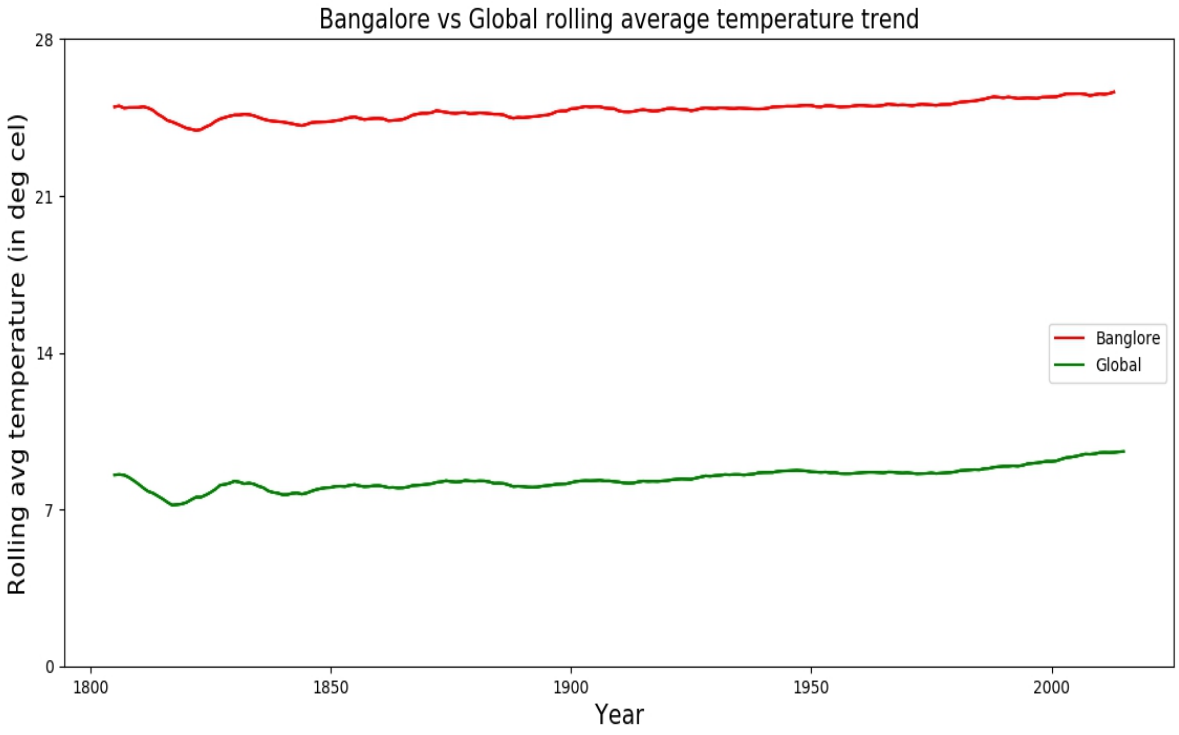
WEATHER TRENDS

Objective:
To analyse and interpret between my closest city's temperature (Bangalore) historian with the global temperature profile along with few major cities around the world.

Tools used: Python and SQL
Packages: Matplotlib and pandas.

Data Acquisition and pre-processing
1. The data has been downloaded individually from SQL workspace using the relevant queries.
2. Basically the data has been checked with the NaN and missing values, and the same has been filled by the average of the respective columns (numerical columns).
3. Then the Rolling average has been identified for window=10, to smoothen out the curve for visualization and analysis.
4. To visualize the last 43 years of temperature across the cities, relevant filtering has been done.

Key takeaways:
1. Similarity between global temperature and Bangalore is the slope of the trend, which defines the rate at which temperature has changed over the years, Bangalore saw an increase of 0.005 deg cel per year whereas the global temperature also saw an increase of close to 0.0048 deg cel every year.

