Exploring Weather Trends

Contents

Target of this Project

In this project,I will analyze the temperatures of Houston(TX), where I live, along with the global temperature and compare trends of them.

Extra the Data

In the Database Schema, there are three tables in the database:
city_list - This contains a list of cities and countries in the database. Look through them in order to find the city nearest to you.
city_data - This contains the average temperatures for each city by year (ºC).
global_data - This contains the average global temperatures by year (ºC).

- **Write a SQL query to extract the global data. Export to CSV.**

  ```
  SELECT *
  FROM global_data
  ;
  ```

- **Write a SQL query to extract the Houston City data. Export to CSV.**

  ```
  SELECT *
  FROM city_data
  WHERE city = 'Houston'
  ;
  ```

Explore temperature datasets

Import several libraries and tools for data analysis

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style

style.use("fivethirtyeight")
```

In [2]:
```
# Global temperature average data set
global_data = pd.read_csv('Global_Temp.csv')

# Houston TX temperture average data set
houston_data = pd.read_csv('Houston_Temp.csv')
```

In [3]:
```
global_data.head()
```

Out[3]:

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

In [4]:
```
global_data.tail()
```

Out[4]:

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

In [5]:
```
global_data.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: 4.2 KB
```

In [6]:
```
global_data['avg_temp'].describe()
```

Out[6]:
```
count    266.000000
mean       8.369474
std        0.584747
min        5.780000
25%        8.082500
50%        8.375000
75%        8.707500
max        9.830000
Name: avg_temp, dtype: float64
```

In [7]:
```
houston_data.head()
```

Out[7]:

