Rachel Layman
11/23/2019

Exploring Weather Trends: Analysis of Two Major US Cities vs. Global Data

Utilizing three datasets provided by Udacity for the Data Analyst Nanodegree program, the objective of this exercise was to analyze average annual temperatures of a major city for the last 150+ years against global annual averages. I live in Salmon, Idaho, and according to the city_list table provided to us, the closest city data available to where I live is from Seattle, Washington. In order to make this determination, it was necessary to first examine the contents of the city_data table. Considering this dataset included a number of global cities, I knew that the first step would be to narrow it down to only those cities located in the United States. I accomplished that by running the following SQL query in the database provided:

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

This provided me with a short list of 52 U.S. cities to quickly read through, out of which I located the closest major city (on a global scale) Seattle, Washington.

I pulled all of the available temperature data for Seattle by using the following SQL query and exported this to a .csv file:

```
SELECT *
FROM city_data
WHERE city = ‘Seattle”
ORDER BY year
```

This .csv provided me with all of the average annual temperature data for Seattle beginning in 1828 until the most recent date in 2013. I uploaded this data into a tab on a new Google Sheet for further analysis later.

I then returned to the Udacity workspace and pulled the global average annual temperature data to compare with the Seattle data. I used the following SQL query to accomplish the data pull:

```
SELECT year, avg_temp
FROM global_data
ORDER BY year
```

I exported this data to a .csv file and uploaded it to a separate tab in the same Google Sheet as I had uploaded the Seattle data. The global average temperature had been recorded for dozens of years earlier than the available data for Seattle. So, once I had the data uploaded into the sheet, I deleted the global data from earlier than 1828 so that the analysis would incorporate the same set of years.

For each tab, both Seattle and global data, I calculated the 7-year moving average on the temperatures, as shown below, in order to more easily visualize trends without dramatic ups and downs from one year to the next.

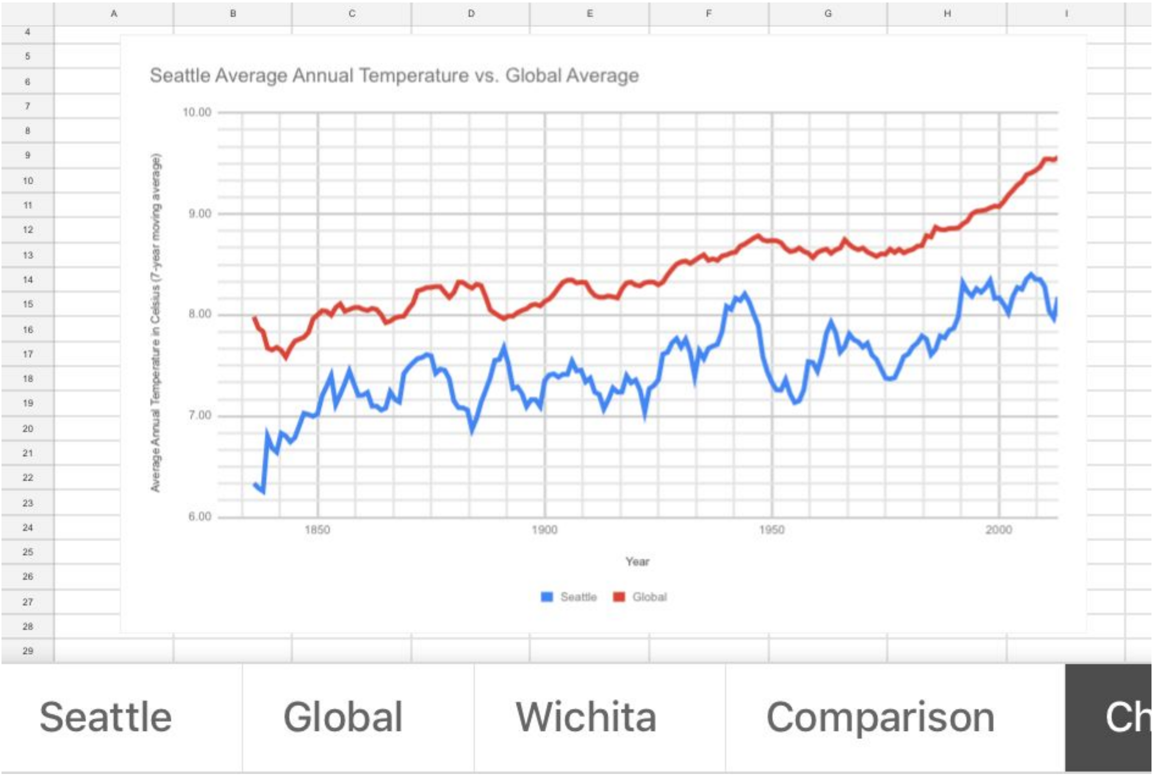
	A	B	C	D
1	year	avg_temp	7 year moving average	
2	1828	7.13		
3	1829	6.8		
4	1832	3.52		
5	1833	7.48		
6	1834	7.1		
7	1835	5.58		
8	1836	6.74	6.34	
9	1837	6.81	6.29	
10	1838	6.59	6.26	
11	1839	7.3	6.80	
12	1840	6.69	6.69	
13	1841	6.81	6.65	
14	1842	6.88	6.83	
15	1843	6.55	6.80	
16	1844	6.41	6.75	
17	1845	6.88	6.79	
18	1847	8.99	7.03	
19	1848	6.6	7.02	
20	1849	6.68	7.00	
21	1850	7.06	7.02	
22	1851	7.79	7.20	
23	1852	7.08	7.30	
24	1853	7.61	7.40	
25	1854	6.99	7.12	
26	1855	7.31	7.22	
27	1856	7.44	7.33	