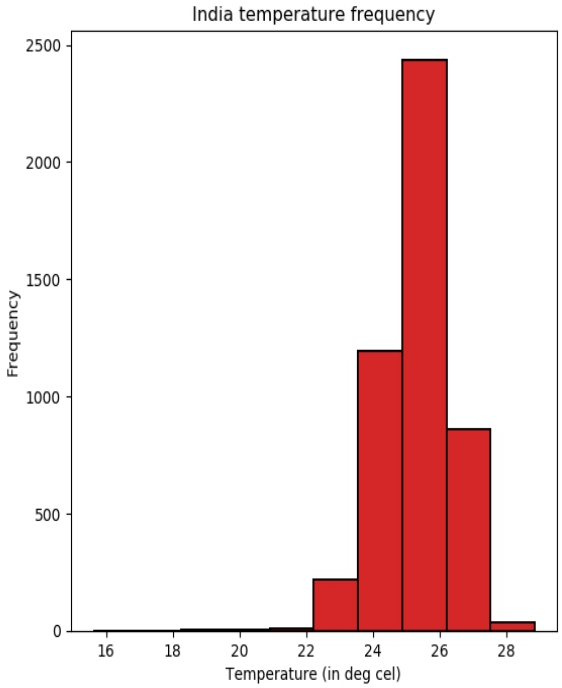


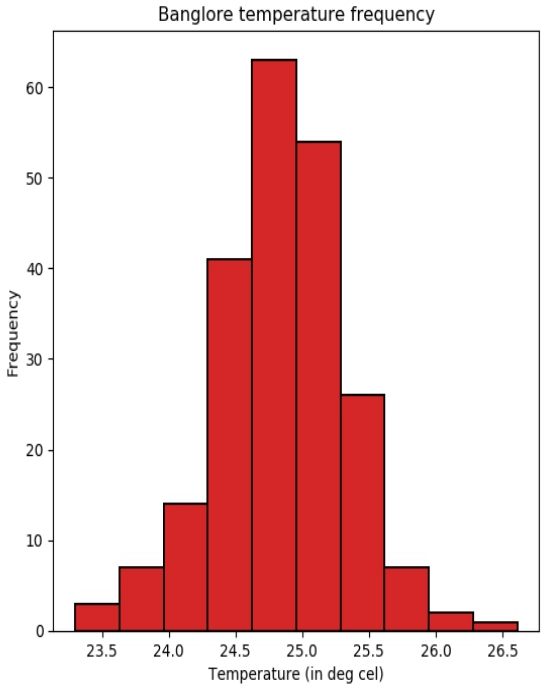
Though the sample size for global temperature has been slightly more, this insight of referring to the rate of change will remain more or less the same.

WEATHER TRENDS

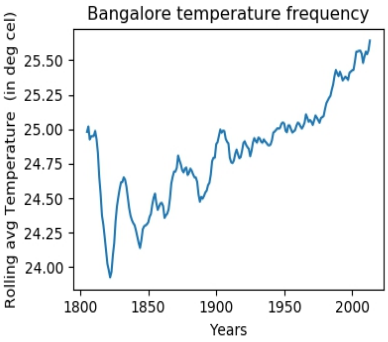
2. The rate of change for overall india and Bangalore has been the same of ~0.005 deg cel per year, which gives an idea of the effect of global warming across the country.

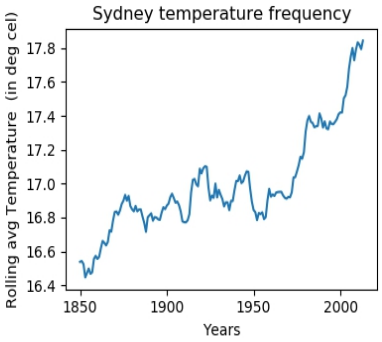
3. The distribution for India is slightly left skewed whereas for Bangalore it is a perfect normal distribution, which means the arithmetic mean will be slightly more than that of
Bangalore.