	year	city	country	avg_temp
0	1820	Houston	United States	19.11
1	1821	Houston	United States	19.57
2	1822	Houston	United States	20.05
3	1823	Houston	United States	19.62
4	1824	Houston	United States	20.19

In [8]:
```
houston_data.tail()
```

Out[8]:

	year	city	country	avg_temp
189	2009	Houston	United States	21.11
190	2010	Houston	United States	20.43
191	2011	Houston	United States	21.69
192	2012	Houston	United States	21.86
193	2013	Houston	United States	22.28

In [9]:
```
houston_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194 entries, 0 to 193
Data columns (total 4 columns):
year        194 non-null int64
city        194 non-null object
country     194 non-null object
avg_temp    194 non-null float64
dtypes: float64(1), int64(1), object(2)
memory usage: 6.1+ KB
```

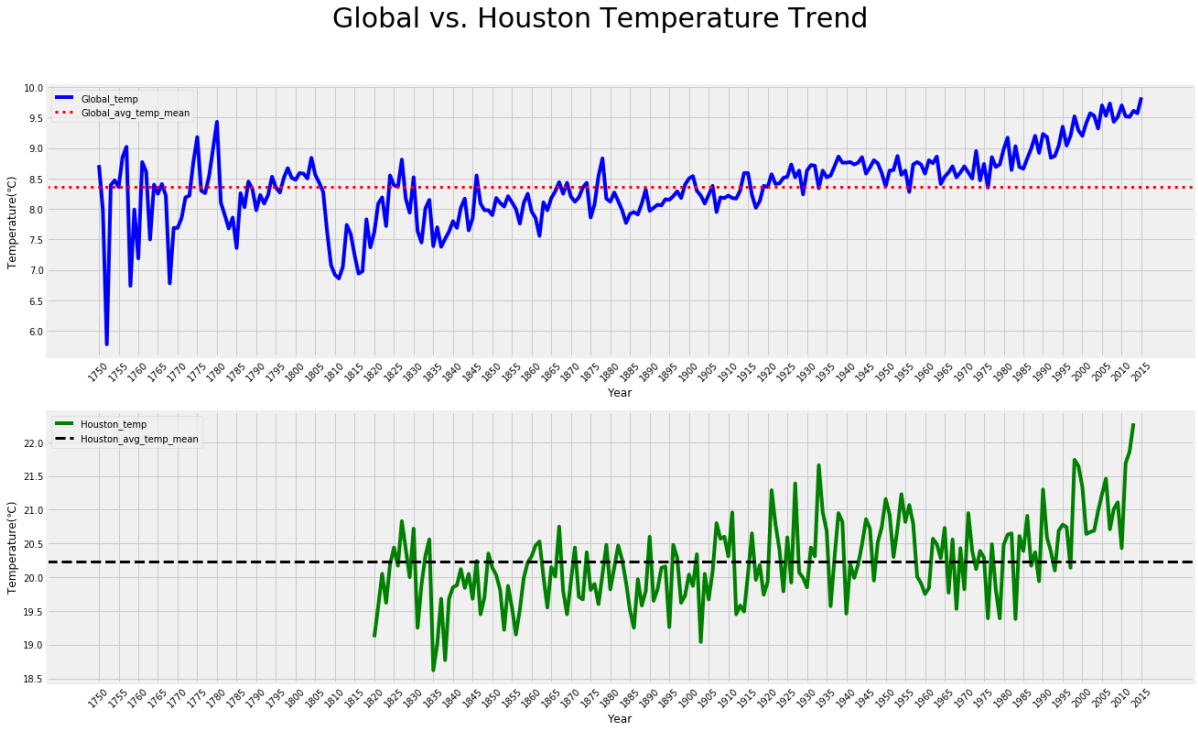
In [10]:
```
houston_data['avg_temp'].describe()
```

Out[10]:
```
count    194.000000
mean      20.231598
std        0.605483
min       18.620000
25%       19.820000
50%       20.185000
75%       20.607500
max       22.280000
Name: avg_temp, dtype: float64
```

From above preliminary dataset analysis, we can see that there are no missing data from both datasets. Global Temperature dataset has started from year 1750 to year 2015 while Houston Temerature data only started from 1820 to 2013. Global average temperature is much lower than Houston's. Considering Houston is in tropical area, this makes sense. It is also worth noting Houston Temperature dataset has a slightly higher standard deviation(0.605) than global dataset's (0.585). Whether this slight change in std is significant, we will find out by plotting the data.

Plots and data visualization

In [11]:
```
fig = plt.figure(figsize = (20, 12))
plt.suptitle('Global vs. Houston Temperature Trend', fontsize = 30)

