Exploring Weather Trends

Overview:

In this project, I will analyze Toronto and global temperature data and compare the temperature trend of Toronto to the global temperature trend.

Goal:

I will visualize and describe the similarities and differences between global temperature and temperature of Toronto in which I live.

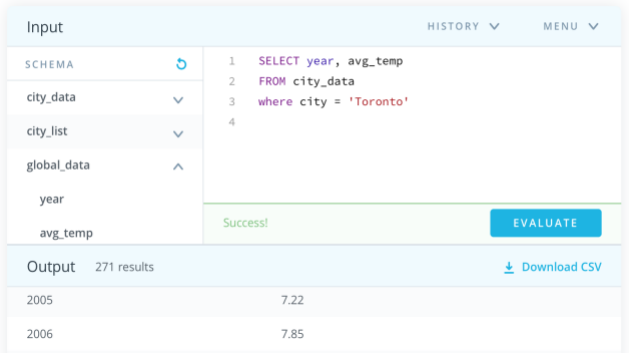
Tools:

To do this, I will follow the steps below:

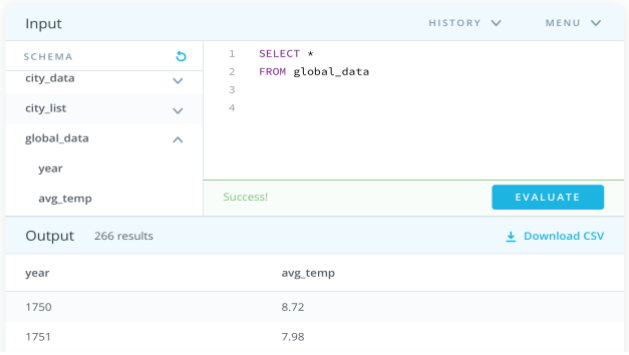
- SQL:
  - To extract data from the weather temperature database
  - To export and save the output to CSV files
- Python:
  - To read CSV files into Pandas' data frame
  - To merge data frame
  - To perform calculation: moving average, slope, avg. temperature, etc.
  - To visualize the trend using matplotlib and seaborn libraries
- Jupyter Notebook:
  - To write Python code, perform data analysis and visualization
- Microsoft word:
  - To compose formal report
  - Export to PDF

1. Data Extraction:

1.1 Use SQL to extract weather data for both Toronto and global temperature
1.2 Download csv and rename files to 'toronto.csv' and 'global.csv'



Graph 1: Extract Toronto Weather



Graph 2: Extract Global Weather

2. Data Preparation:

2.1 Import library to Jupyter Notebook

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
pd.options.display.max_rows = 10
import seaborn as sns
from scipy.stats import linregress
```

2.2 Read CSV files to Pandas data frames

```
toronto_df = pd.read_csv('toronto.csv')
global_df = pd.read_csv('global.csv')`
```

2.3 Rename column names

```
toronto_df=toronto_df.rename(columns = {'avg_temp':'avg_toronto'})
global_df=global_df.rename(columns = {'avg_temp':'avg_global'})
```

2.4 Merge data frames

```
df = pd.merge(toronto_df, global_df, how='inner',on='year')
```

2.5 Display the final data frame which is ready for analysis