Seattle | Global | Wichita

	A	B	C
1	year	avg_temp	7 year moving average
2	1828	8.17	
3	1829	7.94	
4	1830	8.52	
5	1831	7.64	
6	1832	7.45	
7	1833	8.01	
8	1834	8.15	7.98
9	1835	7.39	7.87
10	1836	7.7	7.84
11	1837	7.38	7.67
12	1838	7.51	7.66
13	1839	7.63	7.68
14	1840	7.8	7.65
15	1841	7.69	7.59
16	1842	8.02	7.68
17	1843	8.17	7.74
18	1844	7.65	7.78
19	1845	7.85	7.83
20	1846	8.55	7.96
21	1847	8.09	8.00
22	1848	7.98	8.04
23	1849	7.98	8.04
24	1850	7.9	8.00
25	1851	8.18	8.08
26	1852	8.1	8.11

Seattle | Global | Wichita

I then plotted each individual set of 7-year moving averages across time as a separate x-axis line in one single chart, with the Celsius temperature range along the y-axis. See below:



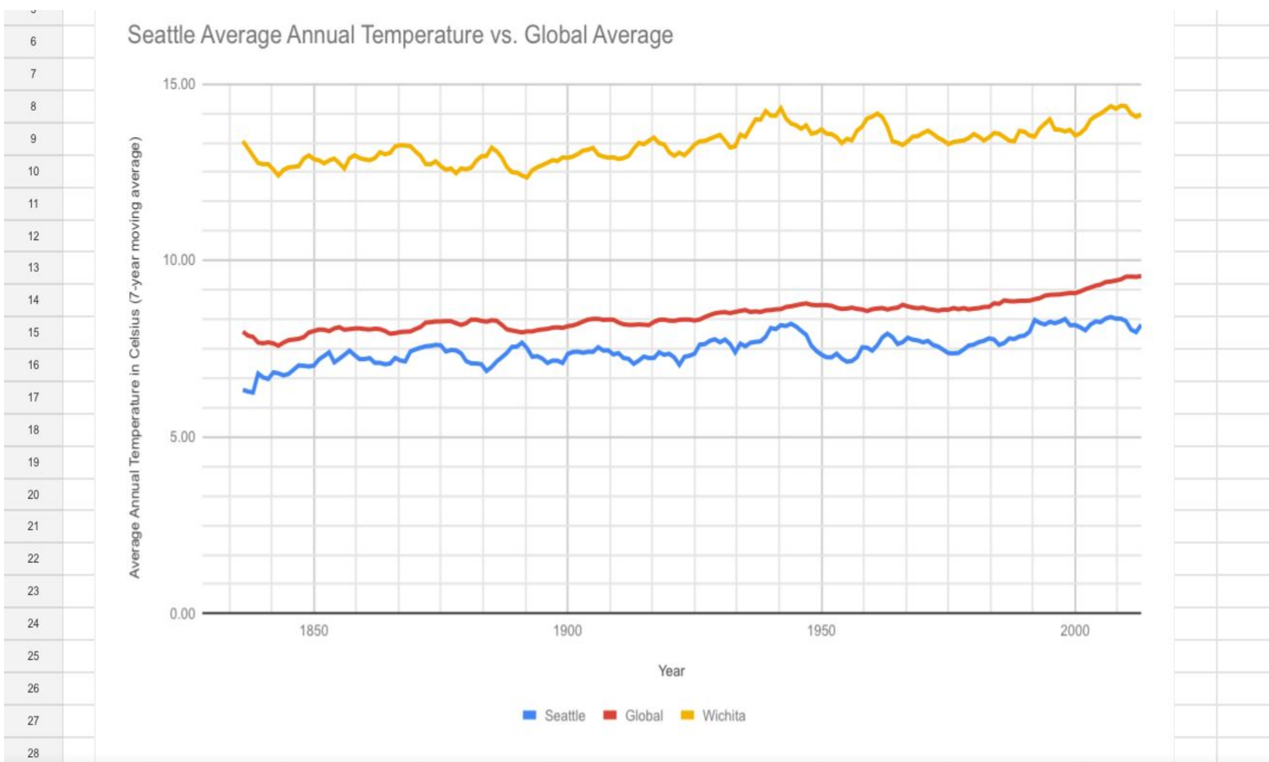
I immediately noticed that across all of these decades shown, the average annual temperature in Seattle is well below the average global temperature. It also appeared from this data visualization, that these two data sets do, in general, seem to trend gradually upward in the same direction across time. It was also curious to me that there are noticeable average temperature dips that occur in both lines approximately every 30 years or so, and that those in the Seattle data tend to happen a bit earlier and dip much more dramatically than those in the average global data.

This realization made me begin to consider the geographic positioning of Seattle and wonder if it is in fact an anomaly of some sort among U.S. cities. I live in Salmon, Idaho, which is technically known as part of the Pacific Northwest. Seattle is the closest city in this dataset, but considering it is located right along the ocean, and it is one of our westernmost U.S. cities, with a maritime climate that I could imagine is probably much more vulnerable to climate fluctuations than a more protected, landlocked Rocky Mountain city like Salmon, where I live, I surmised that Seattle’s data may not be the most accurate representation of the United States as a whole. I have no doubt that my own high desert mountain community experiences much lower average winter temperature, but probably also much higher summertime temperatures than Seattle. Salmon, Idaho is also not on the western edge of the U.S. and as such, would not be the first U.S. city to experience weather changes rolling in from across the ocean. I opted to include an additional dataset into the mix to analyze the annual temperature data of a more landlocked U.S. city.

I went back in to the city_list table and combed back through the 52 U.S. cities for which data was available. I settled on Wichita, Kansas, the most continentally central location that I could find among those cities listed. I pulled the data for Wichita using the following SQL query:

