

# Exploring Weather Trends

Project\_1\_Udacity

Shatakshi Shandilya

## Exploring Weather Trends:

### Content:

1. Goal of the Project
2. Extract Data
3. Explore Temperature datasets
4. Data Visualisation of datasets
5. Moving Average and Comparison
6. Summary
7. Reference

### Goal of this project:

To study the relationship between the Global temperature trends and the temperature trends in the city where I live i.e. San Jose(CA). Average Moving Temperature of 15 years for San Jose needs to be compared with Global Average Moving temperature of 15 years.

### Extract Data:

The data of interest is extracted from the following three tables available in database schema:

**city\_list:** This contains a list of cities in different countries. I found the city San Jose from this list.

**city\_data:** This contains the average recorded temperature(°C) every year for each city in the list.

**global\_data:** This contains the average recorded temperature (°C) every year globally.

**SQL Query to extract the San Jose and Global temperature data. Export the data to a CSV file.**

**SQL Query to look for San Jose(the city I live in):**

```
SELECT*  
FROM city_list where city like '%San% '  
;
```

**SQL Query to extract San Jose temperature record data:**

```
SELECT *  
FROM city_data where city= 'San Jose'  
;
```

**SQL Query to extract global data:**

```
SELECT *  
FROM global_data;
```

# EXPLORE TEMPERATURE DATASETS:

In [ ]:

```
#IMPORT Pandas, Numpy and matplotlib.pyplot for reading the csv and  
visualization of datasets
```

```
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as mplot
```

In [22]:

```
#Global Average Temperature Data
```

```
df=pd.read_csv('global_temp.csv')  
df.head()
```

Out[22]:

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47

In [3]:

```
df.tail()
```

Out[3]:

	year	avg_temp
261	2011	9.52
262	2012	9.51
263	2013	9.61
264	2014	9.57
265	2015	9.83

In [4]:

```
df.describe()
```

Out[4]:

	year	avg_temp
count	266.000000	266.000000
mean	1882.500000	8.369474
std	76.931788	0.584747
min	1750.000000	5.780000
25%	1816.250000	8.082500
50%	1882.500000	8.375000
75%	1948.750000	8.707500
max	2015.000000	9.830000

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 2 columns):
year      266 non-null int64
avg_temp  266 non-null float64
dtypes: float64(1), int64(1)
memory usage: 4.3 KB
```

In [6]:

```
#San Jose Average Temperature Data
df2=pd.read_csv('sj_temp.csv')
df2.head()
```

Out[6]:

	year	city	country	avg_temp
0	1849	San Jose	United States	14.12
1	1850	San Jose	United States	13.80
2	1851	San Jose	United States	14.39
3	1852	San Jose	United States	13.81
4	1853	San Jose	United States	14.40

In [7]:

```
df2.tail()
```

Out[7]:

	year	city	country	avg_temp
160	2009	San Jose	United States	15.02
161	2010	San Jose	United States	14.67
162	2011	San Jose	United States	14.50
163	2012	San Jose	United States	15.05
164	2013	San Jose	United States	16.23

In [26]:

```
df2.describe()
```

Out[26]:

	year	avg_temp	ma5	ma10	ma15
count	165.000000	165.000000	161.000000	156.000000	151.000000
mean	1931.000000	14.450788	14.444335	14.440968	14.434066
std	47.775517	0.504885	0.344856	0.301796	0.275209
min	1849.000000	13.220000	13.826000	13.951000	13.975333
25%	1890.000000	14.120000	14.190000	14.201750	14.213333
50%	1931.000000	14.390000	14.414000	14.413000	14.396000
75%	1972.000000	14.760000	14.692000	14.607000	14.560667
max	2013.000000	16.230000	15.244000	15.148000	15.112000

