## Outline

1. What tools did i use?
   Step 1: Used SQL to extract data into a csv file, then downloaded it
   Step 2: Uploaded the csv file into Python, did all the calculation and visualization using Python

2. How did i calculate the moving average?
   I use the function **DataFrame.rolling(x).mean()** with x as the number of years

3. What were my key considerations when deciding how to visualize the trends?
   I want the visualization to be clear, easy to follow and minimize all the excessive info

In [22]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [3]:

```
import warnings
warnings.filterwarnings("ignore")
```

1. Upload csv file

In [4]:

```
df = pd.read_csv("results.csv")
```

In [5]:

```
df.head(5)
```

Out[5]:

| | type | year | avg_temp |
|---|---|---|---|
| **0** | LOCAL | 1840 | 21.00 |
| **1** | LOCAL | 1841 | 21.30 |
| **2** | LOCAL | 1842 | 21.16 |
| **3** | LOCAL | 1843 | 21.26 |
| **4** | LOCAL | 1844 | 20.78 |

In [6]:

```
local_df = df[df['type']=='LOCAL']
local_df
```

Out[6]:

| | type | year | avg_temp |
|---|---|---|---|
| **0** | LOCAL | 1840 | 21.00 |
| **1** | LOCAL | 1841 | 21.30 |
| **2** | LOCAL | 1842 | 21.16 |
| **3** | LOCAL | 1843 | 21.26 |
| **4** | LOCAL | 1844 | 20.78 |
| **...** | ... | ... | ... |
| **169** | LOCAL | 2009 | 22.59 |
| **170** | LOCAL | 2010 | 22.57 |
| **171** | LOCAL | 2011 | 21.03 |
| **172** | LOCAL | 2012 | 22.09 |
| **173** | LOCAL | 2013 | 23.10 |

174 rows × 3 columns

2. Calculate Moving average

In [13]:

```
local_df['moving_avg'] = local_df['avg_temp'].rolling(10).mean()
local_df.dropna(inplace=True)
local_df.loc[:,]
```

Out[13]:

| | type | year | avg_temp | moving_avg |
|---|---|---|---|---|
| **13** | LOCAL | 1853 | 21.51 | 21.021 |
| **14** | LOCAL | 1854 | 21.45 | 21.088 |
| **15** | LOCAL | 1855 | 21.10 | 21.130 |
| **16** | LOCAL | 1856 | 20.69 | 21.074 |
| **17** | LOCAL | 1857 | 20.89 | 21.080 |
| **...** | ... | ... | ... | ... |
| **169** | LOCAL | 2009 | 22.59 | 22.063 |
| **170** | LOCAL | 2010 | 22.57 | 22.140 |
| **171** | LOCAL | 2011 | 21.03 | 22.064 |
| **172** | LOCAL | 2012 | 22.09 | 22.060 |
| **173** | LOCAL | 2013 | 23.10 | 22.108 |

161 rows × 4 columns

In [14]:

```
global_df = df[df['type']=='GLOBAL']
global_df
```

Out[14]:

| | type | year | avg_temp |
|---|---|---|---|
| **174** | GLOBAL | 1750 | 8.72 |
| **175** | GLOBAL | 1751 | 7.98 |
| **176** | GLOBAL | 1752 | 5.78 |
| **177** | GLOBAL | 1753 | 8.39 |
| **178** | GLOBAL | 1754 | 8.47 |
| **...** | ... | ... | ... |
| **435** | GLOBAL | 2011 | 9.52 |
| **436** | GLOBAL | 2012 | 9.51 |
| **437** | GLOBAL | 2013 | 9.61 |
| **438** | GLOBAL | 2014 | 9.57 |
| **439** | GLOBAL | 2015 | 9.83 |

266 rows × 3 columns

In [15]:

```
global_df['moving_avg'] = global_df['avg_temp'].rolling(10).mean()
global_df.dropna(inplace=True)
global_df
```

Out[15]:

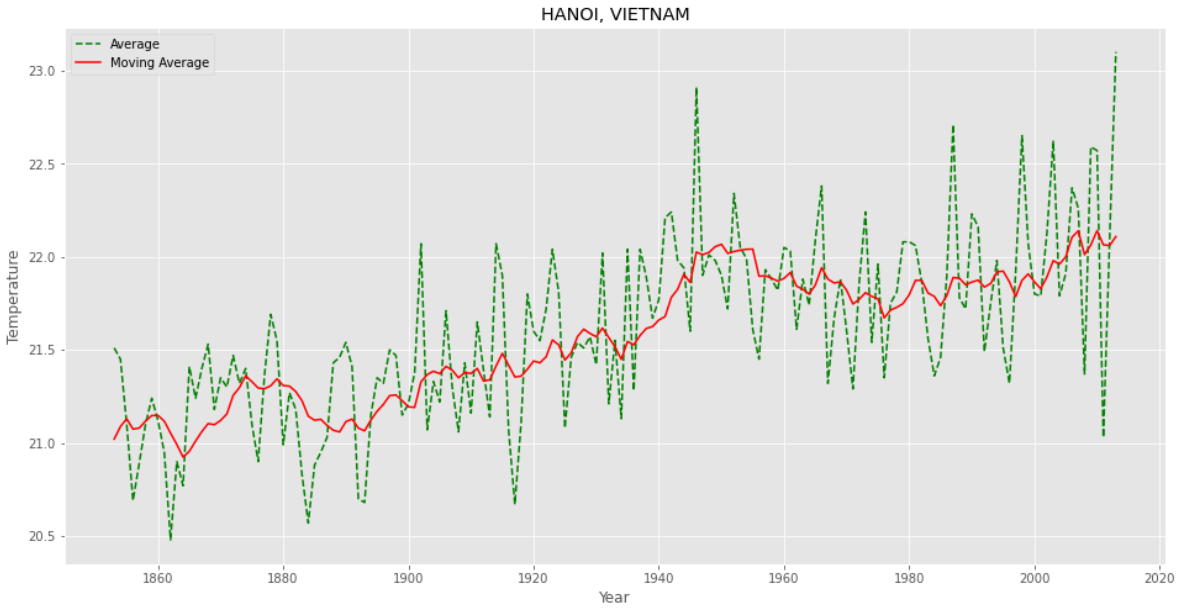
| | type | year | avg_temp | moving_avg |
|---|---|---|---|---|
| **183** | GLOBAL | 1759 | 7.99 | 8.030 |
| **184** | GLOBAL | 1760 | 7.19 | 7.877 |
| **185** | GLOBAL | 1761 | 8.77 | 7.956 |
| **186** | GLOBAL | 1762 | 8.61 | 8.239 |
| **187** | GLOBAL | 1763 | 7.50 | 8.150 |
| **...** | ... | ... | ... | ... |
| **435** | GLOBAL | 2011 | 9.52 | 9.554 |
| **436** | GLOBAL | 2012 | 9.51 | 9.548 |
| **437** | GLOBAL | 2013 | 9.61 | 9.556 |
| **438** | GLOBAL | 2014 | 9.57 | 9.581 |
| **439** | GLOBAL | 2015 | 9.83 | 9.594 |

