udacity_weather_trends

January 17, 2018

1 Udacity - Explore Weather Trends

The goal of this notebook is to create visualizations and present information describing the similarities and differences between global temperature trends and temperature trends in San Francisco, CA.

Temperature data has been extracted from the Udacity DB and put into files specific to location domain: San Francisco and Global Temps, respectively.

Global temperature records exist back well into the 1700s, where as data from San Francisco only started in the mid 1800s.

Temperature data was originally in Celsius and will be converted to Fahrenheit with this formula

Summary The first part of this notebook takes the raw data and make observatios against it. The second part of the notebook will convert the `avg_temp` columns of both datasets to a 7 day moving average (rolling mean) and make further observations.

I used a pandas dataframe to calculate the moving averages.

I used python (pandas/jupyter/matplotlib) to analyze the data

```
In [145]: import pandas as pd
          import matplotlib.pyplot as plt
          %matplotlib inline
```

1.0.1 SF Weather Data

Data based off of average temps and the year in ascending order

```
In [146]: sf_weather_data = pd.read_csv('./data/san_francisco_temps.csv')
          sf_weather_data.head()
```

```
Out[146]:    year  avg_temp
          0  1849    14.12
          1  1850    13.80
          2  1851    14.39
          3  1852    13.81
          4  1853    14.40
```

```
In [147]: # Lets convert the avg_temp to Fahrenheit from Celsius
          sf_weather_data['avg_temp'] = (sf_weather_data['avg_temp'] * 1.8 + 32)
```

```
In [148]: sf_weather_plot_data = sf_weather_data.set_index('year')
          sf_weather_plot = sf_weather_plot_data.plot(title="Avg. Temperature By Year", legend=
          sf_weather_plot.set_xlabel('Year')
          sf_weather_plot.set_ylabel('Temperature (C)')

Out[148]: Text(0,0.5,'Temperature (C)')
```
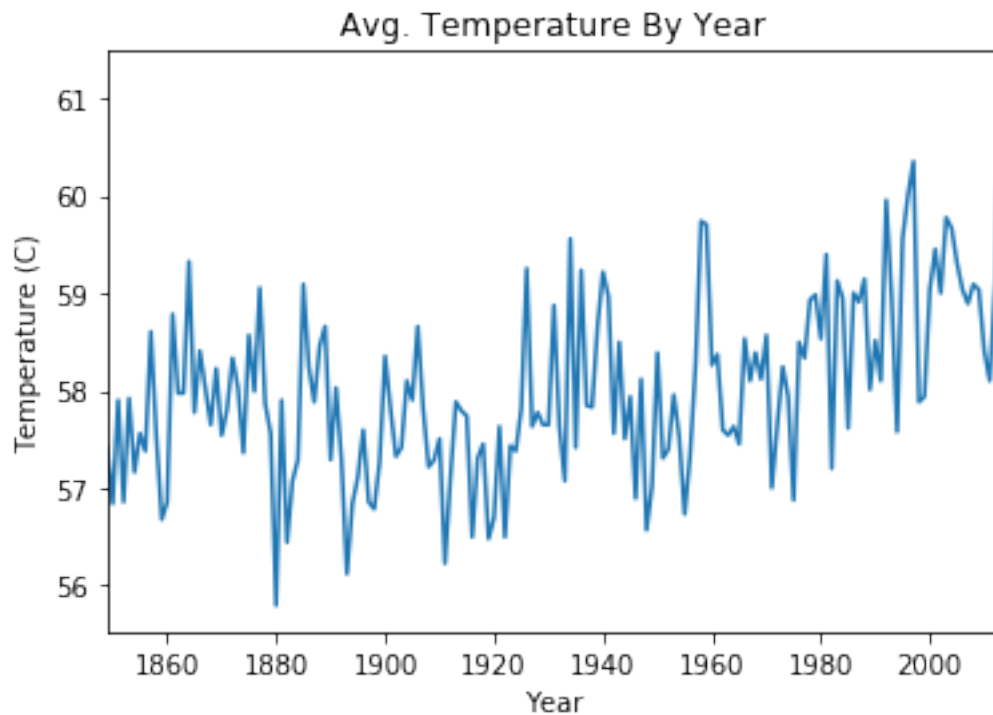

1.0.2 Global Weather Data

Data based off of average temps and the year recorded, in ascending order

```
In [149]: global_weather_data = pd.read_csv('./data/global_temps.csv')
          global_weather_data.head()

Out[149]:    year  avg_temp
          0  1750      8.72
          1  1751      7.98
          2  1752      5.78
          3  1753      8.39
          4  1754      8.47
```

```
In [150]: # Lets convert the avg_temp to Fahrenheit from Celsius
          global_weather_data['avg_temp'] = (global_weather_data['avg_temp'] * 1.8 + 32)
```