ax1 = plt.subplot(211)
plt.plot(global_data.year, global_data.avg_temp,color='blue',label='Global_tem
p')
plt.axhline(global_data.avg_temp.mean(), color='r', linestyle='dotted', linewi
dth=3,label='Global_avg_temp_mean')
plt.legend(loc=0)
plt.xlabel('Year')
plt.ylabel('Temperature(℃)')
plt.xticks(np.arange(1750, 2020, step=5), rotation=45)


ax2 = plt.subplot(212, sharex=ax1)
plt.plot(houston_data.year, houston_data.avg_temp,color='green', label='Housto
n_temp')
plt.axhline(houston_data.avg_temp.mean(), color='k', linestyle='dashed', linew
idth=3,label='Houston_avg_temp_mean')
plt.legend(loc=0)
plt.xlabel('Year' )
plt.ylabel('Temperature(℃)')
plt.xticks(np.arange(1750, 2020, step=5), rotation=45)

plt.show()
```



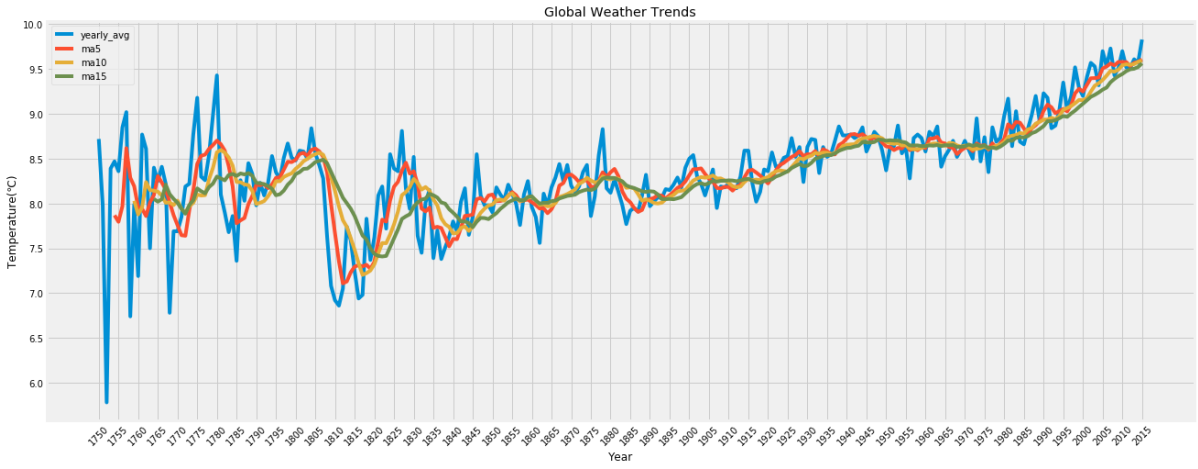
As we can see from the plots, Houston temperature change fluctuate significantly from year to year. Maybe Houston weather and climate are more dramtically affected by local ocean currents and severe weather changes.

Moving averages and comparison

Moving averages are used to smooth out data to make it easier to observe long term trends and not get lost in daily/short-term fluctuations. 5 years,10 years, and 15 years moving average are made for better understand and visualization of our global and local weather trends.

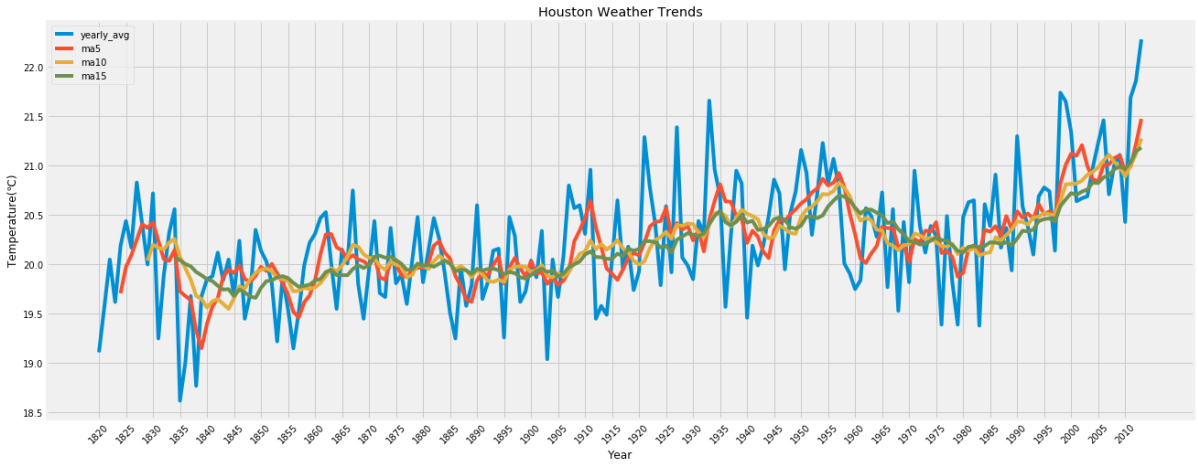
In [12]:
```
n = 5
global_data['ma5'] = global_data['avg_temp'].rolling(window=n).mean()
houston_data['ma5'] = houston_data['avg_temp'].rolling(window=n).mean()

n = 10
global_data['ma10'] = global_data['avg_temp'].rolling(window=n).mean()
houston_data['ma10'] = houston_data['avg_temp'].rolling(window=n).mean()

