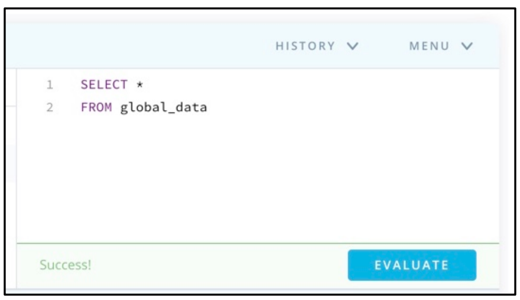
These are the steps I took to complete Project 1

1. Read requirements and check available resources
   1.1. Review Project Instructions
   1.2. Review Project Rubric
   1.3. Review Temperature Database

2. Extract the data
   2.1. Tools used
      2.1.1. I used SQL to extract the required data

   2.2. The SQL queries used to extract the data
      2.2.1. I used the following to extract global data:



      2.2.2. I used the following to extract city data

3. Open up the CSV
   3.1. Tools used

      I used Excel to save the data from a CSV file to an XLS file

4. Create a line chart

4.1. Tools used

In my XLS file I created my data area. I truncated approximately 100 years of global temperature data prior to the first recorded temperature data for my city. When finished, I was able to compare global and city temperature data from 1858-2013. That resulted in 156 units of comparable A:B, moving average data.

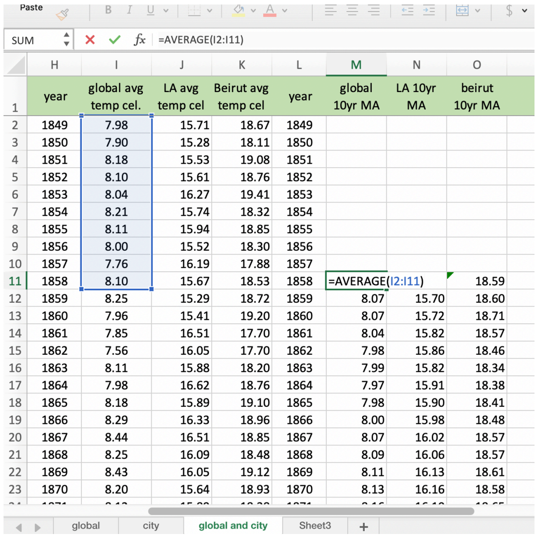
I also created a second series converting all the above data from Celsius to Fahrenheit using the calculation (x°C × 9/5) + 32. Though this woudn't be presented in the final submission because the requirements didn't request it, I analyzed and created primarily from this data set because the values made more sense to me personally.

4.2. Describe calculating the moving average

I watched the video describing how to calculate a moving average. I downloaded the CSV and practiced it in Excel and then I applied what I learned to my temperature data. I chose to make it a 10-year moving average of temperatures as no specific interval was indicated in the requirements. I felt that a 10-year moving average spread over 156 years would provide the necessary smoothing effect and still be able to show significant trends.

REVISION 1:

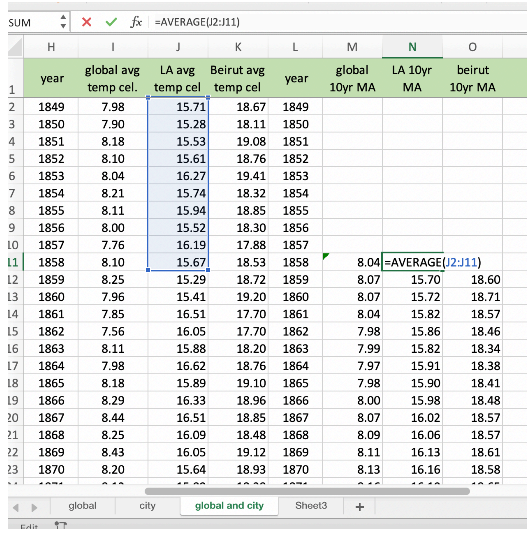
Below are screen shots indicating the formulas used to create the 10-year moving averages in Excel. Once the first one was created for Global and Los Angeles (and eventually Beirut), then it was merely a case of dragging the formula to the bottom of the column ending in the row with the date of 2013. It was these moving averages that were used in the remaining calculations.