2

```
In [151]: global_weather_plot_data = global_weather_data.set_index('year')
          global_weather_plot = global_weather_plot_data.plot(title="Avg. Temperature By Year"
          global_weather_plot.set_xlabel('Year')
          global_weather_plot.set_ylabel('Temperature (C)')

Out[151]: Text(0,0.5,'Temperature (C)')
```
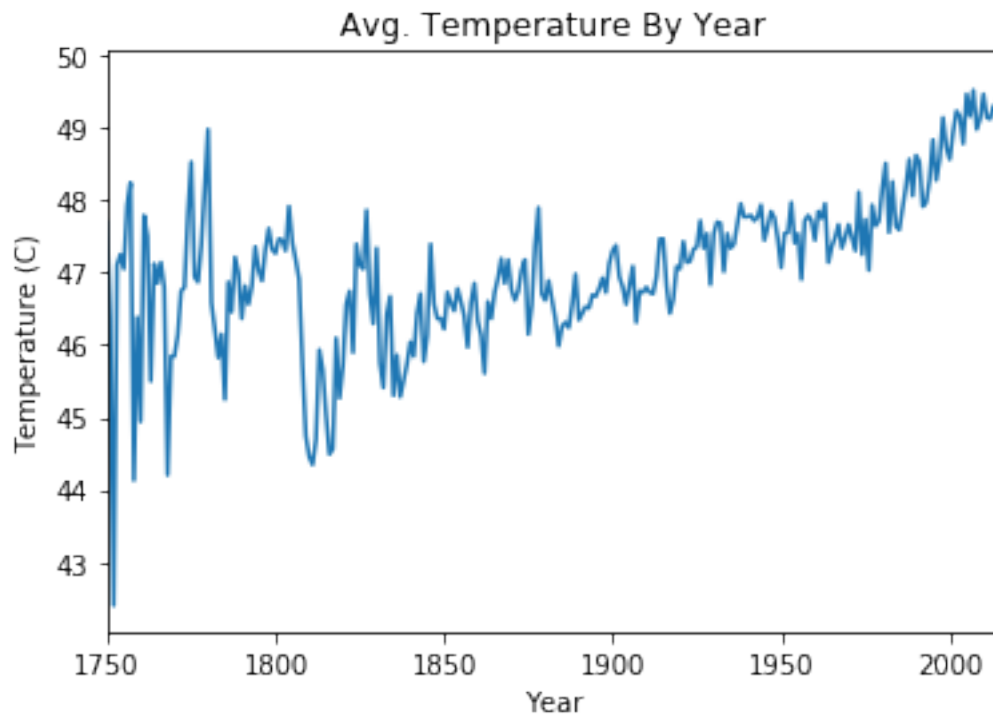

2 Observations

Here we are plotting both sets of data on the same dataframe to visual see any similarities or differences in the data.

I have used this stackoverflow on creating labels for figure with multiple plots which is why there is some weird padding going on
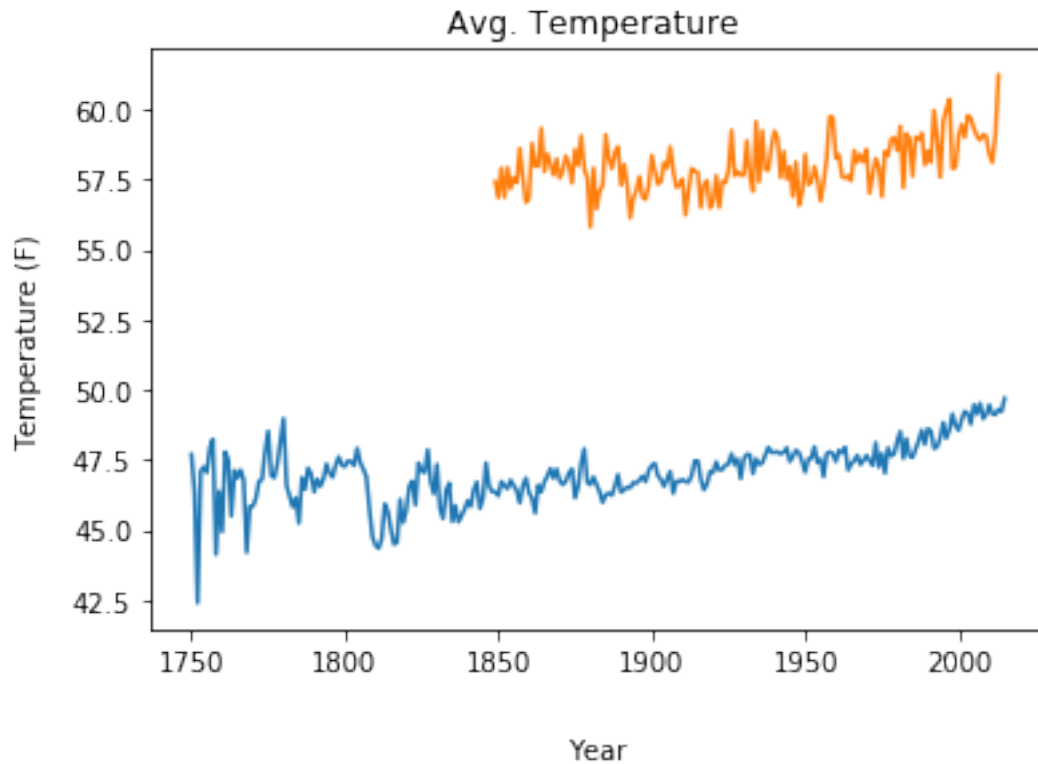
```
In [152]: fig = plt.figure()

          for data in [global_weather_plot_data, sf_weather_plot_data]:
              plt.plot(data)

          labels_subplot = fig.add_subplot(111, frame_on=False)
          labels_subplot.set_xticks([])
          labels_subplot.set_yticks([])
          labels_subplot.set_title('Avg. Temperature ')
```

```
        labels_subplot.set_xlabel('Year', labelpad=40)
        labels_subplot.set_ylabel('Temperature (F)', labelpad=40)
```