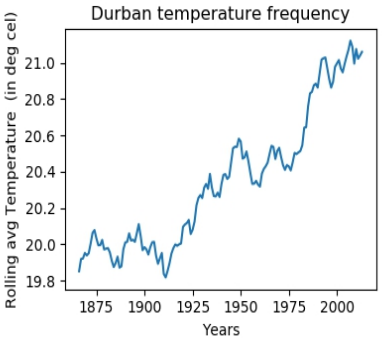


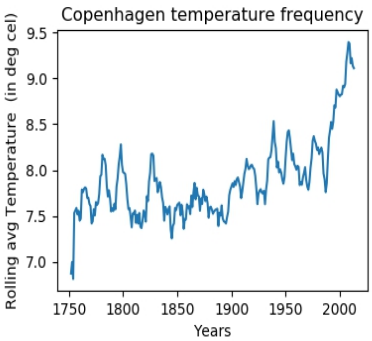


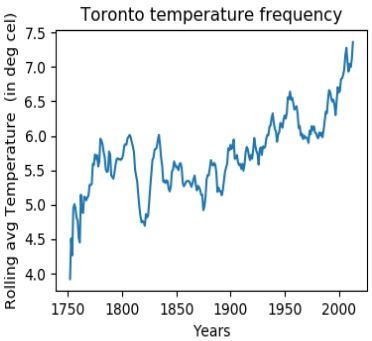
4. The temperature of major cities from each continent are also similar to the above interpretation ie., there is an upward trend in temperature with Durban seeing a max increase of 0.008 deg cel per year.

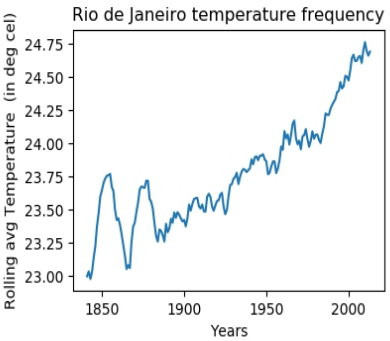












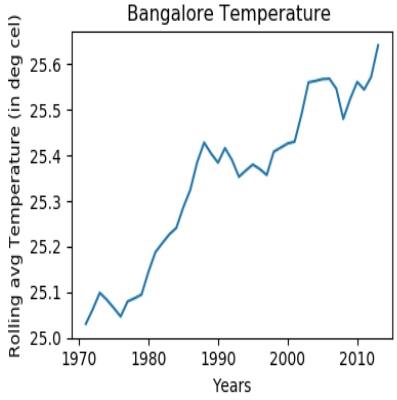
Summarizing the rate of change across the cities

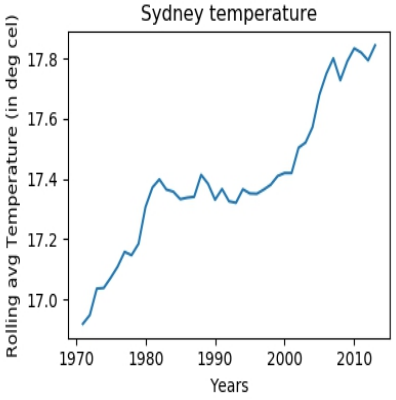
WEATHER TRENDS

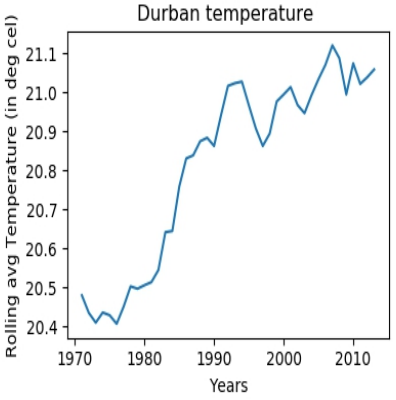
City	Rate of change per year(in deg cel)
Bangalore	0.005
Sydney	0.005
Durban	0.008
Copenhagen	0.004
Toronto	0.006
Rio De Janeiro	0.007

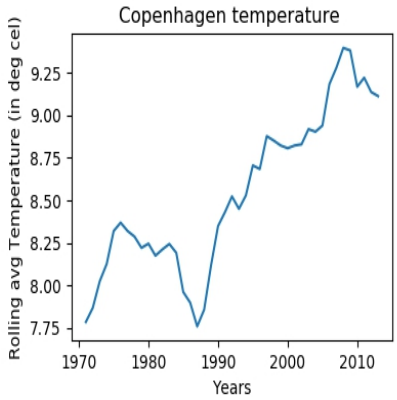
5. It is interesting to note that the last few year (40 years) have witnessed a sharp rise in the temperature in these cities, magnifying a little bit on this would give a fair idea of how global warning has aggravated in these years.