SUM =AVERAGE(I2:I11)

	year	global avg temp cel.	LA avg temp cel	Beirut avg temp cel	year	global 10yr MA	LA 10yr MA	beirut 10yr MA
2	1849	7.98	15.71	18.67	1849			
3	1850	7.90	15.28	18.11	1850			
4	1851	8.18	15.53	19.08	1851			
5	1852	8.10	15.61	18.76	1852			
6	1853	8.04	16.27	19.41	1853			
7	1854	8.21	15.74	18.32	1854			
8	1855	8.11	15.94	18.85	1855			
9	1856	8.00	15.52	18.30	1856			
10	1857	7.76	16.19	17.88	1857			
11	1858	8.10	15.67	18.53	1858	=AVERAGE(I2:I11)		18.59
12	1859	8.25	15.29	18.72	1859	8.07	15.70	18.60
13	1860	7.96	15.41	19.20	1860	8.07	15.72	18.71
14	1861	7.85	16.51	17.70	1861	8.04	15.82	18.57
15	1862	7.56	16.05	17.70	1862	7.98	15.86	18.46
16	1863	8.11	15.88	18.20	1863	7.99	15.82	18.34
17	1864	7.98	16.62	18.76	1864	7.97	15.91	18.38
18	1865	8.18	15.89	19.10	1865	7.98	15.90	18.41
19	1866	8.29	16.33	18.96	1866	8.00	15.98	18.48
20	1867	8.44	16.51	18.85	1867	8.07	16.02	18.57
21	1868	8.25	16.09	18.48	1868	8.09	16.06	18.57
22	1869	8.43	16.05	19.12	1869	8.11	16.13	18.61
23	1870	8.20	15.64	18.93	1870	8.13	16.16	18.58

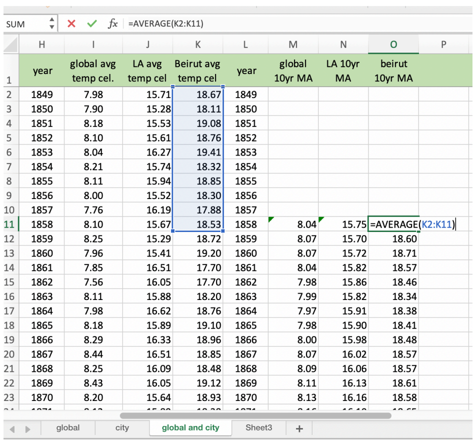
global city global and city Sheet3



SUM =AVERAGE(J2:J11)

	year	global avg temp cel.	LA avg temp cel	Beirut avg temp cel	year	global 10yr MA	LA 10yr MA	beirut 10yr MA
2	1849	7.98	15.71	18.67	1849			
3	1850	7.90	15.28	18.11	1850			
4	1851	8.18	15.53	19.08	1851			
5	1852	8.10	15.61	18.76	1852			
6	1853	8.04	16.27	19.41	1853			
7	1854	8.21	15.74	18.32	1854			
8	1855	8.11	15.94	18.85	1855			
9	1856	8.00	15.52	18.30	1856			
10	1857	7.76	16.19	17.88	1857			
11	1858	8.10	15.67	18.53	1858	8.04	=AVERAGE(J2:J11)	
12	1859	8.25	15.29	18.72	1859	8.07	15.70	18.60
13	1860	7.96	15.41	19.20	1860	8.07	15.72	18.71
14	1861	7.85	16.51	17.70	1861	8.04	15.82	18.57
15	1862	7.56	16.05	17.70	1862	7.98	15.86	18.46
16	1863	8.11	15.88	18.20	1863	7.99	15.82	18.34
17	1864	7.98	16.62	18.76	1864	7.97	15.91	18.38
18	1865	8.18	15.89	19.10	1865	7.98	15.90	18.41
19	1866	8.29	16.33	18.96	1866	8.00	15.98	18.48
20	1867	8.44	16.51	18.85	1867	8.07	16.02	18.57
21	1868	8.25	16.09	18.48	1868	8.09	16.06	18.57
22	1869	8.43	16.05	19.12	1869	8.11	16.13	18.61
23	1870	8.20	15.64	18.93	1870	8.13	16.16	18.58

