Name: Thu Phung Ngoc Minh
Course: Udacity Data Analyst Nanodegree
Date: 29 April 2020

PROJECT 1: WEATHER TRENDS

1. Export data by SQL from database

- I have looked firstly in the city_list, city where I want to export data. I chose Vietnam is my country and city is Ho Chi Minh City

   SELECT * FROM city_list
   WHERE country = 'Vietnam' AND city = 'Ho Chi Minh City'

- Then, I check in city_data that my city is available with avg_temp

   SELECT * FROM city_data

   WHERE city = ' Ho Chi Minh City'

- Later, I recognized both tables city_data and global_data have the same column name 'avg_temp'. Thus, I changed the name of columns in both tables. City_data.avg_temp to CAT and Global_data.avg_temp to GAT
   ALTER TABLE city_data RENAME COLUMN avg_temp TO CAT
   ALTER TABLE global_data RENAME COLUMN avg_temp TO GAT

- After that, I export data from global_data table and join with city_data table, under CSV file
   SELECT global_data.year, global_data.GAT, city_data.CAT
   FROM global_data
   JOIN city_data ON city_data.year = global_data.year
   WHERE city = 'Ho Chi Minh City'

2. Calculate 10 years moving average of CAT and GAT by command '=average (B2:B11)' and then dragging for all years

| E6 | | ⋮ | ✕ | ✓ | *fx* | =AVERAGE(B2:B11) | |

◢	A	B	C	D	E
1	year	cat	gat	10 year MA CAT	10 years MA GAT
2	1825	27.11	8.39		
3	1826		8.36		
4	1827		8.81		
5	1828		8.17		
6	1829		7.94		
7	1830		8.52		
8	1831		7.64		
9	1832		7.45		
10	1833		8.01		
11	1834		8.15	27.11	8.14
12	1835		7.39		8.04
13	1836		7.7		7.98
14	1837		7.38		7.84
15	1838		7.51		7.77
16	1839	26.69	7.63	26.69	7.74
17	1840	26.63	7.8	26.66	7.67
18	1841	27.02	7.69	26.78	7.67

3. Create again Pivot table for combining MA columns to rest of data, in order to make flexible filter the years when doing visualization.

Row Labels ▼	City AVG Temperature	Global AVG Temperature
2003	27.83	9.53
2004	27.69	9.32
2005	27.88	9.7
2006	28.04	9.53
2007	27.87	9.73
2008	27.61	9.43
2009	27.85	9.51
2010	28.28	9.7
2011	27.68	9.52
2012	28.25	9.51
2013	28.46	9.61
Grand Total	**307.44**	**105.09**

4. Visualize for comparison

- The average temperature of Global and Ho Chi Minh City for years 2003 to 2013
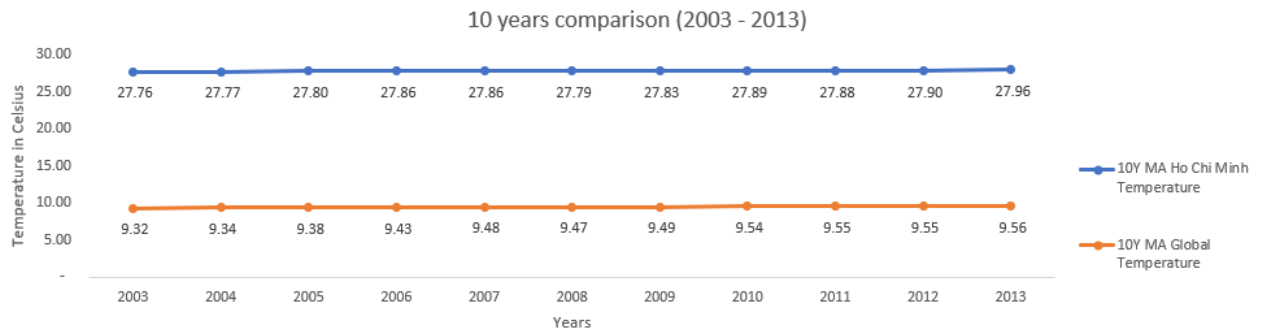
➔ In 10 years 2003 – 2013, the moving average Temperature of Ho Chi Minh City (27.8 to 28 degree) is higher than Global moving average Temperature (9.3 to 9.6 degree). Thus, in 10 years, Ho Chi Minh City is hotter than global

