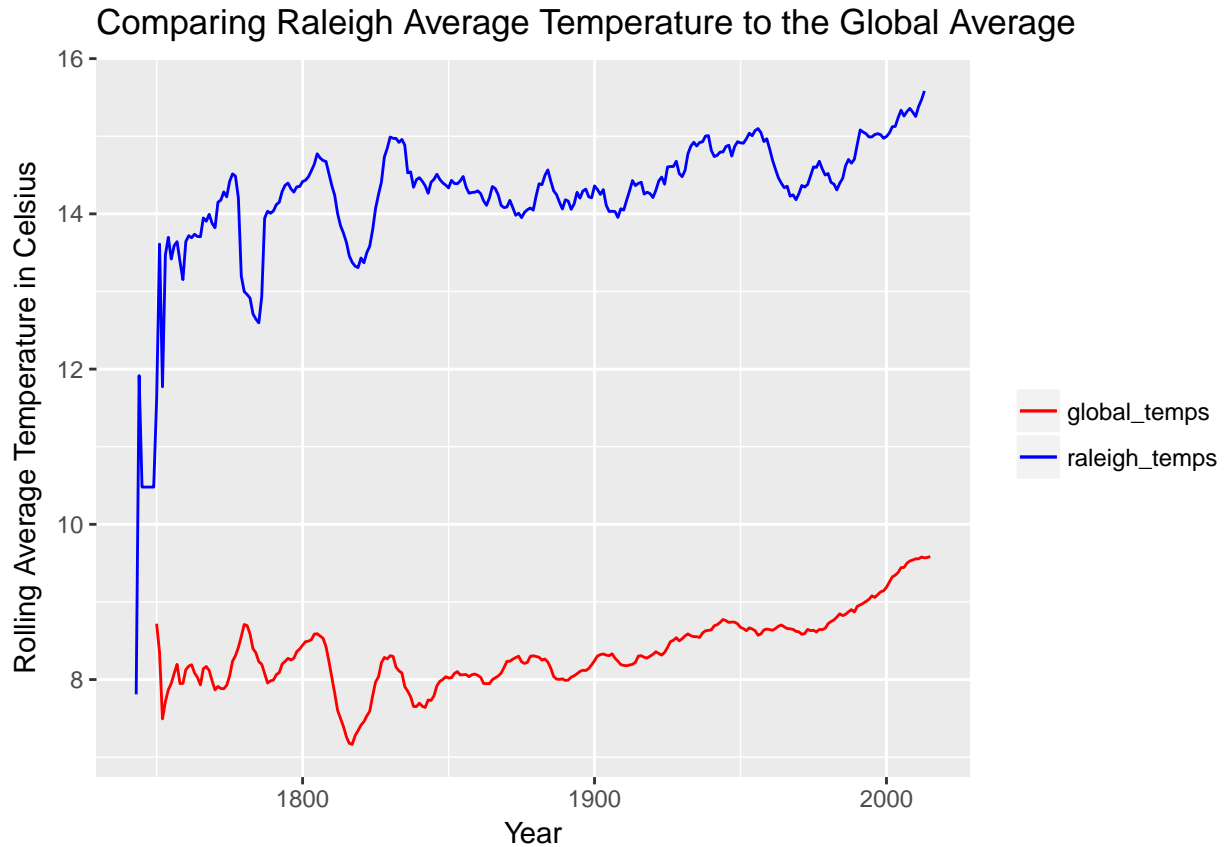
I exported the data from the provided SQL workspace through the “Export” button provided in the user interface. I performed the remaining tasks in R.

## CSV

I imported the CSV into R using the `read.csv` command in the `utils` package.

```
global <- read.csv(file="extract/global.csv", header=TRUE, sep=",")
raleigh <- read.csv(file="extract/raleigh.csv", header=TRUE, sep=",")
```

## Visualization



There are seven years of Raleigh data for which there were no metrics. These were considered NULL values and were included in calculations as such. The same was true for two years of global data.

## Observations