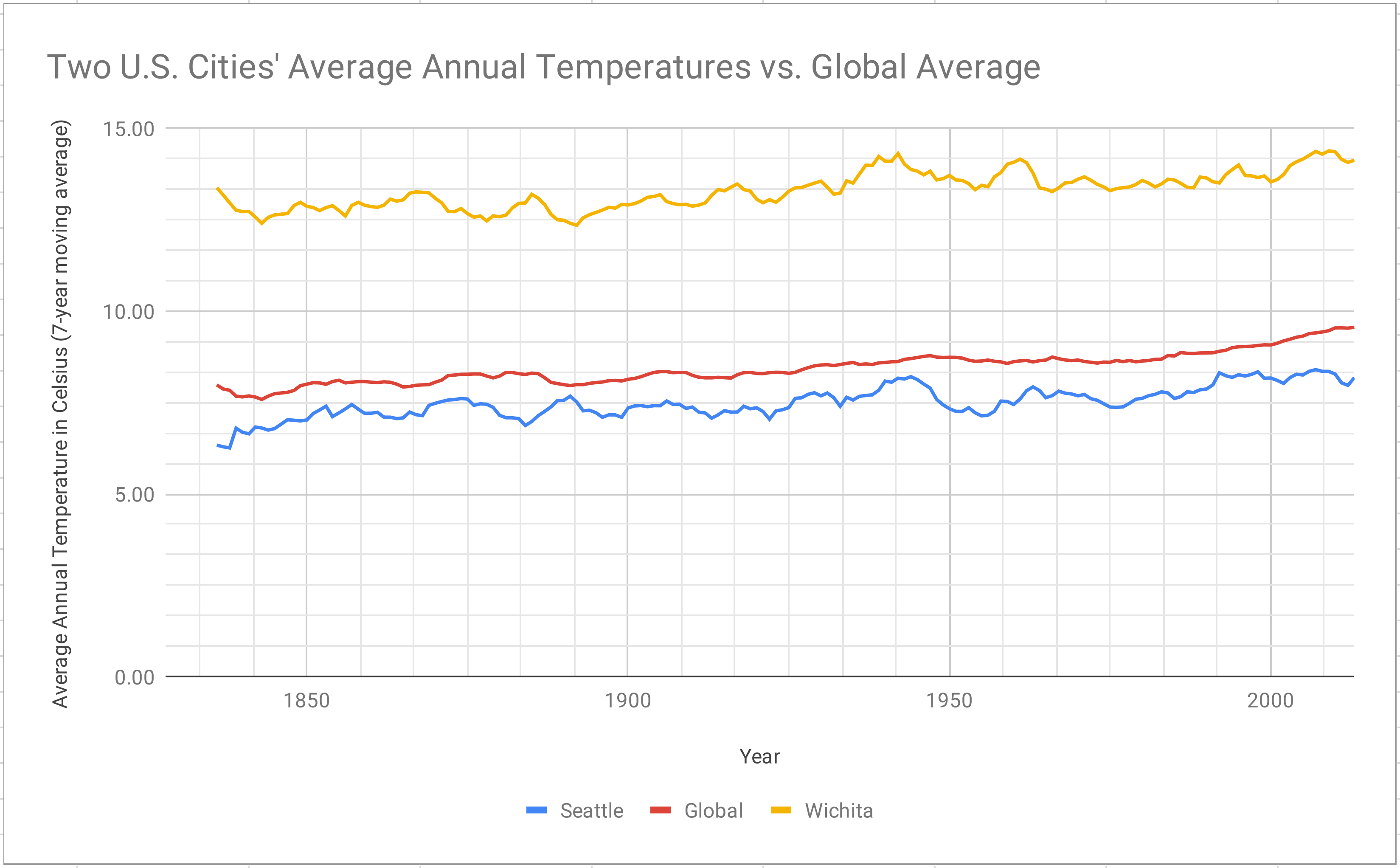
```
SELECT *
FROM city_data
WHERE city = “Wichita’
ORDER BY year
```

Again, I exported the data to a .csv file and uploaded it to another tab in my Weather Data Exploration Google spreadsheet. The Wichita data was recorded as far back as the 1760’s, so again I deleted all of the data prior to 1828, because that year is the earliest recorded information available about Seattle. I created a 3rd data range in my existing line chart so that I could compare the Wichita data to the Seattle and Global annual averages as well. The resulting chart is shown below.



