Explore Weather Trends

The notebook is a python implementation of analyzing the weather trend dataset. Instead of running multiple SQL queries, I executed the following query, which returned just one table with the local temperature of the nearest capital city, Riyadh in Saudi Arabia, and the global temperature.

Tools:

- SQL JOIN: I used JOIN in SQL to get results in one table. Another approach is to export two tables, city and global, from the workspace, and merge them using pandas.DataFrame.merge()
- Python as a programming language
- pandas library for data manipulation
- matplotlib library for plotting
- Jupyter notebook

SQL query

```
select city_data.year, city_data.avg_temp as riyadh, global_data.avg_temp as global from city_data join global_data on city_data.year = global_data.year where city_data.city = 'Riyadh' and city_data.country = 'Saudi Arabia';
```

In [1]:
```
# Import required libraries
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
# Read the csv file. There will be NaN values. I kept them as is and did not fill them out with anything.
temps = pd.read_csv('results.csv')
```

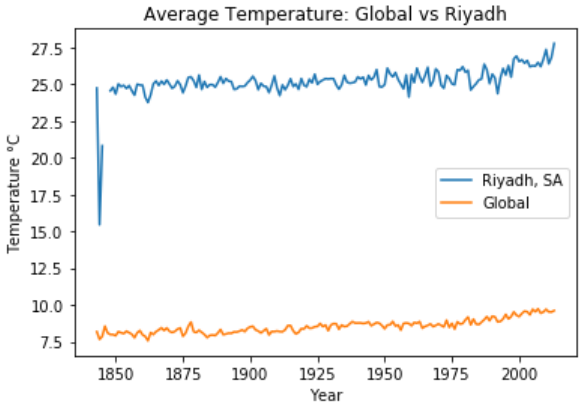
In [3]:
```
# Calculated the moving average mean on 7 and 14 days. Results are appended to the dataframe.
# 7ma = 7-day moving average temperature
# 14ma = 14-day moving average temperature
temps[['riyadh_7ma', 'global_7ma']] = temps[['riyadh', 'global']].rolling(7).mean()
temps[['riyadh_14ma', 'global_14ma']] = temps[['riyadh', 'global']].rolling(14).mean()
```

In [4]:
```
# A sample to check how things look like.
# Data started from 1843 to 2013
temps.sample(5)
```

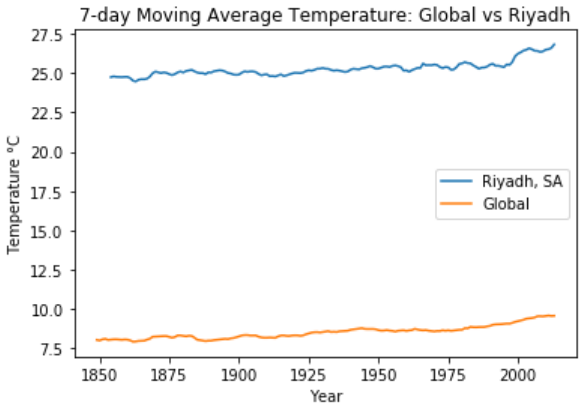
Out[4]:

	year	riyadh	global	riyadh_7ma	global_7ma	riyadh_14ma	global_14ma
131	1974	25.40	8.47	25.395714	8.620000	25.449286	8.642143
130	1973	25.60	8.95	25.320000	8.652857	25.468571	8.650000
42	1885	24.98	7.92	25.042857	8.051429	25.049286	8.188571
52	1895	24.69	8.15	25.105714	8.107143	25.020714	8.035714
85	1928	25.39	8.63	25.287143	8.535714	25.150000	8.430714

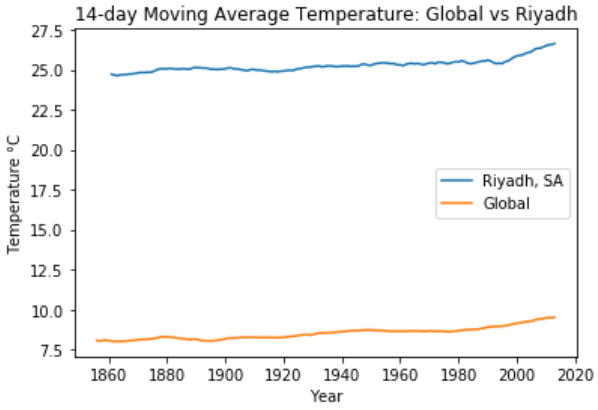
In [5]:
```
# Plotting the temperature average between Riyadh and the global one. I know this is not the requi
red chart.
# However, it's here just to see climate oscillations. The missing part in the beginning of Riya
d's line chart
# is because some of the data are missing. I did not do any data cleanup
plt.plot(temps['year'], temps['riyadh'], label='Riyadh, SA')
plt.plot(temps['year'], temps['global'], label='Global')
plt.title('Average Temperature: Global vs Riyadh')
plt.xlabel('Year')
plt.ylabel('Temperature °C')
plt.legend(loc='center right')
plt.show()
```



In [6]:
```
# 7-day moving average temperature comparison between the global temperature and Riyadh's
plt.plot(temps['year'], temps['riyadh_7ma'], label='Riyadh, SA')
plt.plot(temps['year'], temps['global_7ma'], label='Global')
plt.title('7-day Moving Average Temperature: Global vs Riyadh') # 7-day moving Average
plt.xlabel('Year')
plt.ylabel('Temperature °C')
plt.legend(loc='center right')
plt.show()
```



In [7]:
```
# 14-day moving average temperature comparison between the global temperature and Riyadh's
plt.plot(temps['year'], temps['riyadh_14ma'], label='Riyadh, SA')
plt.plot(temps['year'], temps['global_14ma'], label='Global')
plt.title('14-day Moving Average Temperature: Global vs Riyadh') # 14-day moving Average
plt.xlabel('Year')
plt.ylabel('Temperature °C')
plt.legend(loc='center right')
plt.show()
```



Observations:

1. Global Warming is a fact. Temperature, both locally and globally, has been increasing
2. Local temperature, of Riyadh, and the global one are highly correlated
3. Saudi Arabia's weather is much, much, much hotter than the global average
4. Weather increased by %5.6 in the last 2 decades, compared to an increase of %3 in the first 2 decades in the dataset (using the 7-day moving average of the global temperature) -numbers in below cells-

In [8]:
```
# Temperature change in the first 2 decades in the dataset
(1 - temps['global_7ma'][7]/temps['global_7ma'][27]) * 100
```

Out[8]: 3.0638739830362027

In [9]:
```
# Temperature change in the last 2 decades in the dataset
(1 - temps['global_7ma'][150]/temps['global_7ma'][170]) * 100
```

Out[9]: 5.6409491120728354