global city global and city Sheet3



SUM =AVERAGE(K2:K11)

	year	global avg temp cel.	LA avg temp cel	Beirut avg temp cel	year	global 10yr MA	LA 10yr MA	beirut 10yr MA
2	1849	7.98	15.71	18.67	1849			
3	1850	7.90	15.28	18.11	1850			
4	1851	8.18	15.53	19.08	1851			
5	1852	8.10	15.61	18.76	1852			
6	1853	8.04	16.27	19.41	1853			
7	1854	8.21	15.74	18.32	1854			
8	1855	8.11	15.94	18.85	1855			
9	1856	8.00	15.52	18.30	1856			
10	1857	7.76	16.19	17.88	1857			
11	1858	8.10	15.67	18.53	1858	8.04	15.75	=AVERAGE(K2:K11)
12	1859	8.25	15.29	18.72	1859	8.07	15.70	18.60
13	1860	7.96	15.41	19.20	1860	8.07	15.72	18.71
14	1861	7.85	16.51	17.70	1861	8.04	15.82	18.57
15	1862	7.56	16.05	17.70	1862	7.98	15.86	18.46
16	1863	8.11	15.88	18.20	1863	7.99	15.82	18.34
17	1864	7.98	16.62	18.76	1864	7.97	15.91	18.38
18	1865	8.18	15.89	19.10	1865	7.98	15.90	18.41
19	1866	8.29	16.33	18.96	1866	8.00	15.98	18.48
20	1867	8.44	16.51	18.85	1867	8.07	16.02	18.57
21	1868	8.25	16.09	18.48	1868	8.09	16.06	18.57
22	1869	8.43	16.05	19.12	1869	8.11	16.13	18.61
23	1870	8.20	15.64	18.93	1870	8.13	16.16	18.58