```
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 165 entries, 0 to 164  
Data columns (total 7 columns):  
year      165 non-null int64  
city      165 non-null object  
country   165 non-null object  
avg_temp  165 non-null float64  
ma5       161 non-null float64  
ma10      156 non-null float64  
ma15      151 non-null float64  
dtypes: float64(4), int64(1), object(2)  
memory usage: 9.1+ KB
```

The preliminary datasets analysis tells us that there are no missing data in either of the datasets. Global dataset shows average temperature every year from year 1750 upto year 2015 while San Jose dataset shows average temperature from 1859 till year 2013. It can be noticed that the average temperature in San Jose is higher than temperature recorded globally. It does make sense as San Jose has a warm-summer Mediterranean climate, lying near Pacific Ocean. It has warm to hot, long dry summers and mild to cool, short wet winters. Lying in the San Francisco Bay Area, San Jose is affected by extreme heat waves in Summer as well as Fall in recent years which contributes to higher average temperature.

# VISUALISATION OF DATASETS:

In [10]:

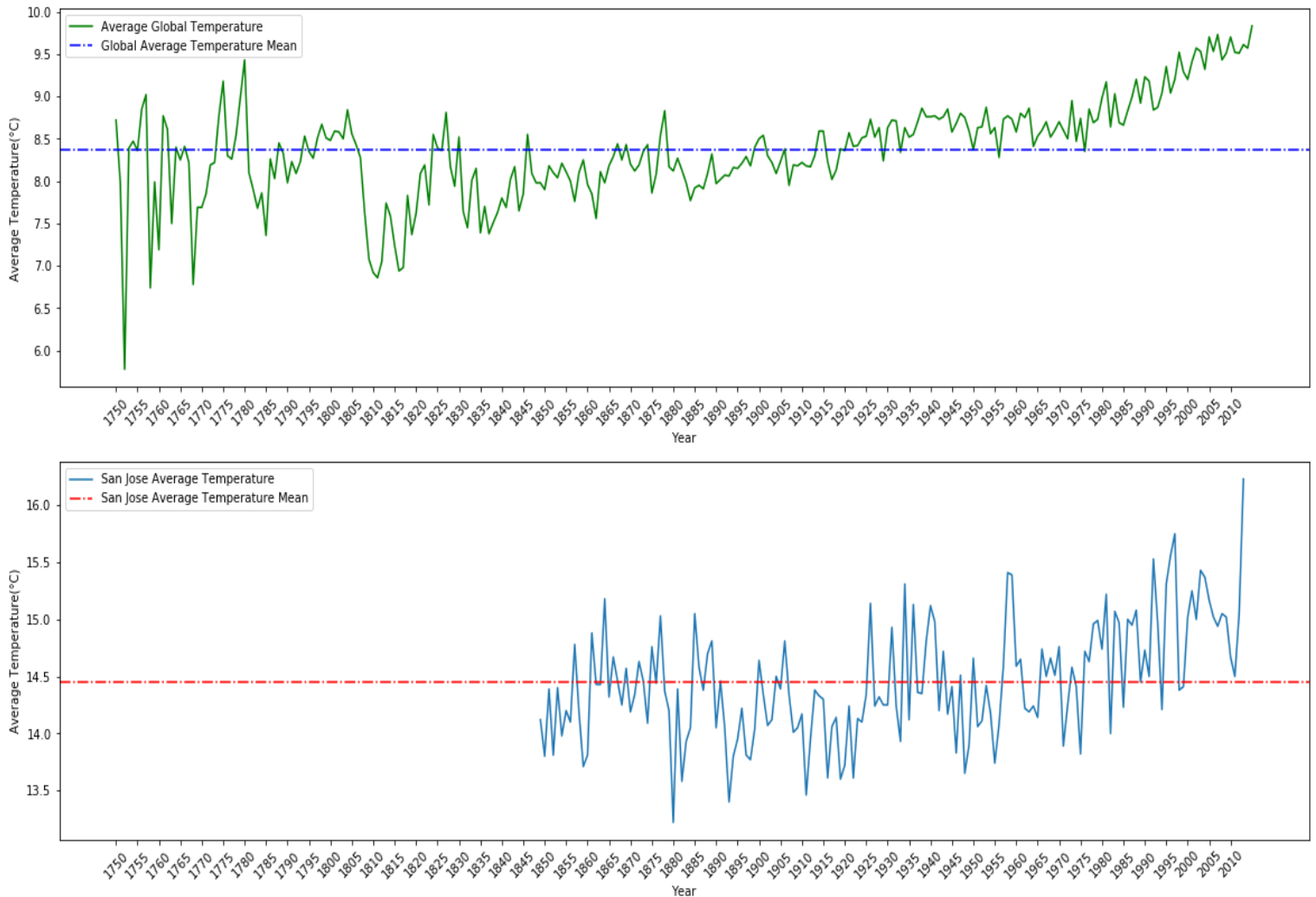
```
fig1=plt.figure(figsize=(20,12))
plt.suptitle('Global Average temperature vs San Jose Average Temperature')

