# Project 1

Bryan Beus

February 15, 2021

## 1 SQL Queries

The following are the SQL queries used to obtain data from the tables.

I live in North Pole, AK, USA. The nearest city available in the `city_data` table is Seattle. Temperatures for Seattle are available from 1828 to 2013. Therefore, when extracting global data I excluded values for years prior to 1828.

```sql
SELECT year, city, avg_temp
FROM city_data
WHERE city='Seattle';

SELECT *
FROM global_data
WHERE year>=1828 AND year<=2013;
```

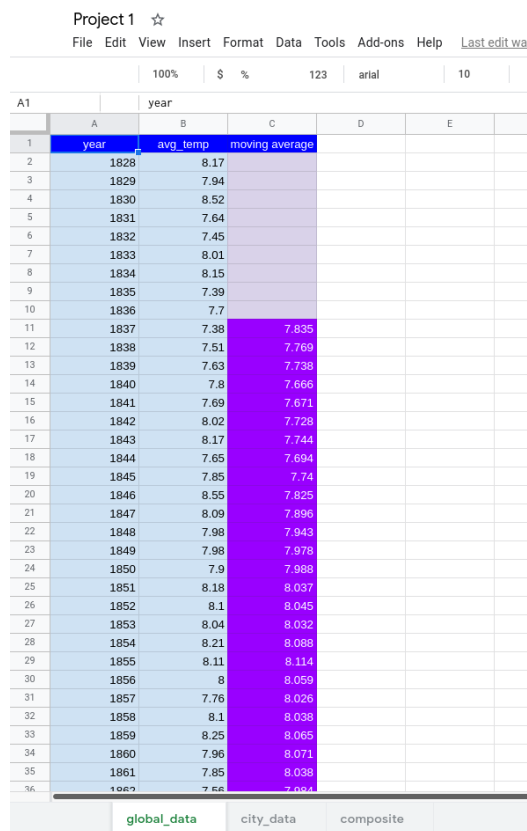# 2   Organizing and Cleaning Data

## 2.1   Global Data

Upon reviewing the CSV data in Google Sheets I found that the `global_data` table provided clean data with no missing values.

I created a `moving average` column and used the following command to create a moving average that considered `10` years at a time.

```
=average(B2:B11)
```

The following is a screenshot of the result.

## 2.2 City Data

The `city_data` table values for Seattle lacked data for several dates.

To repair the missing values, I created an additional column called `wrangled` that used the following spreadsheet command.

```
=IF(C2=0,D1,C2)
```

I applied the above formula to all matching cells in this column. The result was that anywhere data was missing, the cell would automatically fill with the value that was inserted for the prior year.

Using the `wrangled` column as the primary data column, I then performed the same steps as before to obtain a `moving average` column of data, with the following result.

| | A year | B city | C avg_temp | D wrangled | E moving average |
|---|---|---|---|---|---|
| 2 | 1828 | Seattle | 7.13 | 7.13 | |
| 3 | 1829 | Seattle | 6.8 | 6.8 | |
| 4 | 1830 | Seattle | | 6.8 | |
| 5 | 1831 | Seattle | | 6.8 | |
| 6 | 1832 | Seattle | 3.52 | 3.52 | |
| 7 | 1833 | Seattle | 7.48 | 7.48 | |
| 8 | 1834 | Seattle | 7.1 | 7.1 | |
| 9 | 1835 | Seattle | 5.58 | 5.58 | |
| 10 | 1836 | Seattle | 6.74 | 6.74 | |
| 11 | 1837 | Seattle | 6.81 | 6.81 | 6.476 |
| 12 | 1838 | Seattle | 6.59 | 6.59 | 6.422 |
| 13 | 1839 | Seattle | 7.3 | 7.3 | 6.472 |
| 14 | 1840 | Seattle | 6.69 | 6.69 | 6.461 |
| 15 | 1841 | Seattle | 6.81 | 6.81 | 6.462 |
| 16 | 1842 | Seattle | 6.88 | 6.88 | 6.798 |
| 17 | 1843 | Seattle | 6.55 | 6.55 | 6.705 |
| 18 | 1844 | Seattle | 6.41 | 6.41 | 6.636 |
| 19 | 1845 | Seattle | 6.88 | 6.88 | 6.766 |
| 20 | 1846 | Seattle | | 6.88 | 6.78 |
| 21 | 1847 | Seattle | 8.99 | 8.99 | 6.998 |
| 22 | 1848 | Seattle | 6.6 | 6.6 | 6.999 |
| 23 | 1849 | Seattle | 6.68 | 6.68 | 6.937 |
| 24 | 1850 | Seattle | 7.06 | 7.06 | 6.974 |
| 25 | 1851 | Seattle | 7.79 | 7.79 | 7.072 |
| 26 | 1852 | Seattle | 7.08 | 7.08 | 7.092 |
| 27 | 1853 | Seattle | 7.61 | 7.61 | 7.198 |
| 28 | 1854 | Seattle | 6.99 | 6.99 | 7.256 |
| 29 | 1855 | Seattle | 7.31 | 7.31 | 7.299 |
| 30 | 1856 | Seattle | 7.44 | 7.44 | 7.355 |
| 31 | 1857 | Seattle | 7.91 | 7.91 | 7.247 |
| 32 | 1858 | Seattle | 6.91 | 6.91 | 7.278 |
| 33 | 1859 | Seattle | 6.27 | 6.27 | 7.237 |
| 34 | 1860 | Seattle | 7.63 | 7.63 | 7.294 |
| 35 | 1861 | Seattle | 7.18 | 7.18 | 7.233 |
| 36 | 1862 | Seattle | 6.26 | 6.26 | 7.161 |