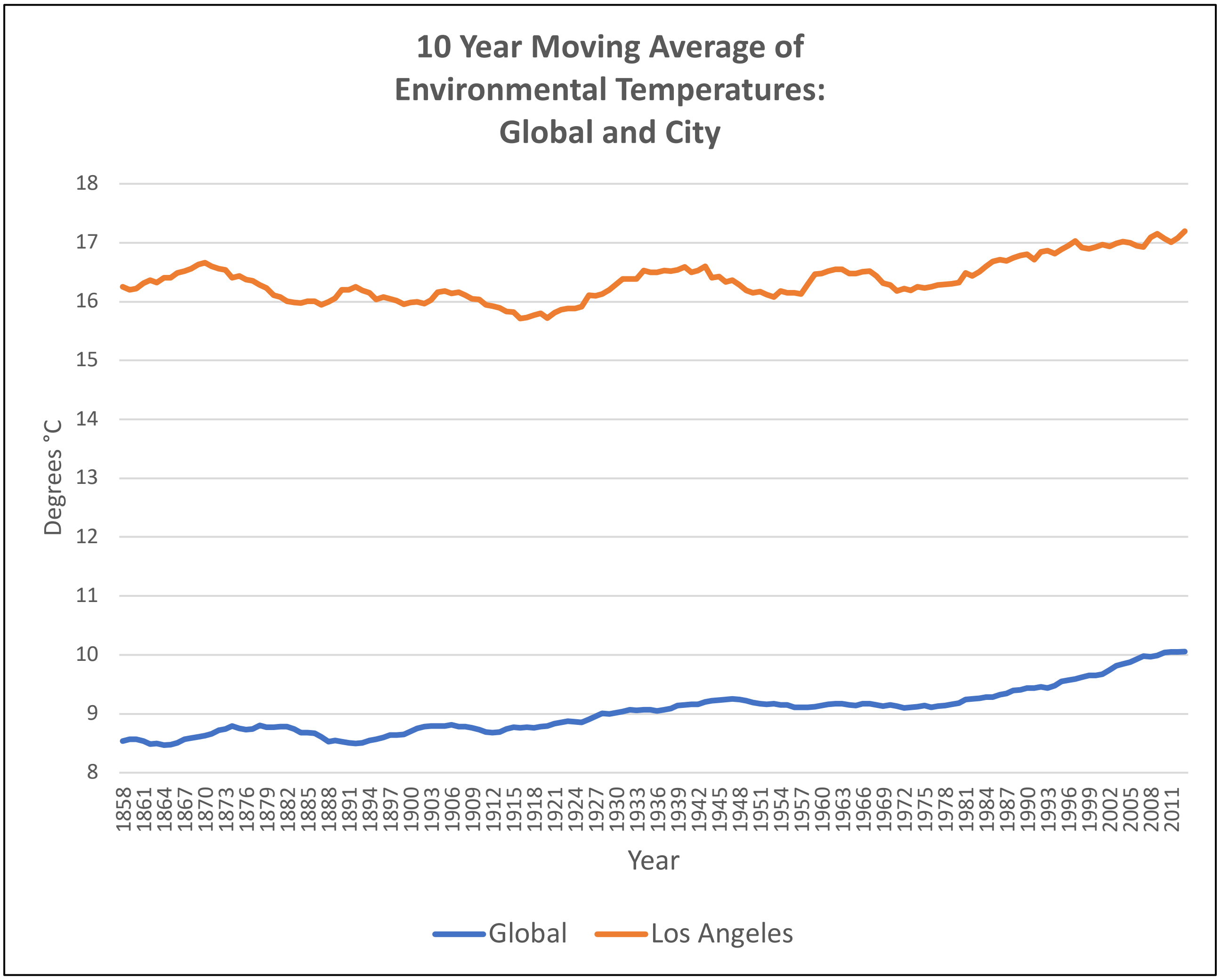
global city global and city Sheet3

4.3. Visualizing the trends

4.3.1. Key Considerations to bear in mind when charting/graphing data

When creating charts, graphs or any visual medium used to convey data, I follow the Tufte philosophy: keep it clean, simple and make sure the data presents in a way that convinces with as little additional clutter as possible. In this case following those rules was simple; the chart's data was very straightforward and required no graphics or callouts that might chance cluttering the graph's environment

4.3.2. Attach line chart, with titled axis and a legend



5. Make observations

5.1. Four similarities and/or differences in the trends

5.1.1. The similarity of the rising temperatures

Both global and local temperature mirrored each other in that they generally rose in temperature (and rose more pronounced in the last 50 years).

5.1.2. The similarity of the change in degrees

Both global and local temperatures rose respectively within 1.5 °C. The spread between the lowest temperature and highest temperature, both globally and locally, stayed within 1.5 °C.

5.1.3. The difference of global and local temperatures

The average global temperature for the years in study was 8.54 °C whereas the average local temperature was 18.86 °C. That was more than a 10°C difference in the two averages which seems like that could hold significance.