```
df.head(10)
```

	year	avg_toronto	avg_global
0	1750	6.29	8.72
1	1751	6.84	7.98
2	1752	-1.10	5.78
3	1753	5.76	8.39
4	1754	5.94	8.47
5	1755	2.81	8.36
6	1756	6.37	8.85
7	1757	5.13	9.02
8	1758	4.37	6.74
9	1759	5.27	7.99

3. Data Processing:

3.1 Moving Average

- Python's built-in 'rolling' function is used to calculate the moving average of temperature
- The moving average's rolling window is set as 10 years
- The purpose of using moving average instead of single year data is to smooth out the time series to make it easier to be observed analyzed

In [7]:
```
#parse all column names into a name list
name=list(df)
```

In [8]:
```
#for loop to create Moving Average
#rolling window is set as 10 years
for i in range(len(name)):
    df['MA_{}'.format(name[i])] = df[name[i]].rolling(window=10).mean()
```

In [9]:
```
df
```

Out[9]:

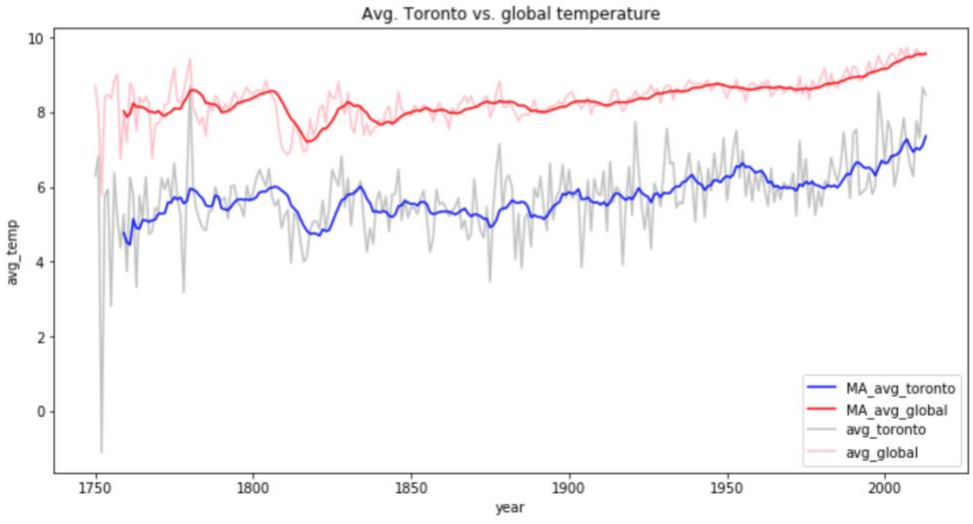
year	avg_toronto	avg_global	MA_avg_toronto	MA_avg_global
1750	6.29	8.72	NaN	NaN
1751	6.84	7.98	NaN	NaN
1752	-1.10	5.78	NaN	NaN
1753	5.76	8.39	NaN	NaN
1754	5.94	8.47	NaN	NaN
...	...	...	...	...
2009	6.28	9.51	6.933	9.493
2010	7.77	9.70	7.043	9.543
2011	7.30	9.52	6.997	9.554
2012	8.66	9.51	7.115	9.548
2013	8.46	9.61	7.359	9.556

264 rows × 4 columns

4. Data Visualization:

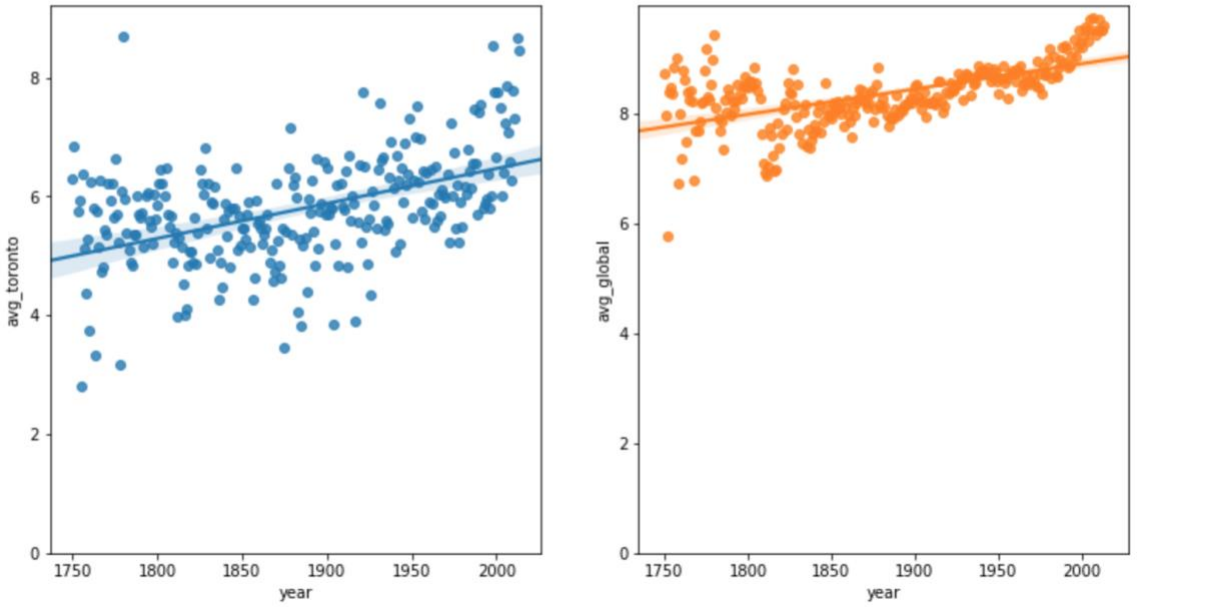
4.1 Line chart for Toronto and global temperature (based on moving average of 10 years)

In [10]:
```
ax=df[['MA_avg_toronto','MA_avg_global']].plot(figsize=(12,6),color =('blue','red'),zorder=1)
ax.set_xlabel("x label")
ax2=df[['avg_toronto','avg_global']].plot(ax=ax,color =('silver','pink'),zorder=0)
plt.ylabel('avg_temp')
plt.title('Avg. Toronto vs. global temperature')
plt.show()
```



4.2 Regression plots for Toronto and global temperature

In [14]:
```
#Create two regplots side by side
fig, (ax1, ax2) = plt.subplots(ncols=2, sharey=False)
#Set the size of each subplots
fig.set_size_inches(12.5, 6.5)
f=sns.regplot(x=aa['year'], y=aa['avg_toronto'], ax=ax1)
f.set(ylim=(0, None))
g=sns.regplot(x=aa['year'], y=aa['avg_global'], ax=ax2)
g.set(ylim=(0, None))
plt.show()
```



5. Statistical Calculation:

5.1 Calculate the slope of regression lines for Toronto and global temperature

5.1.1 Slope of Toronto temperature

linregress(df.index,df.avg_toronto)