ax1=plt.subplot(211)
plt.plot(df.year,df.avg_temp,color='green',label='Average Global Temperature')
plt.axhline(df.avg_temp.mean(),label='Global Average Temperature
Mean',linestyle='dashdot',color='blue')
plt.xlabel('Year')
plt.ylabel('Average Temperature(°C)')
plt.legend(loc=0)
plt.xticks(np.arange(1750,2015,step=5),rotation=45)

ax2=plt.subplot(212,sharex=ax1)
plt.plot(df2.year,df2.avg_temp,label='San Jose Average Temperature')
plt.axhline(df2.avg_temp.mean(),color='red',linestyle='dashdot',label='San
Jose Average Temperature Mean')
plt.xlabel('Year')
plt.ylabel('Average Temperature(°C)')
plt.legend(loc=0)
plt.xticks(np.arange(1750,2015,step=5),rotation=45)
plt.show()
```

*####sharex=ax1 is used in ax2 subplot as we need the x axis to show all the years available in ax1 that are not present in ax2*

Global Average temperature vs San Jose Average Temperature



It is noted from the plot above that both San Jose and Global Average temperature have dramatical changes in their average temperature every year. The fluctuation in the average temperature is noticeable every year. The temperatures we experience locally and in short periods can fluctuate significantly due to predictable cyclical events (night and day, summer and winter) and hard-to-predict wind and precipitation patterns.

To observe long term trends of the average temperature, Moving averages are calculated and used. This smooths out the data eliminating short term fluctuations. 5 years, 10 years and 15 years averages of average temperature data are calculated for better understanding of the temperature trends globally and locally.

# MOVING AVERAGES AND COMPARISON:

In [15]:

```
#Moving Average and Comparison

n=5
df['ma5']=df['avg_temp'].rolling(window=n).mean()
df2['ma5']=df2['avg_temp'].rolling(window=n).mean()

n=10
df['ma10']=df['avg_temp'].rolling(window=n).mean()
df2['ma10']=df2['avg_temp'].rolling(window=n).mean()