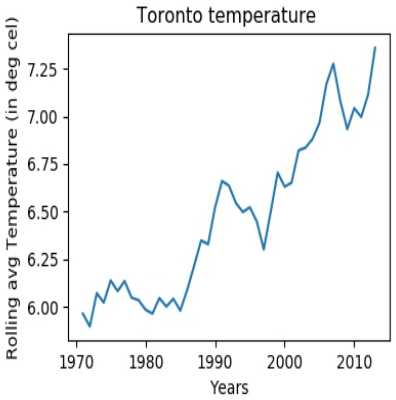
City	Rate of change per year(in deg cel)
Bangalore	0.016
Sydney	0.017
Durban	0.016
Copenhagen	0.027
Toronto	0.040
Rio De Janeiro	0.020
Global	0.026

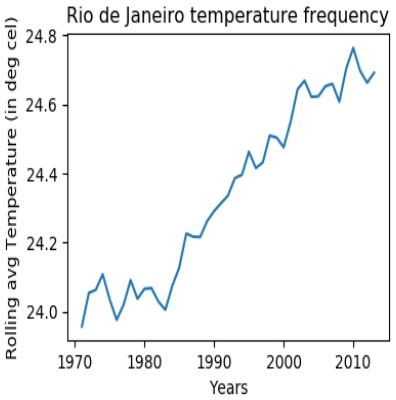












WEATHER TRENDS

To statistically infer the temperature historian of all the above major cities aong with the global temperature, the below table furnishes the details:

Cities/location								
		Global	**Bangalore**	**Sydney**	**Durban**	**Copenhagen**	**Toronto**	**Rio de Janeiro**
Parameters	**Mean (deg cel)**	8.36	24.85	17.00	20.35	7.88	5.72	23.79
	Mode(deg cel)	7.98	24.85	16.86	19.73	7.88	5.7	23.79
	Std Dev (deg cel)	0.58	0.47	0.43	0.48	0.98	1.21	0.58
	Min & Max(deg cel)	(5.78, 9.83)	(23.3,26.6)	(16.08, 18.18)	(19.31, 21.64)	(0.09,9.87)	(-3.96, 8.7)	(22.29, 25.19)

- Since the standard deviation tells us the spread of the data, Canada has shown some largest variance over the years with a value of 1.21.
- Max temperature has understandably on the higher side being in a tropical region for Bangalore with other cities as well as the global maximum temperature falling below 25 deg cel.
- It is also obvious to see that the mode which reflects the frequently occurred temperature, is almost equal to the mean of the respective location.

WEATHER TRENDS

Appendix

Python Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#%%
df_b=pd.read_csv('A:/Work/udacity/data_analysis/course-1/bangalore.csv')
df_g=pd.read_csv('A:/Work/udacity/data_analysis/course-1/global.csv')
df_i=pd.read_csv('A:/Work/udacity/data_analysis/course-1/india.csv')
df_c=pd.read_csv('A:/Work/udacity/data_analysis/course-1/copenhagen.csv')
df_d=pd.read_csv('A:/Work/udacity/data_analysis/course-1/durban.csv')
df_t=pd.read_csv('A:/Work/udacity/data_analysis/course-1/toronto.csv')
df_r=pd.read_csv('A:/Work/udacity/data_analysis/course-1/rio.csv')
df_s=pd.read_csv('A:/Work/udacity/data_analysis/course-1/sydney.csv')

df_b.head(5)
df_b.isna().sum()
df_b['avg_temp'].fillna(df_b['avg_temp'].mean(),inplace=True)
df_b['avg_temp']=df_b['avg_temp'].astype('float64')
df_b['rolling_mean']=df_b.iloc[:,3].rolling(window=10).mean()

df_g.isna().sum()
df_g['rolling_mean']=df_g.iloc[:,1].rolling(window=10).mean()