```
slope=0.0058759269283697145
```

5.1.2 Slope of global temperature

linregress(df.index, df.avg_global)

```
slope=0.00461111074878203
```

5.2 Calculate the mean temperature for Toronto and globe

print("Toronto's avg. temp: " + "{0:.2f}".format(df['avg_toronto'].mean()))

print("Global avg. temp: " + "{0:.2f}".format(df['avg_global'].mean()))

```
Toronto's avg. temp: 5.77
```

```
Global avg. temp: 8.36
```

6. Observation:

6.1 Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?

Based on the line chart in section 4.1, Toronto is colder on average compared to the global average temperature. The difference is consistent about 2.6 degree Celsius. Toronto is colder than global temperature because Toronto is located in a high latitude area.

6.2 How do the changes in your city's temperatures over time compare to the changes in the global average?"

Toronto temperature is raising along with the global temperature, but it is more volatile (having a higher variance) than global temperature.

6.3 What does the overall trend look like?

Overall, Toronto and global temperature is increasing due to global warming.

6.4 Is the world getting hotter or cooler?

The world is getting hotter, we can see that in the line plot and also from the slope of regression plots of Toronto and global temperature. The slope for Toronto is 0.00587, and the slope for the world is 0.00461.

6.5 Has the trend been consistent over the last few hundred years?

The uprising trend of global and Toronto's temperature is consistent over the years, this shows a strong evidence of global warming. Toronto's slope of trend is 0.0012 higher than the global trend. Therefore, Toronto's temperature is rising a bit faster than the global temperature.