# Weather Trends

To gather the data required first a city in the data set had to be selected. In order to get list of available cities I first ran the following queries to first get how many cities were included for the United States (to understand the burden of manually scrolling) and to ultimately get the list of actual cities.

Select count(city) FROM city_list Where country Like 'United States'

Select * FROM city_list Where country Like 'United States'

As only 52 cities were returned, I decided I would scroll through the data and visibly make the decision, had the count of cities been significant I would have queried the table using a list of possible cities close to me to see what which of those were included. The closest city in the list ended up being Louisville. To be sure there were no other 'Louisville' in any other country in the dataset I ran the following query and received 1 as the result, confirming I need not worry about pulling data from the city_data table for more than one 'Lousiville'

Select count(city) FROM city_list Where city Like 'Louisville'

Now to pull the data I joined both the global and the city data on the year with a query and downloaded the CSV. Query used is as follows-

Select global_data.year, global_data.avg_temp As global_avg, city_data.avg_temp As louisville_avg from global_data LEFT JOIN city_data ON global_data.year = city_data.year Where city_data.city Like 'Louisville'

Although a bit of overkill for this exercise I decided to utilize python, along with the libraries pandas and matplotlib, for analysis and visualizations. After loading in the CSV file into a "DataFrame" I decided to generate 3 additional columns of the moving average – 10 year, 25 year and 100 year. To do this I used the built-in pandas rolling function and then the pandas mean function on the returned rolling object. The "rolling" function returns an object of pandas "window" which is a slice of data of the size you provide in the function. Example in this case to generate the 10 year moving averages a size of 10 was provided for the rolling function, as each record is for a year, and pandas generates a "window" for each year including record it is on and the 9 trailing records (ie if it was on 1920 it would include 1920, 1919,…1911). Then applying the mean function in pandas which calculates the mean of each window in the rolling object. Finally, for each of these moving averages which were calculated I joined them into the main DataFrame on the year.

One important note with the data is that in 1780 the city_data for Louisville, and at the least a few other cities I checked, was a blank value. In order to be able to calculate the moving averages without throwing out all data around this data point, which would be quite a bit for the 100 year moving average, I implemented the rolling function to allow for the minimum window to be 1 smaller than the working size, ie for the 100 year average it would still calculate as long as there was 99 records in the window. I made this decision as the effect of this missing record would reduce as the moving average size increased and would allow us to view a larger size of uninterrupted data when visualizing.

Finally, I added one additional column which was just the delta of the 100 year moving average so that we could view the change each year.

For visualization I generated 4 line graphs for each of the moving averages and the 100 year average delta. From the graphs the y-axis is the moving average temperature and x-axis is the year. This seemed to be the most straight-forward way to display the data being that it is a time series, and these line graphs show you the data moving through time.