5.1.4. The difference of the smooth and jagged rises

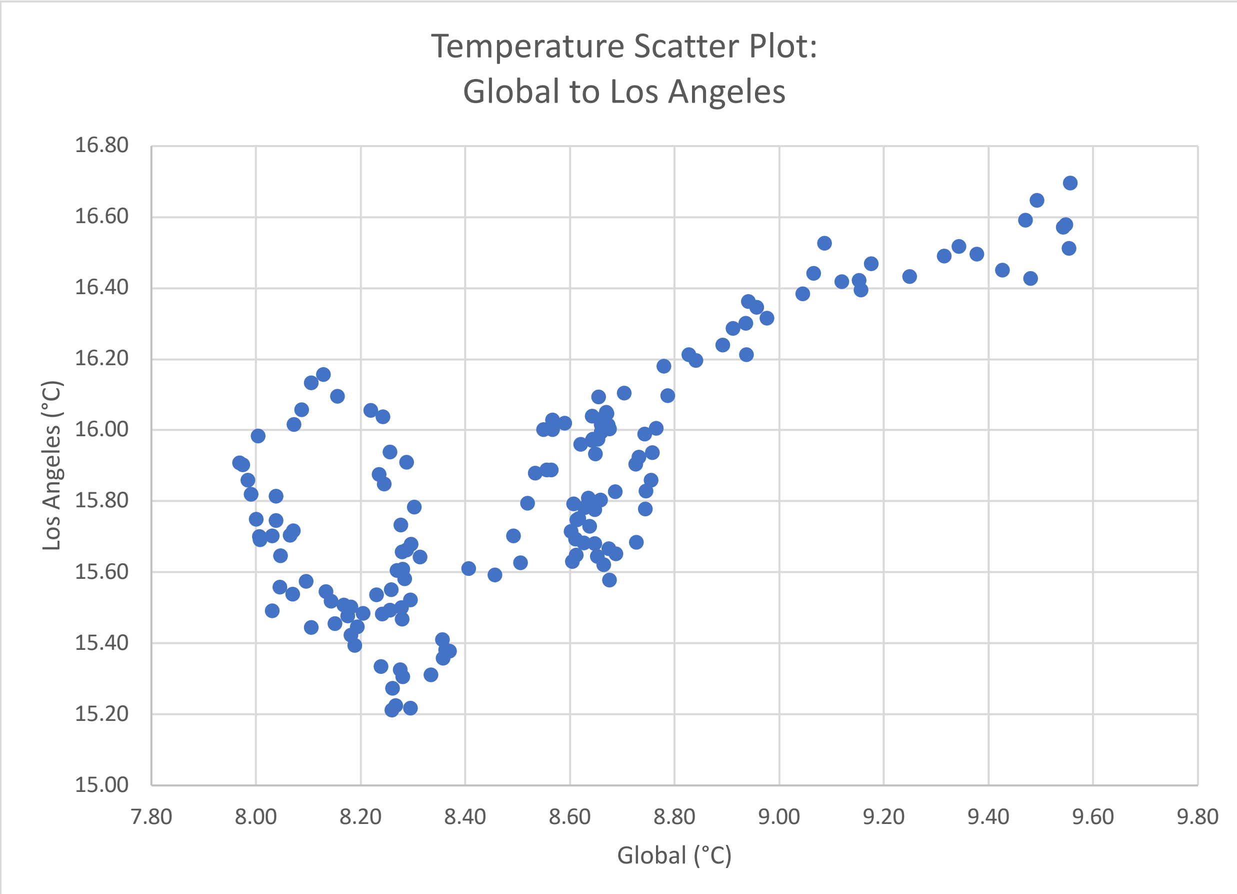
Even with the smoothing effects of the 10-year moving average created for both global and local temperatures, the former had a smoother rise in temperature as opposed to the local temperatures more jagged rise in temperature.

5.2. Additional observations

5.2.1. The correlation coefficient: can a linear relationship between global temperatures and Los Angeles temperature during the 155-year time period be established?

The line charts are interesting and useful to a degree, but they really only scratch the surface of the data. That's because they only visually describe the data. Determining the correlation coefficient between bivariate data can help determine whether a linear relationship exists between the two. In this case, we can determine if a linear relationship exists between global temperature and Los Angeles' local temperature for the years under study.