df_i.isna().sum()
df_i['avg_temp'].fillna(df_i['avg_temp'].mean(),inplace=True)
df_i['avg_temp']=df_i['avg_temp'].astype('float64')
df_i['rolling_mean']=df_i.iloc[:,3].rolling(window=10).mean()

df_c['avg_temp'].fillna(df_c['avg_temp'].mean(),inplace=True)
df_c['avg_temp']=df_c['avg_temp'].astype('float64')
df_c['rolling_mean']=df_c.iloc[:,3].rolling(window=10).mean()

df_d['avg_temp'].fillna(df_d['avg_temp'].mean(),inplace=True)
df_d['avg_temp']=df_d['avg_temp'].astype('float64')
df_d['rolling_mean']=df_d.iloc[:,3].rolling(window=10).mean()

df_t['avg_temp'].fillna(df_t['avg_temp'].mean(),inplace=True)
df_t['avg_temp']=df_t['avg_temp'].astype('float64')
df_t['rolling_mean']=df_t.iloc[:,3].rolling(window=10).mean()
df_r['avg_temp'].fillna(df_r['avg_temp'].mean(),inplace=True)
df_r['avg_temp']=df_r['avg_temp'].astype('float64')
df_r['rolling_mean']=df_r.iloc[:,3].rolling(window=10).mean()

df_s['avg_temp'].fillna(df_s['avg_temp'].mean(),inplace=True)
df_s['avg_temp']=df_s['avg_temp'].astype('float64')
df_s['rolling_mean']=df_s.iloc[:,3].rolling(window=10).mean()
```

```
plt.figure(1)
plt.subplot(121)
plt.plot(df_b['year'][9:],df_b['rolling_mean'][9:],'red')
plt.xlabel('year')
plt.ylabel('Rolling avg temperature')
plt.title('Bangalore rolling average temperature trend')

plt.subplot(122)
plt.plot(df_g['year'][9:],df_g['rolling_mean'][9:],'red')
plt.xlabel('year')
plt.ylabel('Rolling avg temperature')
plt.title('Global rolling avergae temperature trend')

plt.figure(2)
plt.subplot(121)
plt.plot(df_b['year'],df_b['avg_temp'])
plt.xlabel('year')
plt.ylabel('Average temperature')
plt.title('Bangalore temperature trend')

plt.subplot(122)
plt.plot(df_g['year'],df_g['avg_temp'])
plt.xlabel('year')
plt.ylabel('Average temperature')
plt.title('Global temperature trend')

plt.figure(3)
plt.scatter(df_i['year'][9:],df_i['rolling_mean'][9:])
plt.xlabel('year')
plt.ylabel('Rolling avg temperature')
plt.title('India temperature trend')

plt.figure(4)
plt.subplot(121)
plt.hist(df_i['avg_temp'],linewidth=1.2,edgecolor = 'black')
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('India temperature frequency')

plt.subplot(122)
plt.hist(df_b['avg_temp'],linewidth=1.2,edgecolor = 'black')
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('Banglore temperature frequency')

plt.figure(5)
plt.subplot(121)
plt.hist(df_g['avg_temp'])
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('Global temperature frequency')
```

WEATHER TRENDS

```
plt.figure(6)
plt.subplot(231)
plt.plot(df_b['year'][9:],df_b['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Bangalore temperature frequency')
plt.subplot(232)
plt.plot(df_s['year'][9:],df_s['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Sydney temperature frequency')
plt.subplot(233)
plt.plot(df_d['year'][9:],df_d['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Durban temperature frequency')
plt.subplot(234)
plt.plot(df_c['year'][9:],df_c['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Copenhagen temperature frequency')
plt.subplot(235)
plt.plot(df_t['year'][9:],df_t['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Toronto temperature frequency')
plt.subplot(236)
plt.plot(df_r['year'][9:],df_r['rolling_mean'][9:])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Rio de Janeiro temperature frequency')
plt.subplots_adjust(top=0.92, bottom=0.08, left=0.10,
right=0.95, hspace=0.35,
                    wspace=0.35)
plt.show()
from scipy import stats
slope1=stats.linregress(df_b['year'],df_b['avg_temp'])
slope2=stats.linregress(df_g['year'],df_g['avg_temp'])
slope3=stats.linregress(df_i['year'],df_i['avg_temp'])
slope4=stats.linregress(df_s['year'],df_s['avg_temp'])
slope5=stats.linregress(df_d['year'],df_d['avg_temp'])
slope6=stats.linregress(df_c['year'],df_c['avg_temp'])
slope7=stats.linregress(df_t['year'],df_t['avg_temp'])
slope8=stats.linregress(df_r['year'],df_r['avg_temp'])
```

