Extract the data

- Find the nearest large city

```sql
SELECT *
FROM city_list
WHERE country='United States'
```

Then I found that the large city nearest to Gainesville in FL is Jacksonville.

  - Write a SQL query to extract the city level data. Export to CSV.

```sql
SELECT year, avg_temp
FROM city_data
WHERE city='Jacksonville'
```

Then I exported it to a csv file.

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

```sql
SELECT *
FROM global_data
```

Line charts of moving averages

- Steps
  - I used python packages pandas and matplotlib to calculate the moving averages and to plot the lines.
  - The key consideration when deciding how to visualize the trends is how to choose the window for averaging. I'd like the curves to be fairly smooth without losing too many details at the same time. Therefore, moving averages over 10 years might be a good choice.
  - The following is the python code I used to calculate the moving averages and plot.

```python
import pandas as pd
import matplotlib.pyplot as plt

def ma(file):
    df = pd.read_csv(file)
    df['10_ma'] = df['avg_temp'].rolling(10).mean()
```
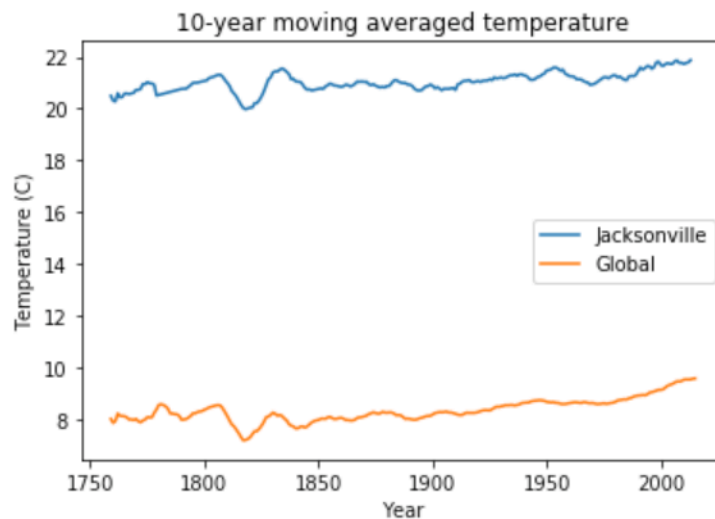
```
        df.dropna(inplace=True)
        return df

    df_city = ma('city_data.csv')
    df_global = ma('global_data.csv')
    plt.plot(df_city['year'],df_city['10_ma'],label='Jacksonville')
    plt.plot(df_global['year'],df_global['10_ma'],label='Global')
    plt.legend()
    plt.xlabel('Year')
    plt.ylabel('10-year moving averaged temperature (C)')
```

- The code above generates the figure below.

Make observations

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

  Jacksonville is consistently hotter than the global average.

- *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?*

  Overally the worlder is getting hotter. There were a few drops, e.g., in around 1820, but the overal trend is that it's getting hotter.

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

  The temperatures of both Jacksonville and global average are increasing, and we can see which one is increasing more rapidly by using a linear regression on both of the two datasets. The following code adds trendlines.

```
import numpy as np
def trend(df):
    x = df['year']
    y = df['ma']
    z = np.polyfit(x,y,1)
    p = np.poly1d(z)
    print(p)
    plt.plot(x , p(x))
trend(df_city)
trend(df_global)
```


10-year moving averaged temperature
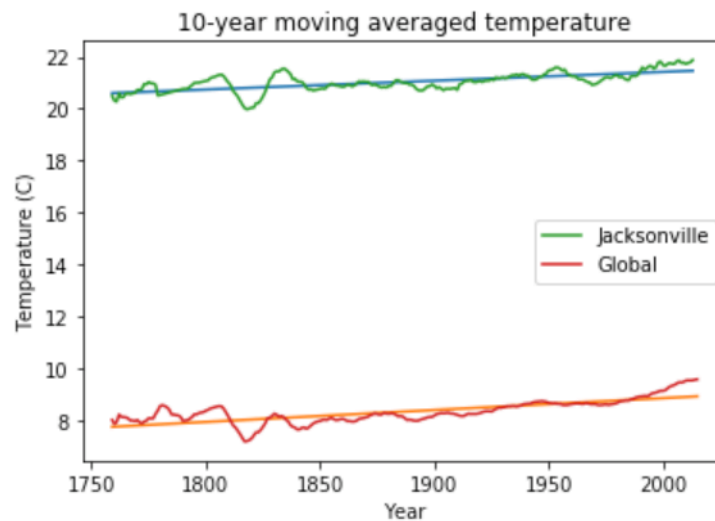
It also outputs the slopes: 0.003385 for Jacksonville and 0.004572 for the global average. Therefore, the global is getting hotter more rapidly.

- *How correlated are the temperatures of Jacksonville and the global average?*

We can use the code below to plot the temperature of Jacksonville vs. the global average and also calculate the correlation coefficient.

```
import statsmodels.formula.api as sm
df_all = pd.merge(df_city, df_global,left_index=True, right_index=True,
how='inner')
plt.plot(df_all['ma_y'],df_all['ma_x'],'.')
plt.xlabel('Temp of global')
plt.ylabel('Temp of Jacksonville')
result = sm.ols(formula="ma_y ~ ma_x", data=df_all).fit()
print(result.summary())
```
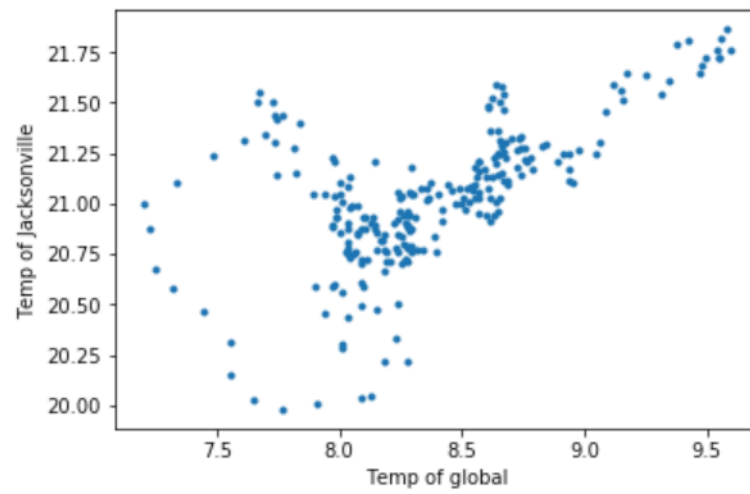
We obtained $r^2 \approx 0.38$. This means a moderate positive correlation.