Explore Weather Trends

Miya(Yi) Su

Abstract: The aim of this project is to analyze and visualize the local and global temperature data. I have considered "Los Angeles" as my local city and compared it with overall global temperature trends. This report focuses on providing five interesting observation about the change in temperature trends. The trends about the weather data (local and global) have been visualized using Python 3, where Excel / SQL query is used to extract the data.

Introduction: The change in weather trends has always been an interesting topic among scientist, politicians, environmentalists and others. The goal of this report is to compare and analyze the similarities and dissimilarities between the Los Angeles and global temperature data. The observations can be drawn by visualizing the temperature data.

Goals: 1. Data extraction (Excel / SQL)
2. Data manipulation (Python 3)
3. Data visualization (Python 3)

Data Extraction: I download data from Github and process in Excel. I use the filter function to get the Los Angeles's local temperature from local data. Then I use the Vlookup function to combine the global data and local data by the year to get the combined data. If the data is in the database, the SQL would be a better approach to extract the data.

SQL code:

Select global data:

```sql
SELECT *
FROM global_data
ORDER BY year;
```

Select local data:

```sql
SELECT *
FROM local_data
WHERE country = 'United States' and city = 'Los Angeles'
ORDER BY year;
```

Select combined data:

```sql
SELECT l.year, l.city, l.country, l.avg_temp as city_avg_temp, g.avg_temp as global_avg_temp
FROM local_data l
JOIN global_data g
ON l.year = g.year
WHERE l.country = 'United States' AND l.city = 'Los Angeles' AND l.avg_temp is not null
ORDER BY l.year;
```

Data manipulation: I have used Python 3 for data manipulation. I used Pandas package to import the data to Jupyter Notebook and created two new columns with “la_mv” and “global_mv” for calculating the moving average in 10 years window.

```
In [4]: avg_temp = temp_data[['city_avg_temp', 'global_avg_temp']]
        print(avg_temp.describe())
```

```
       city_avg_temp  global_avg_temp
count     165.000000       165.000000
mean       15.881152         8.554545
std         0.568754         0.460165
min        14.360000         7.560000
25%        15.500000         8.190000
50%        15.800000         8.530000
75%        16.240000         8.770000
max        18.120000         9.730000
```

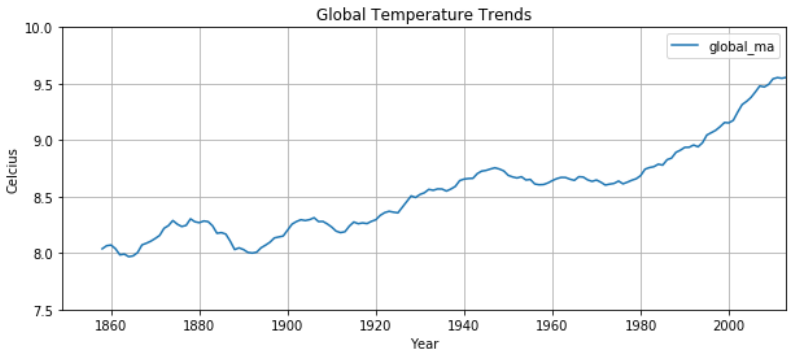
```
In [5]: #Data
        temp_data['la_ma']=temp_data['city_avg_temp'].rolling(window=10).mean()
        temp_data['global_ma']=temp_data['global_avg_temp'].rolling(window=10).mean()
        temp_data.tail()
```

Out[5]:

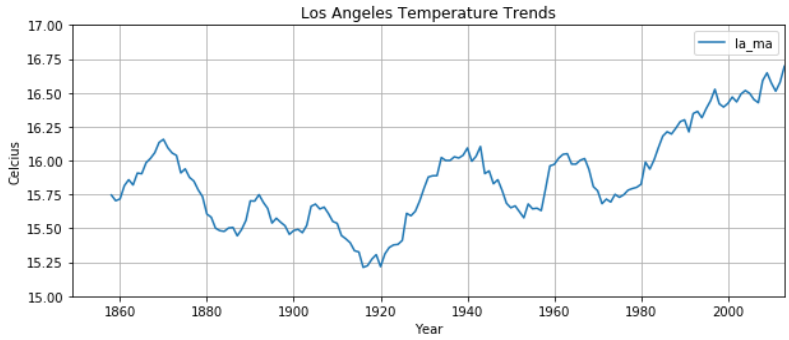
	year	city	country	city_avg_temp	global_avg_temp	la_ma	global_ma
160	2009	Los Angeles	United States	16.68	9.51	16.647	9.493
161	2010	Los Angeles	United States	15.89	9.70	16.572	9.543
162	2011	Los Angeles	United States	15.87	9.52	16.512	9.554
163	2012	Los Angeles	United States	17.09	9.51	16.578	9.548
164	2013	Los Angeles	United States	18.12	9.61	16.696	9.556

Data Visualization: I have analyzed data by plotting the 3 charts. i. Local moving average of temp vs. years. ii. Global moving average of temp vs. years. iii. Comparison between local and global moving average temp.

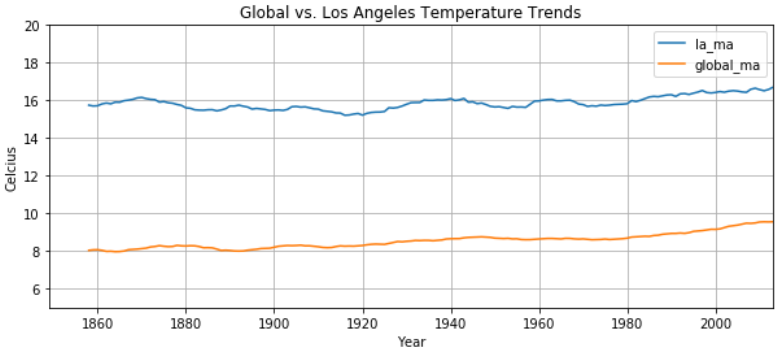
```
#Global moving average of temp vs. year
fig,ax = plt.subplots(nrows=1, ncols=1, figsize=(10,4))
temp_data.plot(kind='line', x='year', y='global_ma', ax=ax)
ax.set(title='Global Temperature Trends', xlabel='Year', ylabel='Celcius', ylim=(7.5,10))
ax.grid(True)
```



```
#Local moving average of temp vs. year
fig,ax = plt.subplots(nrows=1, ncols=1, figsize=(10,4))
temp_data.plot(kind='line', x='year', y='la_ma', ax=ax)
ax.set(title='Los Angeles Temperature Trends', xlabel='Year', ylabel='Celcius', ylim=(15,17))
ax.grid(True)
```




```
#Los Angeles vs. Global moving average of temp
fig,ax = plt.subplots(nrows=1, ncols=1, figsize=(10,4))
temp_data.plot(kind='line', x='year', y=['la_ma','global_ma'], ax=ax)
ax.set(title='Global vs. Los Angeles Temperature Trends', xlabel='Year', ylabel='Celcius', ylim=(5,20))
ax.grid(True)
```



Data Interpretation: There following observations were drawn after analyzing the data from line charts:

1) The Los Angeles's temperature is higher than global average temperature from 1860 to 2010 and the difference is decreasing.
2) The overall trend shows increase in temperature in both global and Los Angeles level. This trends is consistent over last hundred years.
3) The lowest global temperatures are recorded around 1862 and 1885. And the lowest Los Angeles's temperature is recorded on 1920.
4) Los Angeles's temperature is higher than more than half of areas in the world.
5) The upward trend after analyzing the last century data shows that it could be because of the industrialization or heavy use of fossil fuel.