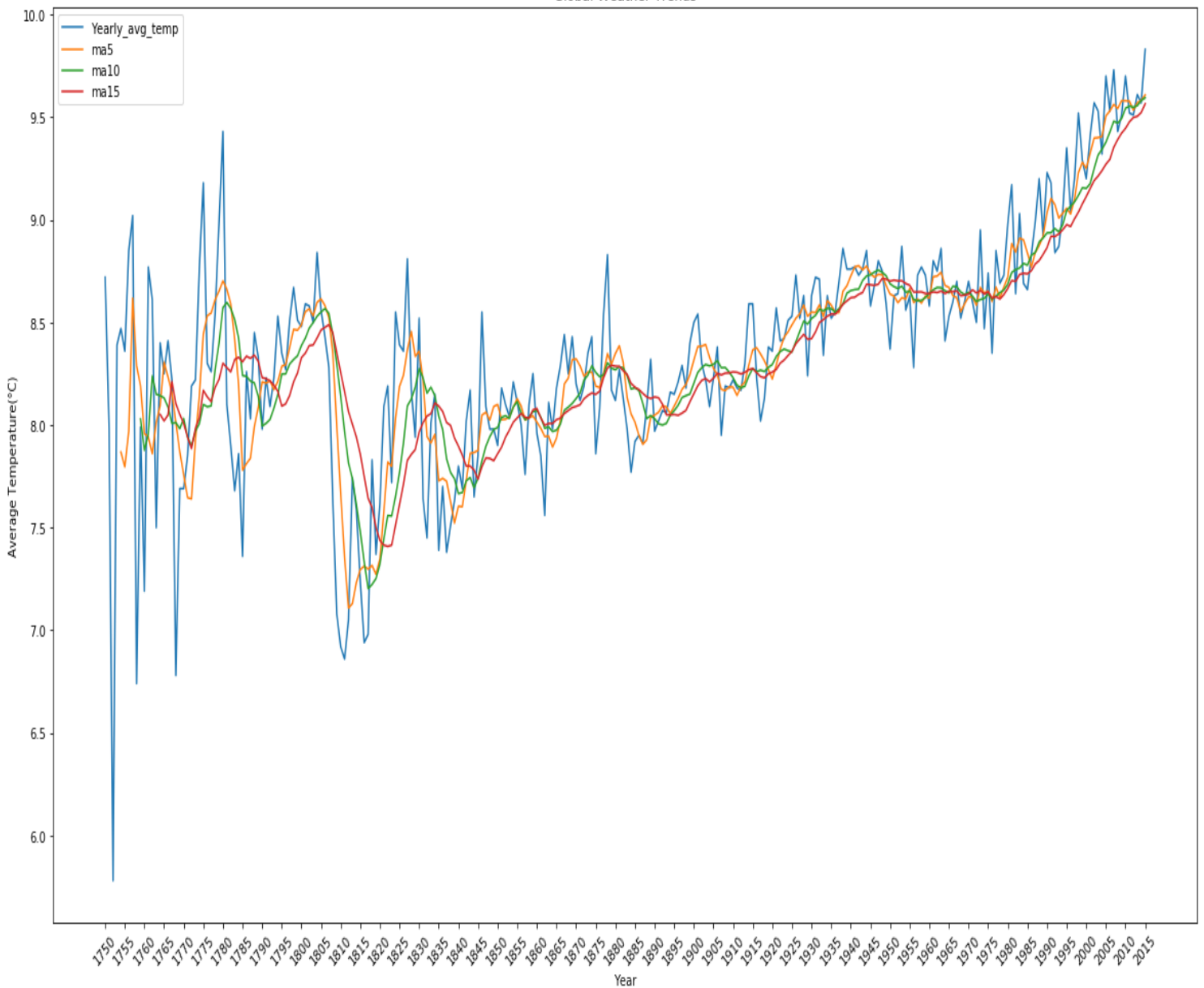
n=15
df['ma15']=df['avg_temp'].rolling(window=n).mean()
df2['ma15']=df2['avg_temp'].rolling(window=n).mean()
```

In [17]:

```
#Plot Global Average Temperature trends
matplotlib.figure(figsize=(20,12))
matplotlib.plot(df.year,df.avg_temp,label="Yearly_avg_temp")
matplotlib.plot(df.year,df.ma5,label="ma5")
matplotlib.plot(df.year,df.ma10,label="ma10")
matplotlib.plot(df.year,df.ma15,label="ma15")
matplotlib.title('Global Weather Trends')
matplotlib.xlabel('Year')
matplotlib.xticks(np.arange(1750,2020,step=5),rotation=45)
matplotlib.ylabel('Average Temperature(°C)')
matplotlib.legend()
matplotlib.show()
```



Global Weather Trends



when the moving averages of average temperature for 5, 10 and 15 years are plotted together, we find that the moving average for 15 years(red line) gives us a smoother line and helps us understand the temperature trend better. The red line here tells us that the temperature has been rising globally since around 1820. We notice that the average temperature has increased by at least 2 degrees globally over the last 200 years.