n = 15
global_data['ma15'] = global_data['avg_temp'].rolling(window=n).mean()
houston_data['ma15'] = houston_data['avg_temp'].rolling(window=n).mean()
```

In [13]:
```
plt.figure(figsize=(20,8))
plt.plot(global_data.year, global_data.avg_temp, label='yearly_avg')
plt.plot(global_data.year, global_data.ma5, label='ma5')
plt.plot(global_data.year, global_data.ma10, label='ma10')
plt.plot(global_data.year, global_data.ma15, label='ma15')
plt.title('Global Weather Trends')
plt.xlabel('Year')
plt.xticks(np.arange(1750, 2020, step=5), rotation=45)
plt.ylabel('Temperature(℃)')
plt.legend()
plt.show()
```



Those differen lines are overlapping each other. I can see the moving average of 15 year(green line) shows a smoother and more clear upward trending of global temperature within the last 40 years. Despite ups and downs from year to year, global average surface temperature is rising. The rate of temperature increase has nearly doubled in the last 40 to 50 years. Temperatures are certainly going up further. This may be due to the increasing greenhouse effect by human consumption of more fossil fuels and increased activities in cutting down carbon-absorbing forests in the last 40 years. You can also clearly see the global tempereature increased by 2 degree from mid-1800's to now. This may be due to massive use of fossil fuels with a steam engine to an industrial level.

In [14]:
```
plt.figure(figsize=(20,8))
plt.plot(houston_data.year, houston_data.avg_temp, label='yearly_avg')
plt.plot(houston_data.year, houston_data.ma5, label='ma5')
plt.plot(houston_data.year, houston_data.ma10, label='ma10')
plt.plot(houston_data.year, houston_data.ma15, label='ma15')
plt.legend()
plt.title('Houston Weather Trends')
plt.xlabel('Year')
plt.xticks(np.arange(1820, 2015, step=5), rotation=45)
plt.ylabel('Temperature(℃)')
plt.show()
```



Houston also follows the similar trend as global temperature changes. Ever since crude oil wells were found in west Texas as early 1800 century, Houston has become world oil center for crude oil transportation and downstream oil refinery. Despite ups and downs from year to year, Houston average surface temperature is rising. There are two major temperature rising peaks in the last 100 years. One peak is around 1950's and the other is yet to come. This may be in line with the fact that there was mass oil production during the world warII and the latest economic and technology booming is still largely based on the fussil fuel usage.

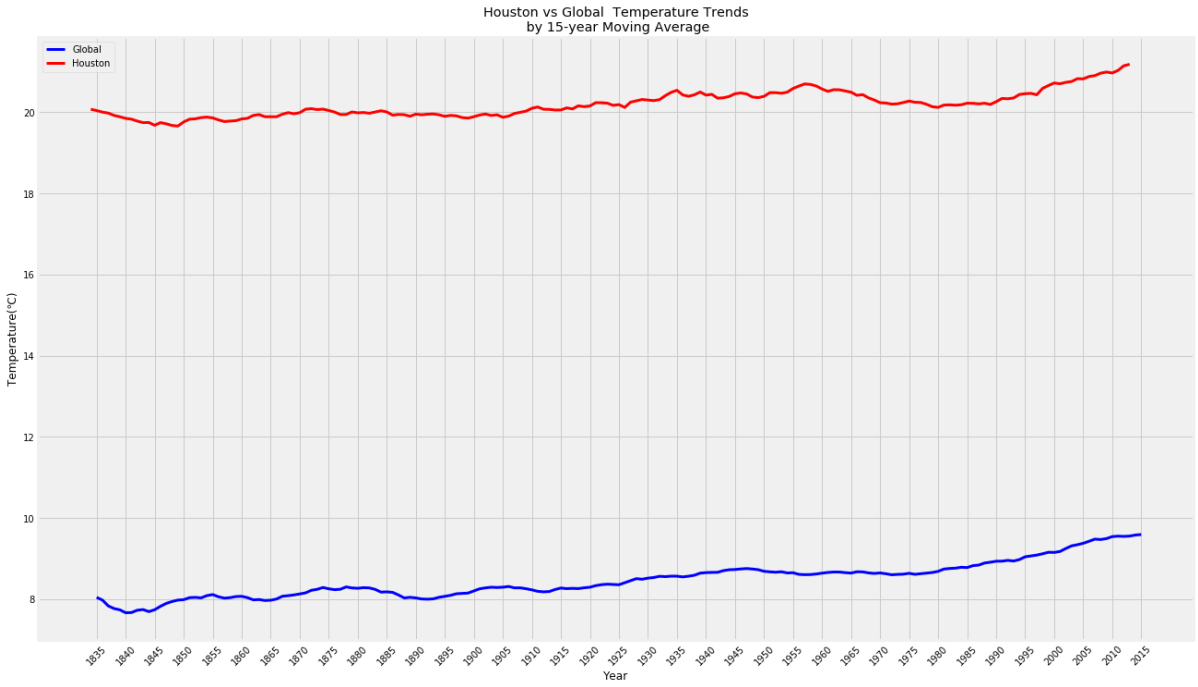
Because global Temperature dataset has started from year 1750 to year 2015 while Houston Temerature data only started from 1820 to 2013, I would like to compare 15-year moving averages of both datasets from 1835(1820+15) to the latest.

In [15]:
```