Change Over Year (100 Year Moving Average)
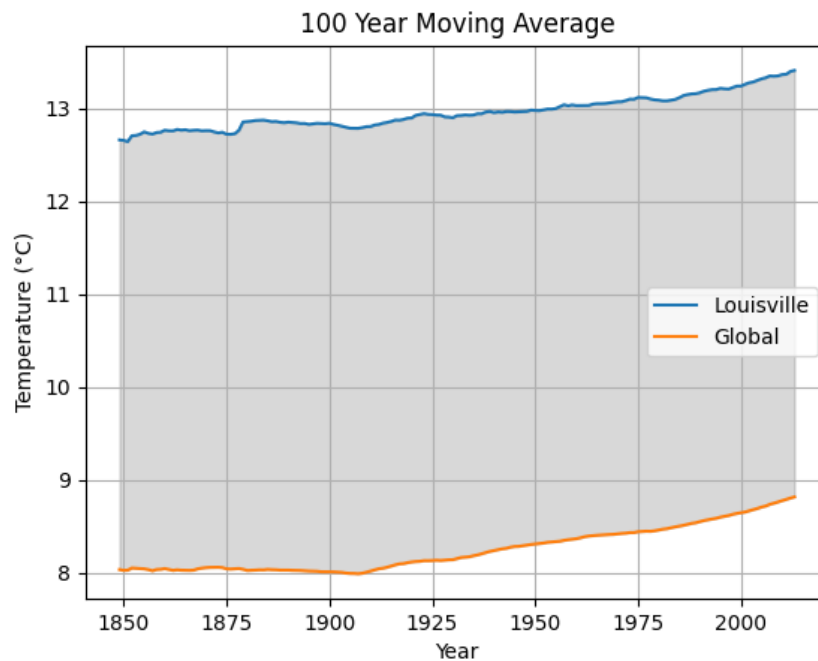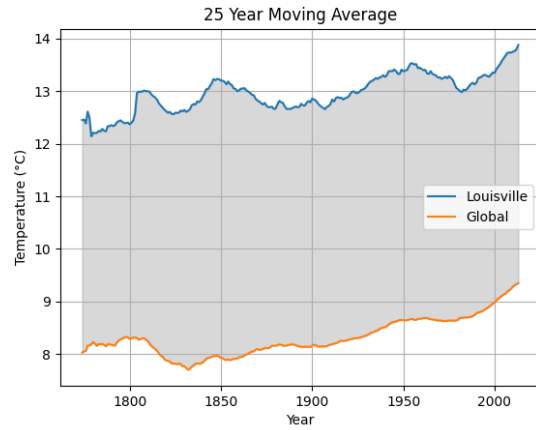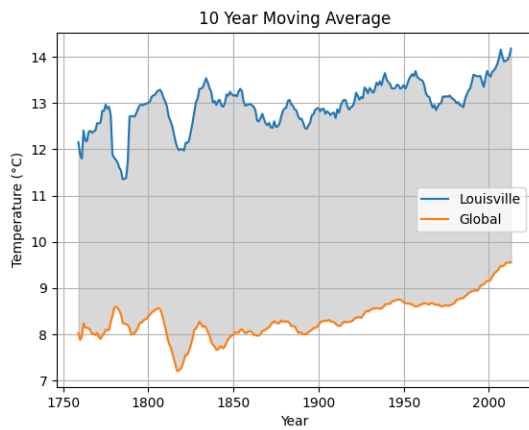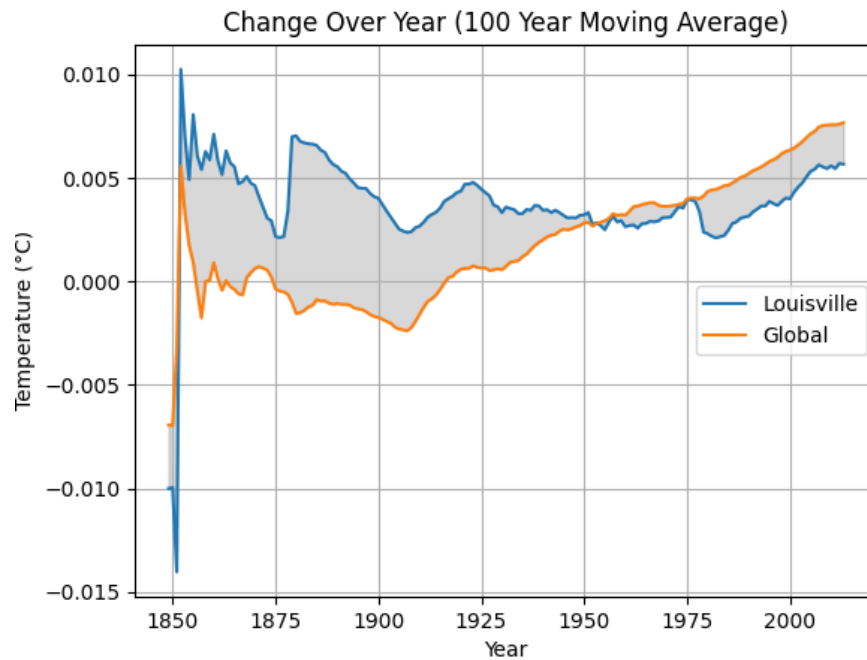
Observations

All three moving averages show the same upward trend in both global and city average yearly temperature. As the 100 year moving average shows the most smoothing I believe it is the graph that best shows the upward trend. One notable detail is you can see the increasing in temperature take off somewhere around 1910.

Focusing on the 100 year average the difference between the global and city temperature has stayed stable. When looking at the trend of upward movement the difference compares close to the 1850s difference, 2013 difference in global to city was 4.592 and 1850 was 4.614923.

Louisville temperature has been consistently warmer than the global temperature and that trend has stayed the same over the time of the data set.

Finally, the delta of 100 year moving average shows some interesting trends. Louisville temperature had a large change around the 1880s however the global temperature does not mirror this change and shows globally the change was a cool down. Additionally, Louisville had a temperature change each year greater than that of the global change however starting in around 1910 the global change started outpacing the Louisville change and sometime close to 1950 the global change became greater than the local yearly change. From 1980 and onward both show a similar change in temperature with the global temperature maintain a larger increase in average temperature.