257 rows × 4 columns

3. Plot the results

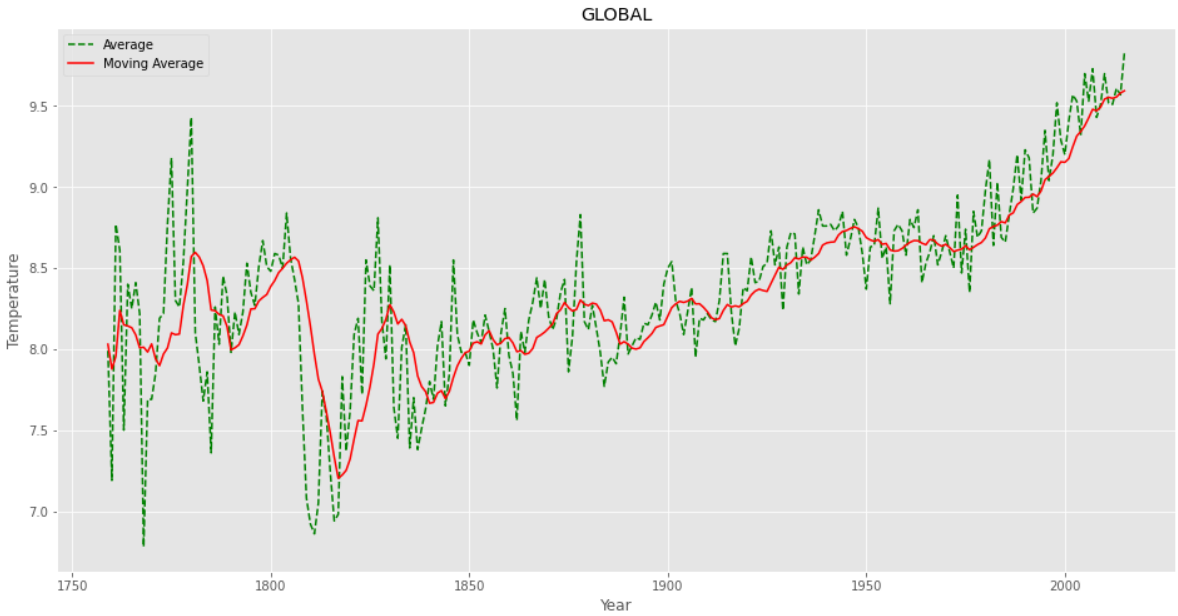
In [20]:

```
# Set names for each category
year = local_df['year']
avg_temp = local_df['avg_temp']
moving_avg_temp = local_df['moving_avg']
test = local_df[['avg_temp','moving_avg']]

plt.subplots(figsize = (16,8))

# Plot 1st set of data
plt.plot(year, avg_temp, 'g--', label = 'Average')

# Plot 2nd set of data
plt.plot(year, moving_avg_temp , 'r-', label = 'Moving Average')

# Create legend
plt.legend(loc = 'upper left') # để legend ở góc trên bên trái

plt.style.use('ggplot')
plt.title('HANOI, VIETNAM')
plt.xlabel('Year')
plt.ylabel('Temperature')

plt.show()
```



In [24]:

```
# Set names for each category
year = global_df['year']
avg_temp = global_df['avg_temp']
moving_avg_temp = global_df['moving_avg']

plt.subplots(figsize = (16,8))

# Plot sets of data
plt.plot(year, avg_temp, 'g--', label = 'Average')
plt.plot(year, moving_avg_temp, 'r-', label = 'Moving Average')

# Create legend
plt.legend(loc = 'upper left')

plt.style.use('ggplot')
plt.title('GLOBAL')
plt.xlabel('Year')
plt.ylabel('Temperature')

plt.show()
```




## 4. Observations

The above charts compare the temperature between Hanoi (Vietnam) and the global. Here are some observations obtained from the charts.

- Overall, both Hanoi and the world witnessed a rise in temperature over the years.
- Hanoi's temperature was much hotter than the world's, ranging from 20.5 to above 23 Celcius degree, whereas the global temperature ranges from 6.5 to nearly 10 Celcius degree.
- While temperature trend in Hanoi tended to increase moderately, the world had seen a fluctuation in temperature during the period of 1770 to 1830.
- Over a few hundred years, Hanoi grew by 1 degree, from 21 to 22, whereas the world grew by nearly 2.6 degree, from 8 to 9.6 approximately