Scott Haluck
Data Analyst Nanodegree
Project 1 - Explore Weather Trends

Outline
Extract the data

The data was gathered with an SQL query to gather the Kansas City data set and JOIN it to the global data set according to the year. This produced a single CSV file with both data sets side-by-side and ready for analysis. The query limits the year between 1768 and 2013 due to the significant amount of missing data in Kansas City before that point. The gaps and inconsistency in the pre-1768 data would give a misrepresentation of the comparisons.

```
SELECT cd.year, cd.city, cd.avg_temp kc_temp, gd.avg_temp
glob_temp
FROM city_data cd
JOIN global_data gd
ON cd.year = gd.year
WHERE cd.city = 'Kansas City' AND cd.year BETWEEN '1768'
AND '2013';
```

Figure 1: SQL Query for Gathering Local and Gloabl Temperature Data

Calculate 10-year moving average

I used a 10-year moving average given the 200+ data points, because it provides enough granularity to identify trends and smooths the data enough to avoid drastic (and misleading) jumps in the data. See Figure 2 for the formula used to calculate 10-year moving average.

		fx	=AVERAGE(E2:E11)		
	A	B	C	D	E	F
1	year	city	kc_temp	Kansas City, MO	glob_temp	World
2	1768	Kansas City	10.57		6.78	
3	1769	Kansas City	10.92		7.69	
4	1770	Kansas City	11.42		7.69	
5	1771	Kansas City	12.04		7.85	
6	1772	Kansas City			8.19	
7	1773	Kansas City	13.1		8.22	
8	1774	Kansas City	4.69		8.77	
9	1775	Kansas City	12.72		9.18	
10	1776	Kansas City	11.26		8.3	
11	1777	Kansas City	11.49	10.91222222	8.26	8.093
12	1778	Kansas City	8.28	10.65777778	8.54	8.269
13	1779	Kansas City	1.11	9.567777778	8.98	8.398
14	1780	Kansas City		9.33625	9.43	8.572

Figure 2: Initial Data for 1768-1780, Formula used for calculating 10-year moving average
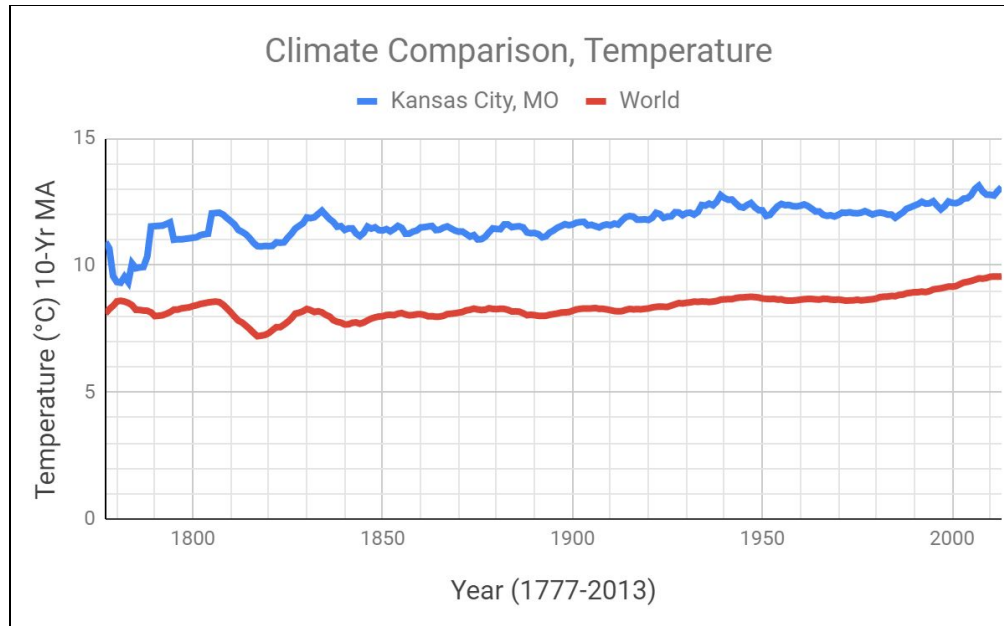
Data Visualization

Create a line chart

Figure 3: Average Temperature Data Line Graph, 10-year Moving Average

Observations

1) The data for Kansas City fluctuates more due to its isolation (one data point averaged over 10-years) as compared to the data for the entire globe (thousands of data points averaged over 10-years). Kansas City's data is more susceptible to deviation due to outliers (e.g. historically hot year in Kansas City), while the data for the globe experiences less variable due to a single outlier (historically hot year in Kansas City would get averaged in with thousands of data points).

2) Despite inconsistent beginnings, the data for Kansas City follows the basic trends of the global data from 1800 onwards. For example, the relative maximum global temperatures of the mid-1800's and early-1830's and relative minimum in the late-1810's are reflected in the local Kansas City graph as well.

3) Kansas City's average yearly temperature is consistently 3-degrees higher or more than that of the global average since 1810.

4) There has been a consistent increase in the global data since 1920, reaching 1 degree higher over that time. This stands in contrast to the relatively consistent and predictable behavior between 1850 and 1920. The increase after 1920 is less prevalent in the Kansas City data. Furthermore, the rate of increase has become more significant between 1980 and 2013 in both local and global data, both data demonstrating an almost 1-degree increase over the last 30-40 years of data. In this most recent trend, the Kansas City data demonstrates the increasing rate more significantly than the global data.