It is apparent at a glance that the average annual temperatures in Wichita are considerably higher than those in Seattle or the global average. With Wichita’s data now in the picture, it becomes clear that Seattle’s weather is much closer to the global average than Wichita, Kansas. Being a northern latitude Pacific Northwest city, Seattle’s average temperatures are much lower. And being a landlocked midwest community, Wichita’s are much higher on average. However, upon examination of the patterns made by both the Seattle and Wichita lines in the chart, it became clear that the general fluctuations from year to year do match up amazingly well, despite the dramatic average temperature spread from one city to the next. Seattle is not necessarily more vulnerable to global climate shifts, in fact, both of these cities seem to be experiencing the same shifts, on a surprisingly proportionate scale.

On average, between 1828 - 2013, the trend for both of these U.S. cities as well as the global average indicates that temperatures are gradually climbing. Based on this data, one might conclude that in general, temperatures are getting warmer on a global scale.



Two U.S. Cities' Average Annual Temperatures vs. Global Average
Average Annual Temperature in Celsius (7-year moving average)
15.00
10.00
5.00
0.00
1850
1900
1950
2000
Year
Seattle
Global
Wichita

year	Seattle	Global	Wichita
1828			
1829			
1832			
1833			
1834			
1835			
1836	6.34	7.98	13.38
1837	6.29	7.87	13.18
1838	6.26	7.84	12.96
1839	6.80	7.67	12.76
1840	6.69	7.66	12.72
1841	6.65	7.68	12.73
1842	6.83	7.65	12.58
1843	6.80	7.59	12.40
1844	6.75	7.68	12.57
1845	6.79	7.74	12.63
1847	7.03	7.78	12.67
1848	7.02	7.83	12.89
1849	7.00	7.96	12.98
1850	7.02	8.00	12.87
1851	7.20	8.04	12.84
1852	7.30	8.04	12.75
1853	7.40	8.00	12.83
1854	7.12	8.08	12.89
1855	7.22	8.11	12.75
1856	7.33	8.04	12.60
1857	7.45	8.06	12.89
1858	7.32	8.07	12.98
1859	7.21	8.08	12.90
1860	7.21	8.06	12.86
1861	7.24	8.05	12.84

1862	7.10	8.07	12.90
1863	7.10	8.06	13.07
1864	7.06	8.00	13.01
1865	7.08	7.93	13.05
1866	7.24	7.94	13.23
1867	7.17	7.97	13.26
1868	7.14	7.98	13.25
1869	7.42	7.99	13.24
1870	7.48	8.06	13.09
1871	7.53	8.12	12.96
1872	7.57	8.24	12.73
1873	7.58	8.25	12.72
1874	7.61	8.27	12.81
1875	7.60	8.27	12.67
1876	7.42	8.28	12.57
1877	7.47	8.28	12.61
1878	7.45	8.23	12.47
1879	7.37	8.18	12.61
1880	7.15	8.22	12.58
1881	7.09	8.33	12.63
1882	7.08	8.32	12.83
1883	7.06	8.29	12.95
1884	6.87	8.27	12.96
1885	6.98	8.31	13.20
1886	7.14	8.29	13.09
1887	7.26	8.18	12.92
1888	7.38	8.05	12.65
1889	7.55	8.02	12.50
1890	7.56	7.99	12.49
1891	7.68	7.96	12.40
1892	7.52	7.99	12.35

1893	7.27	7.99	12.55
1894	7.29	8.03	12.64
1895	7.22	8.05	12.70
1896	7.10	8.06	12.77
1897	7.16	8.10	12.84
1898	7.16	8.11	12.82
1899	7.10	8.09	12.92
1900	7.35	8.14	12.91
1901	7.41	8.16	12.94
1902	7.42	8.21	13.01
1903	7.39	8.27	13.11
1904	7.42	8.32	13.14
1905	7.41	8.35	13.19
1906	7.54	8.35	13.00
1907	7.45	8.32	12.94
1908	7.46	8.33	12.91
1909	7.34	8.32	12.92
1910	7.38	8.24	12.87
1911	7.24	8.19	12.90
1912	7.22	8.18	12.96
1913	7.07	8.18	13.17
1914	7.16	8.19	13.33
1915	7.28	8.18	13.29
1916	7.24	8.17	13.39
1917	7.24	8.26	13.48
1918	7.40	8.32	13.33
1919	7.33	8.33	13.28
1920	7.36	8.30	13.06
1921	7.26	8.29	12.97
1922	7.05	8.32	13.05
1923	7.27	8.33	12.98