Out[152]: Text(0,0.5,'Temperature (F)')

Similarities

1. Both SF and Global temperatures have been steadily rising since data has been recorded

Differences

1. SF average temperatures are way warmer than the global average temperature.
2. Average temperatures in SF fluctuate (change) more on a per year basis than Global average temperatures

Other Observations

1. Global temperatures in the beginning of the data set (1750s-1800) seem to have been more sparadic, where as time went on the temperatures appear to smooth out (lack volatility)

3 Observations (with moving average)

```
In [153]: global_weather_data['avg_temp'] = global_weather_data['avg_temp'].rolling(7).mean()
          global_weather_data = global_weather_data[global_weather_data['year'] > 1917]
          global_weather_plot_data = global_weather_data.set_index('year')

In [154]: sf_weather_data['avg_temp'] = sf_weather_data['avg_temp'].rolling(7).mean()
          sf_weather_data = sf_weather_data[sf_weather_data['year'] > 1917]
          sf_weather_plot_data = sf_weather_data.set_index('year')

In [155]: fig = plt.figure()

          for data in [global_weather_plot_data, sf_weather_plot_data]:
              plt.plot(data)

          labels_subplot = fig.add_subplot(111, frame_on=False)
          labels_subplot.set_xticks([])
          labels_subplot.set_yticks([])
          labels_subplot.set_title('Avg. Temperature for Years since 1918')
          labels_subplot.set_xlabel('Year', labelpad=40)
          labels_subplot.set_ylabel('Temperature (F)', labelpad=40)

Out[155]: Text(0,0.5,'Temperature (F)')
```
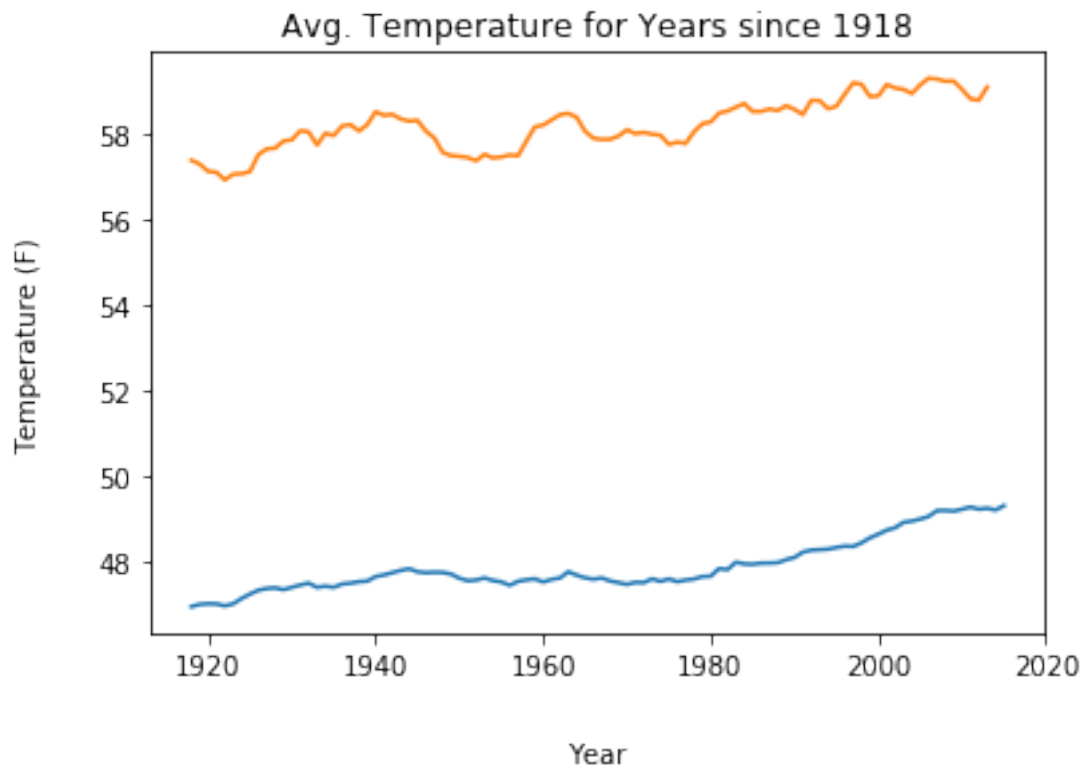
Other Observations

1. Global temperature remained steady from the late 1930s to about the 1980s and have been on the incline since