```
#Plot San Jose Average Temperature Trends
```

```
mpl.figure(figsize=(18,10))
```

```
mpl.plot(df2.year,df2.avg_temp,label="yearly_avg_temp")
```

```
mpl.plot(df2.year,df2.ma5,label="ma5")
```

```
mpl.plot(df2.year,df2.ma10,label="ma10")
```

```
mpl.plot(df2.year,df2.ma15,label="ma15")
```

```
mpl.title('San Jose Weather Trends')
```

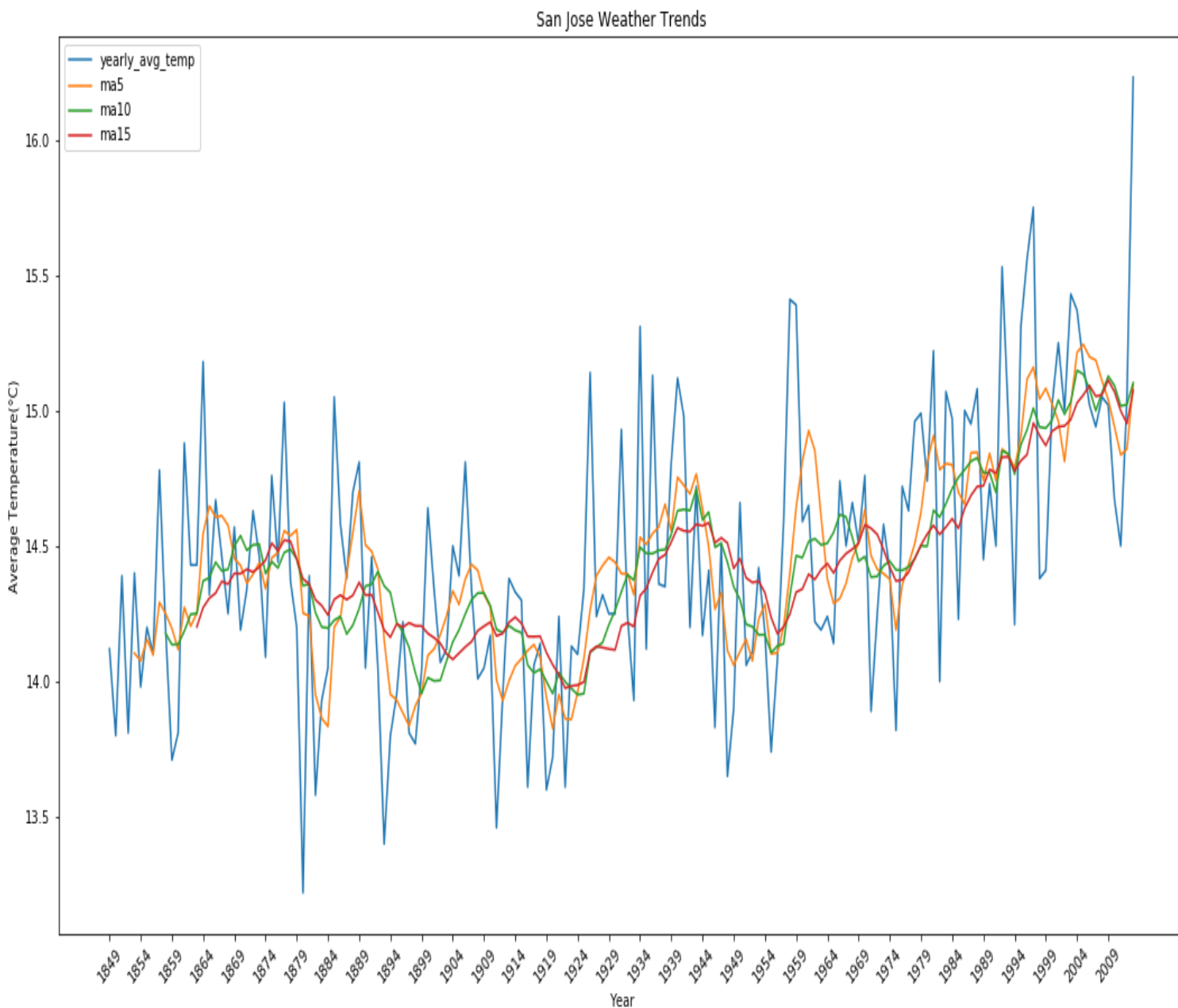
```
mpl.xlabel('Year')
```

