Tyson Bryant
Data Analysis ND Term 1
2017/09/18

Project 1: Explore Weather Trends

Step 1: Extract the data

(Process)
I used the following SQL query to pull up the list of major cities around the world:

```
SELECT *
FROM city_list;
```

Since, the schema box was blank, I scrolled down to see if how they described the United States. Since they simply used “United States” I added a line to my query to pull up only US cities:

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

I found that Portland, Oregon was the closest city to me. Now, to extract and familiarize myself with the city_data table, I did a little something like this:

```
SELECT *
FROM city_data;
```

In hindsight, this wasn’t the best way to do it because it’s such a big table that it took a while, leading me to worry that the page would crash. So I added “LIMIT 1;” just so that I could see how the table is formatted, valuable information for my next query. Since I then knew for sure that city column was simply “city” I had the information I needed to write the following SQL queries in order to pull up the temperature data for Portland and the globe and download the data as CSV files (answers):

```
SELECT *
FROM city_data
WHERE city = 'Portland';
```

```
SELECT *
FROM global_data;
```

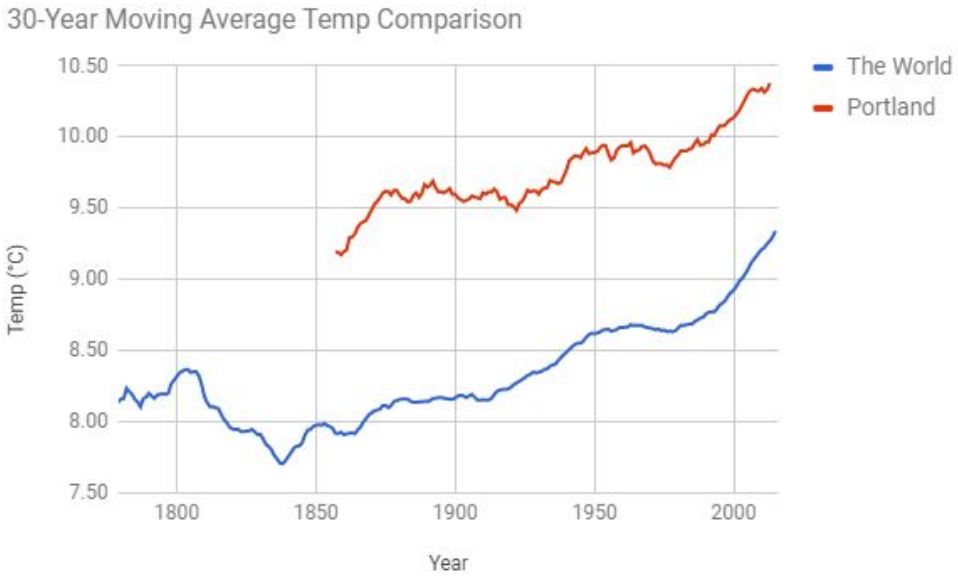
Step 2: Open up the CSV

I decided to open the CSVs in Google Sheets. I then copied the Portland temperature data into the same sheet as the world temperature data, across from the appropriate years, so that I could make the chart for comparison.

Step 3: Create a line chart

First off, I labeled two new columns “The World” and “Portland” to represent the 30-year averages for those respective temperature values. I also thought these labels would look good and clearly explain what the lines on my chart mean (Sheets uses row 1 to label the lines in the legend). I chose to use a 30-year average because I learned in the carbonbrief.org article by Roz Pidcock titled “Explainer: How do Scientists Measure Global Temperature?” that a 30-year average is the contemporary standard. In order to create the moving average values I used formulae like this: =AVERAGE(B2:B31)

I made sure the formula ranged from the beginning of available data and averaged the first 30 years of data available. I then dragged the formula to the end of available data to create moving averages for remaining years. With my two 30 year average columns ready to go I inserted a chart. I set the range of the data for the chart to include the “year” column, which I used as the horizontal axis, the “The World” column, and the “Portland” column. I made sure that the latter two aforementioned columns were included as “sets” in the chart, set the min range for the x-axis to where the moving average for global temperature started at 1779, changed the title to describe what the chart illustrates, and added a label for the y-axis that showed that values were in degrees Celcius. I made the above change to the minimum range on the x-axis in order to eliminate useless dead space in the chart and to unscrunch the lines, making the chart easier to read. Here is what I came up with:



Step 4: Make observations

1. Portland is about 1 °C warmer than the global average. This pattern holds for as long as temperature data for Portland is available.
2. The changes in Portand's temperature over time are very similar to changes in global temperature over the same period, basically following the same hills and troughs. A spike in temperature occurs during the late 1800s for both Portland and the world, followed a by a leveling out during the early 1900s, another spike which leveled out then slightly dipped in the mid 1900s, and finally a huge spike in the late 1900s leading to our current temperatures.
3. The range for moving average temperature data for Portland is (10.38-9.17) = 1.21 °C and the range for moving average temperature data for the world during the same period is (9.27-7.92) = 1.35 °C. This means that while temperature changes for Portland and the world were very similar, a difference of only .14 °C, the world appears to have warmed a little bit more than Portland.
4. A final point of difference between global and Portland temperature data is that the global temperature data is more complete. It goes back further by about 80 years and if you look at the averages Portland is missing points of data from 1830, 1831, and 1846 whereas there is no such lapse in average global temperature data. Now, I double-checked to make sure and the AVERAGE function in Sheets does ignore those values and does not treat them as 0. For example, if 2 of 5 values that Sheets is supposed to average are missing, it will add the remaining 3 and divide by 3, not 5. However, the Portland data could be less accurate for 30 year averages that include the missing years in its range if those missing values were different enough from the usual temperatures for that time period.