global_data2 = global_data[global_data.year >= 1835]
global_data2.head()
```

Out[15]:

	year	avg_temp	ma5	ma10	ma15
85	1835	7.39	7.728	8.044	8.092000
86	1836	7.70	7.740	7.978	8.066000
87	1837	7.38	7.726	7.835	8.012000
88	1838	7.51	7.626	7.769	7.998000
89	1839	7.63	7.522	7.738	7.936667

In [16]:
```
# Houston vs global moving average temperture
plt.figure(figsize=(20,12))
plt.plot(global_data2.year, global_data2.ma10, label='Global', color='blue', l
inewidth=3)
plt.plot(houston_data.year, houston_data.ma15, label='Houston', color='red', l
inewidth=3)
plt.legend()
plt.title('Houston vs Global  Temperature Trends\n by 15-year Moving Average')
plt.xlabel('Year')
plt.ylabel('Temperature(℃)')
plt.xticks(np.arange(1835, 2020, step=5), rotation=45)
plt.show()
```



Summary

After studying those two datasets, I would like to summarzie my study by answering the following questions raised by Udacity.

1.Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?
Houston is classified as humid subtropical climate, with tropical influences. That's why it is much hotter than global average. Rainfall is ample throughout the year Severe weather of Houston mostly takes the form of flooding, supercell thunderstorms, and tropic cyclones which occur most commonly in the months of spring to summer, which is the city's wet season. Despite ups and downs from year to year, global average surface temperature and our local(Houston) temperature are rising. The rate of temperature increase in both senarios has nearly doubled in the last 40 to 50 years. Temperatures are certainly going up further both locally and globally.

2.“How do the changes in your city’s temperatures over time compare to the changes in the global average?”
Despite global temperature slight rise as well as Houston temperature rise(both cases, there is about 2 degree Celsius increase over 200 years) there are two major temperature rising peaks in the last 100 years. One peak is around 1950's and the other is yet to come. This may be in line with the fact that there was mass oil production during the world warII and the latest economic and technology booming is still largely based on the fussil fuel usage.

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?
From our line plot, the overal trend is upward, which means the world is getting hotter. Despite ups and downs from year to year, global average surface temperature and our local(Houston) temperature are rising over the last hundred years.

Reference

Rolling statistics - Data Analysis with Python and Pandas Tutorial (https://www.youtube.com/watch?v=FRzfD1FtrsQ&t=125s)
Pandas moving average (https://pandas.pydata.org/pandas-docs/version/0.17.0/generated/pandas.rolling_mean.html)
Houston Climate (https://en.wikipedia.org/wiki/Climate_of_Houston)