```
print('The rate of change through the years for bangalore
is',slope1.slope,
      'The rate of change through the years for Global
temperature is',slope2.slope,
      'The rate of change through the years for India
temperature is',slope3.slope)
print('The rate of change through the years for Sydney
temperature is',slope4.slope)
print('The rate of change through the years for Durban
temperature is',slope5.slope)
print('The rate of change through the years for Copenhagen
temperature is',slope6.slope)
print('The rate of change through the years for Toronto
temperature is',slope7.slope)
print('The rate of change through the years for Rio De
Janeiro temperature is',slope8.slope)
print('The max temp in bangalore through the
years',max(df_b.avg_temp))
print('The max temp in bangalore through the
years',min(df_b.avg_temp))
print('The max temp in global through the
years',max(df_g.avg_temp))
print('The max temp in Global through the
years',min(df_b.avg_temp))
print('The avg temperature for India through the
years',df_i['avg_temp'].mean())
print('The avg temperature for bangalore through the
years',df_b['avg_temp'].mean())
#%%

bang_50=df_b[df_b['year']>1970]
syd_50=df_s[df_s['year']>1970]
dur_50=df_d[df_d['year']>1970]
cop_50=df_c[df_c['year']>1970]
tor_50=df_t[df_t['year']>1970]
rio_50=df_r[df_r['year']>1970]

slope1a=stats.linregress(bang_50['year'],bang_50['avg_tem
p'])
slope2a=stats.linregress(syd_50['year'],syd_50['avg_temp'])
slope3a=stats.linregress(dur_50['year'],dur_50['avg_temp'])
slope4a=stats.linregress(cop_50['year'],cop_50['avg_temp'])
slope5a=stats.linregress(tor_50['year'],tor_50['avg_temp'])
slope7a=stats.linregress(rio_50['year'],rio_50['avg_temp'])

print('The rate of change through the years for Bangalore
temperature is',slope1a.slope)
print('The rate of change through the years for Sydney
temperature is',slope2a.slope)
print('The rate of change through the years for Durban
temperature is',slope3a.slope)
print('The rate of change through the years for Copenhagen
temperature is',slope4a.slope)
print('The rate of change through the years for Toronto
temperature is',slope5a.slope)
print('The rate of change through the years for Rio De
Janeiro temperature is',slope7a.slope)
#%%
```

WEATHER TRENDS

```
plt.figure(7)
plt.subplot(231)
plt.plot(bang_50['year'],bang_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Bangalore Temperature ')

plt.subplot(232)
plt.plot(syd_50['year'],syd_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Sydney temperature')


plt.subplot(233)
plt.plot(dur_50['year'],dur_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Durban temperature ')

plt.subplot(234)
plt.plot(cop_50['year'],cop_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Copenhagen temperature ')

plt.subplot(235)
plt.plot(tor_50['year'],tor_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Toronto temperature ')

plt.subplot(236)
plt.plot(rio_50['year'],rio_50['rolling_mean'])
plt.xlabel('Years')
plt.ylabel('Rolling avg Temperature')
plt.title('Rio de Janeiro temperature frequency')

plt.subplots_adjust(top=0.92, bottom=0.08, left=0.10, right=0.95, hspace=0.35,
                    wspace=0.35)

plt.show()
```

WEATHER TRENDS

SQL queries

```
Select * from city_list where country='India';
SELECT * from city_data where city=’Bangalore’;
SELECT * from global_data;
SELECT * from city_data where city=’Sydney’
SELECT * from city_data where city=Durban;
SELECT * from city_data where city=Toronto;
SELECT * from city_data where city=’Copenhagen’;
SELECT * from city_data where city=’Rio de Janeiro’;
```