```
mpl.xticks(np.arange(1849,2013,step=5),rotation=45)
```

```
mpl.ylabel('Average Temperature(°C)')
```

```
mpl.legend()
```

```
mpl.show()
```



The red line here gives a clear picture of the change in the average temperature since 1849. We can see that average temperature has increased by 1 degree over the last 170 years. The increase in temperature can be attributed to Global Warming caused by the increased Greenhouse emissions in the city throughout the time period. The typical home in San Jose produced nearly 9,000 pounds of carbon dioxide every month, responsible for a quarter of all greenhouse emissions. Thus, the city of San Jose will become the US's largest city to go all electric requiring electrification of all new residential construction since 2019. Big multi-family buildings and single-family homes must be built without natural gas. This would help us reduce the greenhouse emissions and would also help in reduction of Global Warming in future.

Although we have global Average Temperature data from 1750 to 2015, we have data only from year 1849, we do not have data for San Jose average temperature for the years 1750 to 1848. Thus to compare the 15-year moving average of both datasets, I create a new dataset for Global Average Temperature from the year 1864(1849+15).

In [20]:

```
#Global data from year 1864(1849+15)
```

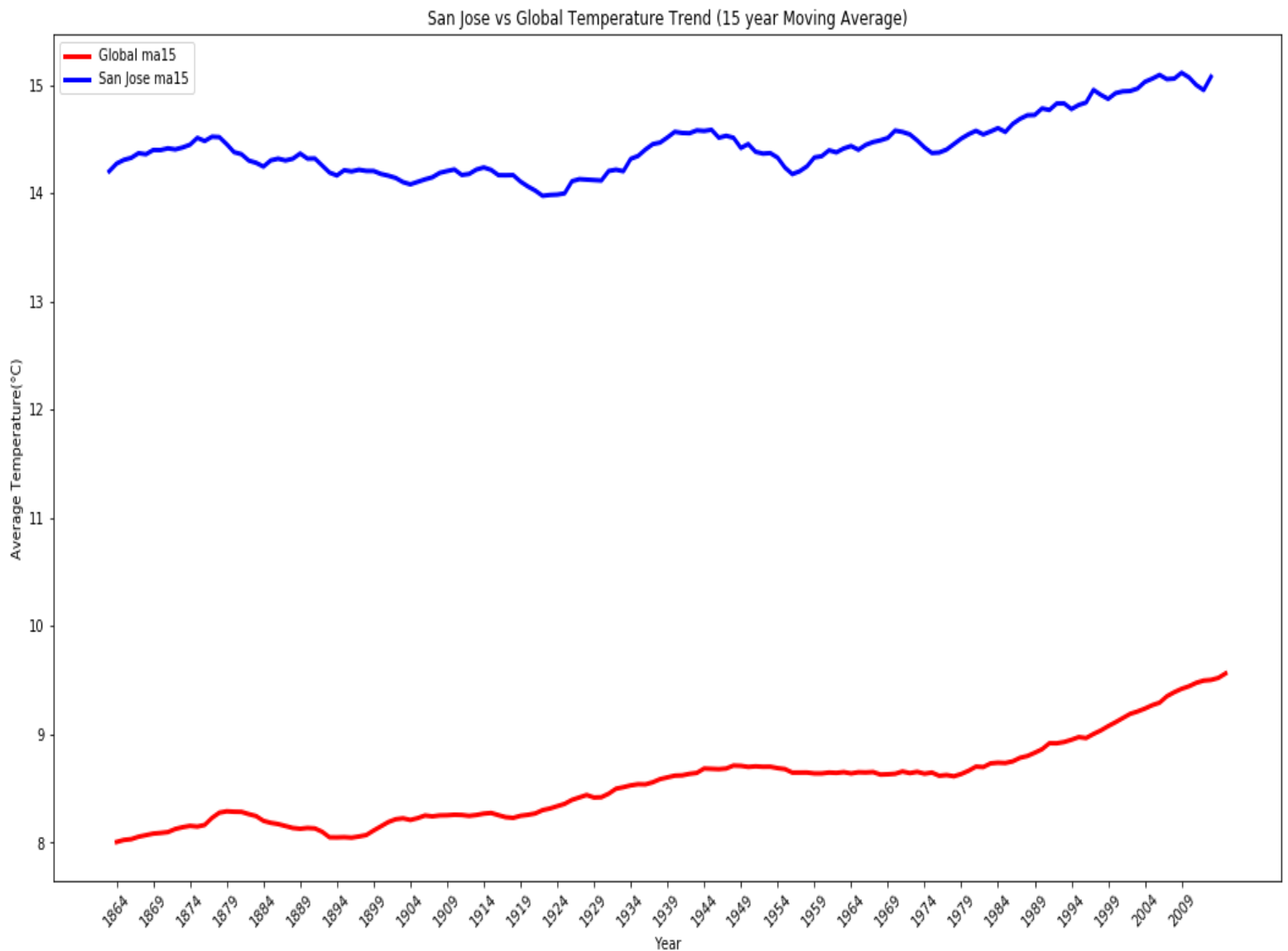
```
df_new=df[df.year>=1864]
df_new.head()
```

Out[20]:

	year	avg_temp	ma5	ma10	ma15
114	1864	7.98	7.892	7.968	8.007333
115	1865	8.18	7.936	7.975	8.026000
116	1866	8.29	8.024	8.004	8.033333
117	1867	8.44	8.200	8.072	8.056000
118	1868	8.25	8.228	8.087	8.070000

In [21]:

```
matplotlib.figure(figsize=(18,10))
matplotlib.title("San Jose vs Global Temperature Trend (15 year Moving Average)")
matplotlib.plot(df_new.year,df_new.ma15,label="Global ma15",color='red',linewidth=3)
matplotlib.plot(df2.year,df2.ma15,label="San Jose ma15",color='blue',linewidth=3)
matplotlib.legend()
matplotlib.xlabel('Year')
matplotlib.ylabel('Average Temperature(°C)')
matplotlib.xticks(np.arange(1864,2013,step=5), rotation =45)
matplotlib.show()
```



The red line here shows how average temperature changed globally while the red line shows the changes in average temperature from year 1849 till year 2013. It can be noticed that the average temperature in San Jose is at least 6 degrees higher throughout than temperature recorded globally. It does make sense as San Jose has a warm-summer Mediterranean climate, lying near Pacific Ocean. It has warm to hot, long dry summers and mild to cool, short wet winters. Lying in the San Francisco Bay Area, San Jose is affected by extreme heat waves in Summer as well as Fall in recent years which contributes to higher average temperature. It is noticeable that the average temperature has increased by 1 degree since 1849, in San Jose and by 1.5 degrees globally. The increase in temperature can be attributed to Global Warming caused by the increased Greenhouse emissions in the world through the time period. We also notice that the peak of average temperature was around 1945 both globally and locally before another peak that is still to be reached (the average temperature seems to be rising gradually in recent years). The peak around the year 1945 can be attributed to greenhouse gas pollution caused as industrialized cities and nations continued burning coal to power factories and trains. The

environmental pollution and massive oil production in World War II also led to increase in the atmospheric temperature. The recent increase in the average temperature is being caused by increased fossil fuel usage. Carbon dioxide, a byproduct of fossil fuel combustion, is the principal greenhouse gas contributing to global warming in the recent decades. Other greenhouse gases including methane, nitrous oxide, and a number of industrial-process gases also are important contributors to the increase in the average temperature

# Summary:

Analysis of the global and local Average temperature datasets can help me answer the following questions by Udacity:

1. Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?

On average, my city (**San Jose**) is around **6 degrees hotter** than the global average. It does make sense as San Jose has a warm-summer Mediterranean climate, lying near Pacific Ocean. It has warm to hot, long dry summers and mild to cool, short wet winters. Lying in the San Francisco Bay Area, San Jose is affected by extreme heat waves in Summer as well as Fall in recent years which contributes to higher average temperature.

The difference in the temperature locally and globally has been around 6 degrees and has been **consistent** over time. With similar ups and downs through 170 years, the only **exceptions are found through 1884-1889 and through 1920-1925**. In the time period between 1884 and 1889, average temperature globally seems to be decreasing while in San Jose, the average temperature was increasing. However the situation was reversed in the time period 1920-1925, when the average temperature in San Jose was increasing, with average temperature increasing globally.

2. "How do the changes in your city's temperatures over time compare to the changes in the global average?"

We observe that the average temperature moved in the **same direction** both in San Jose and globally, increasing by around 1.5 degrees over the years. It is noticed that the average temperature has increased by 1 degree since 1849, in San Jose and by 1.5 degrees globally.