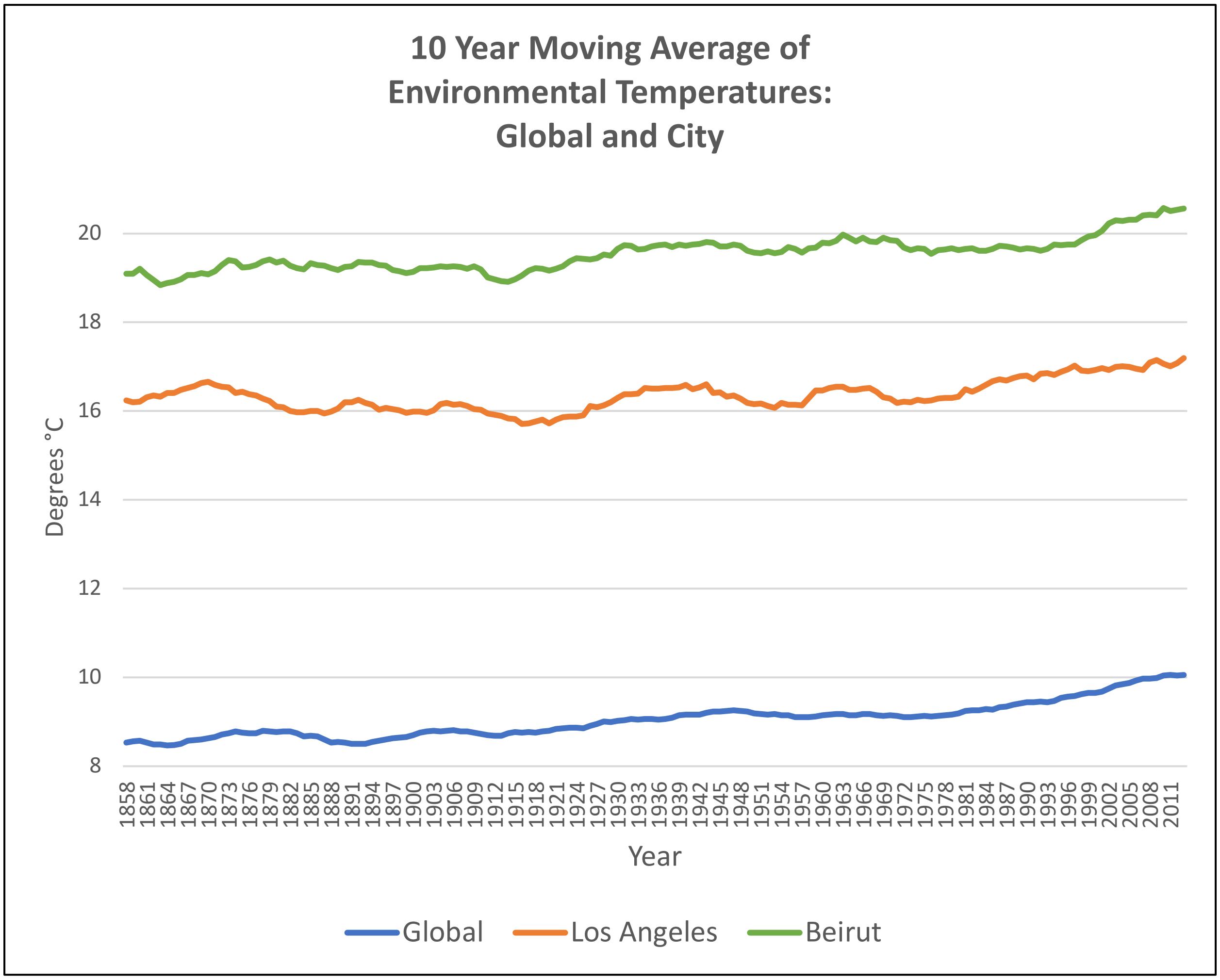
Sometimes a scatter plot of the bivariate data clearly shows linearity. Often times though, a scatter plot at best, shows a general direction of linearity if it appears existent at all. This seemed to be the case with the temperature data:



Visually, there seemed to be potential for some type of positive (albeit meandering) linear relationship. Calculating the correlation coefficient for the data would determine the existence or absence of that linear relationship with finality. Using a correlation coefficient calculating table I created in Excel, I entered the X & Y temperature data for the 155 years (n) of moving averages and determined the correlation coefficient was 0.764816124. The critical value when $n \geq 30$ is 0.361. The absolute value of the correlation coefficient is larger than the critical value therefore a linear relationship exists between global and the local temperature of Los Angeles.

Note: all supporting Excel documentation is available upon request.

5.2.2. Beirut, Lebanon: the additional city



I chose Beirut as an additional city to investigate. I chose Beirut because, upon research, I discovered that Beirut falls on the 34th North Longitudinal Parallel, the same as Los Angeles. I posited that perhaps they'd hold similar temperature data since they shared the same parallel. Though Beirut and Los Angeles were closer to each other than either was close to the average global temperature, visually, they didn't appear as close as I thought they would. Further analysis, of course outside the scope of this project, could prove with statistical significance whether how they "appear" to each other is a valid indicator of whether they are close to each other.