global_data · city_data · composite

# 3 Analyzing the Data

## 3.1 Initial Compilation

In a new spreadsheet I imported the `year` column and the `moving average` columns from both the global and city data.

Using the following commands I calculated the correlation coefficient.

```
=correl(B11:B187,C11:C187)
```

To obtain the min, overall average, max, and standard deviation of the global temperatures, I executed the following spreadsheet commands.

```
=MIN(global_data!B2:B187)
=AVERAGE(global_data!B2:B187)
=MAX(global_data!B2:B187)
=STDEV(global_data!B2:B187)
```

To obtain the Seattle equivalents, I performed the same commands on the matching data (not shown for brevity).
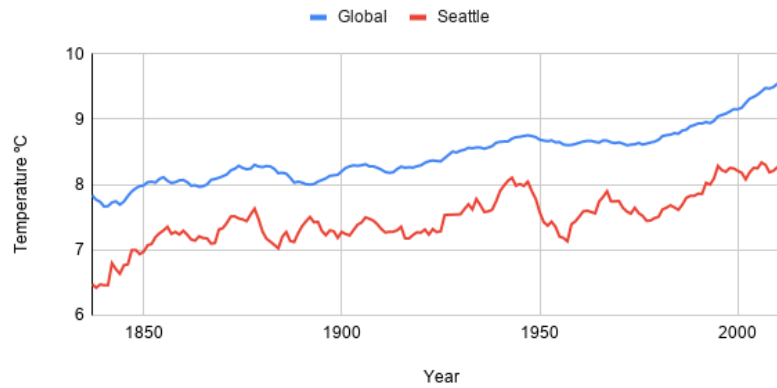
4

## 3.2   Line Graph

Using the default Chart Tool in Google Sheets, I generated a line graph to display the data over the indicated years.
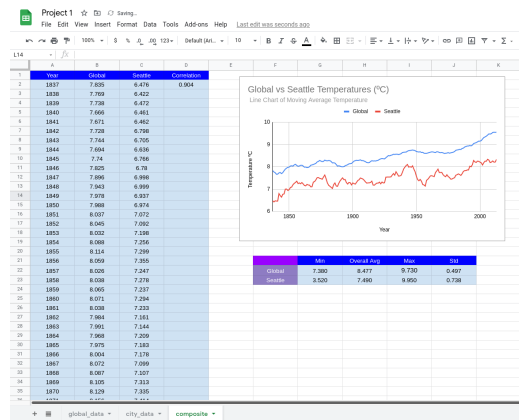




Global vs Seattle Temperatures (°C)

Line Chart of Moving Average Temperature

## 3.3 Observations

- Seattle's overall average of 7.49 °C is consistently cooler than the global average of 8.477 °C. Perhaps this is due to Seattle's northern latitude.

- Seattle's standard deviation of 0.738 °C is far higher than the global standard devision of 0.497 °C. Perhaps this is due to Seattle's frequent inclement weather.

- A correlation coefficient of `0.904` indicates that Seattle's temperature is highly correlated with the global temperature. Perhaps this is due to Seattle's location near the sea. This has the effect of bringing Seattle's temperature into alignment with a far greater surface area due to the manner in which the sea currents conduct temperatures from abroad.

- The temperature does seem to be increasing over the past approximately two hundred years. Seattle's temperature begins at approximately 6.5 °C, rises consistently for the given time span, and ends at approximately 8.2 °C. The global temperature begins at approximately 7.8 °C, rises consistently for the given time span, and ends at approximately 9.6 °C. The cause for this rise in temperature is not clear and is hotly debated, sometimes out of a sincere desire to forestall global catastrophe, and sometimes due to political and corporate agendas.