The increase in temperature can be attributed to Global Warming caused by the increased Greenhouse emissions in the world through the time period.

We also notice that the **peak of average temperature was around 1945 both globally and locally** before another peak that is still to be reached (the average temperature seems to be rising gradually in recent years). The peak around the year 1945 can be attributed to greenhouse gas pollution caused as industrialized cities and nations continued burning coal to power factories and trains. The environmental pollution and massive oil production in World War II also led to increase in the atmospheric temperature. The recent increase in the average temperature is being caused by increased fossil fuel usage. We see that the average temperature moved in the same direction both in San Jose and globally. The only exceptions are found through 1884-1889 and through 1920-1925. Through 1884-1889, Average Temperature globally seems to be decreasing while in San Jose, the temperature is increasing. However the situation was reversed in the time period 1920-1925.

3. What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?

It is noticed through the line plot that the average temperature has increased by 1 degree since 1849, in San Jose and by 1.5 degrees globally. Though there have been ups and downs in average temperature both locally and globally, the overall temperature trend is that of a **rise** in temperature.

The world seems to be **getting hotter** and the increase in temperature can be attributed to Global Warming caused by the increased Greenhouse emissions in the world through the time period.

The trend of rise in temperature has been **consistent** in spite of the ups and down over the last few hundred years. We notice that the peak of average temperature was around 1945 both globally and locally before another peak that is still to be reached (the average temperature seems to be rising gradually in recent years). The peak around the year 1945 can be attributed to greenhouse gas pollution caused as industrialized cities and nations continued burning coal to power factories and trains. The environmental pollution and massive oil production in World War II also led to increase in the atmospheric temperature.

The recent increase in the average temperature is being caused by increased fossil fuel usage. Carbon dioxide, a byproduct of fossil fuel combustion, is the principal greenhouse gas contributing to global warming in the recent decades. Other greenhouse gases including methane, nitrous oxide, and a number of industrial-process gases also are important contributors to the increase in the average temperature.

# References:

<https://stackoverflow.com/questions/38666527/what-is-the-necessity-of-plt-figure-in-matplotlib>

<https://stackoverflow.com/questions/17071871/how-to-select-rows-from-a-dataframe-based-on-column-values>

<https://stackoverflow.com/questions/19125722/adding-a-legend-to-pyplot-in-matplotlib-in-the-simplest-manner-possible>

[https://matplotlib.org/3.1.0/api/\\_as\\_gen/matplotlib.pyplot.legend.html](https://matplotlib.org/3.1.0/api/_as_gen/matplotlib.pyplot.legend.html)

[https://matplotlib.org/3.3.1/api/\\_as\\_gen/matplotlib.pyplot.hlines.html](https://matplotlib.org/3.3.1/api/_as_gen/matplotlib.pyplot.hlines.html)

<https://stackoverflow.com/questions/3584805/in-matplotlib-what-does-the-argument-mean-in-fig-add-subplot111>

<https://towardsdatascience.com/complete-guide-to-data-visualization-with-python-2dd74df12b5e>

<https://hackersandslackers.com/plotting-data-seaborn-pandas/>