1924	7.30	8.33	13.12
1925	7.36	8.30	13.28
1926	7.62	8.33	13.37
1927	7.63	8.40	13.38
1928	7.73	8.45	13.45
1929	7.77	8.50	13.50
1930	7.68	8.53	13.56
1931	7.76	8.54	13.39
1932	7.63	8.51	13.20
1933	7.40	8.54	13.23
1934	7.65	8.57	13.57
1935	7.57	8.60	13.50
1936	7.67	8.54	13.75
1937	7.69	8.56	13.99
1938	7.71	8.54	13.99
1939	7.84	8.59	14.23
1940	8.09	8.60	14.10
1941	8.06	8.62	14.10
1942	8.17	8.62	14.31
1943	8.14	8.68	14.02
1944	8.21	8.70	13.88
1945	8.13	8.73	13.83
1946	8.01	8.76	13.73
1947	7.90	8.78	13.83
1948	7.59	8.74	13.59
1949	7.44	8.73	13.62
1950	7.33	8.74	13.71
1951	7.26	8.74	13.59
1952	7.26	8.72	13.58
1953	7.36	8.66	13.49
1954	7.22	8.63	13.32

1955	7.14	8.64	13.44
1956	7.15	8.66	13.40
1957	7.26	8.63	13.68
1958	7.54	8.61	13.79
1959	7.53	8.57	14.02
1960	7.45	8.62	14.07
1961	7.60	8.64	14.16
1962	7.82	8.65	14.06
1963	7.93	8.61	13.78
1964	7.83	8.65	13.37
1965	7.63	8.66	13.34
1966	7.69	8.75	13.27
1967	7.81	8.70	13.37
1968	7.76	8.67	13.51
1969	7.74	8.65	13.52
1970	7.68	8.66	13.61
1971	7.72	8.62	13.67
1972	7.61	8.60	13.58
1973	7.56	8.58	13.47
1974	7.47	8.61	13.40
1975	7.38	8.60	13.30
1976	7.37	8.65	13.35
1977	7.38	8.62	13.38
1978	7.48	8.65	13.40
1979	7.59	8.62	13.46
1980	7.62	8.64	13.58
1981	7.69	8.65	13.50
1982	7.73	8.68	13.40
1983	7.79	8.69	13.48
1984	7.76	8.79	13.61
1985	7.61	8.77	13.59

1986	7.66	8.87	13.50
1987	7.79	8.85	13.39
1988	7.78	8.84	13.37
1989	7.85	8.86	13.67
1990	7.87	8.86	13.64
1991	7.98	8.86	13.54
1992	8.32	8.90	13.51
1993	8.23	8.93	13.74
1994	8.19	9.00	13.87
1995	8.26	9.03	14.00
1996	8.23	9.03	13.71
1997	8.27	9.04	13.70
1998	8.34	9.06	13.65
1999	8.17	9.08	13.70
2000	8.17	9.07	13.54
2001	8.11	9.12	13.60
2002	8.02	9.19	13.74
2003	8.18	9.23	13.98
2004	8.27	9.29	14.09
2005	8.26	9.32	14.16
2006	8.36	9.39	14.26
2007	8.40	9.41	14.37
2008	8.35	9.43	14.30
2009	8.35	9.47	14.38
2010	8.28	9.54	14.37
2011	8.04	9.54	14.15
2012	7.97	9.54	14.07
2013	8.18	9.56	14.13
		9.59	
		9.56	
		9.57	