➔ In 10 years, the moving average Temperature of Ho Chi Minh City and Global not significantly increasing, both lines are almost flat.
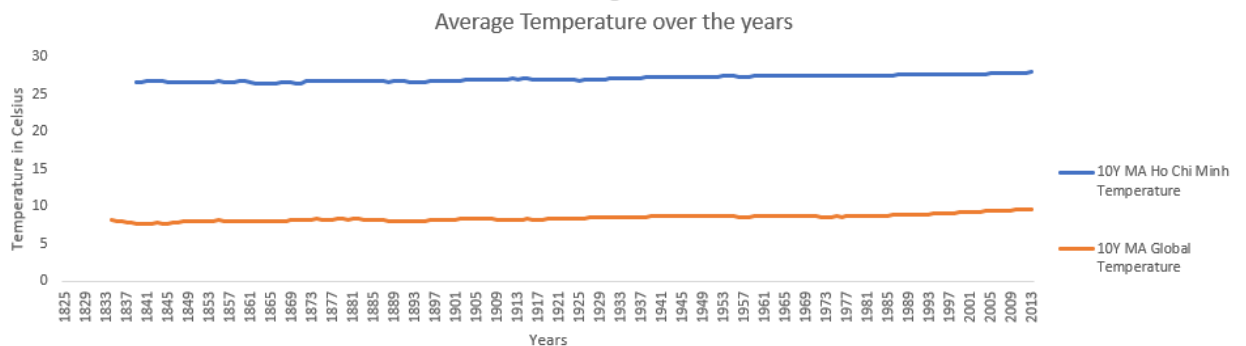
➔ For expanding whole years of database, the AVG Temperature of Ho Chi Minh City is still higher than Global AVG Temperature. That means Ho Chi Minh City is always hotter than Global.

Name: Thu Phung Ngoc Minh
Course: Udacity Data Analyst Nanodegree
Date: 29 April 2020

Ho Chi Minh AVG Temperature vs. Trendline

Global AVG Temperature vs. Trendline
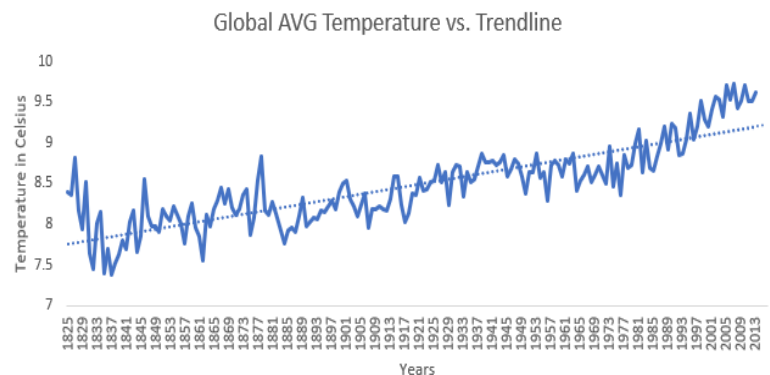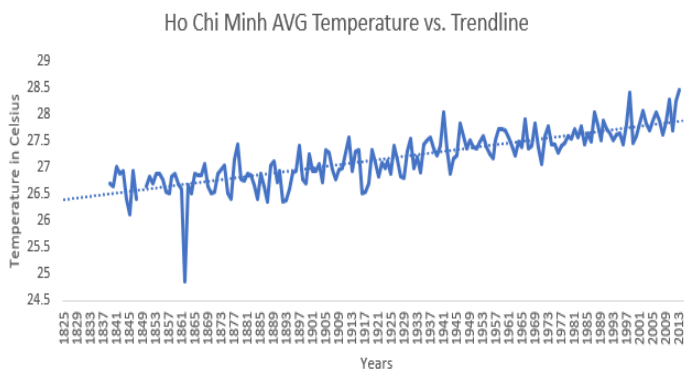
➔ When I separately into two line charts with trendline forecast, I see that the AVG temperature trends over the years are fluctuated in increasing way. The increasing of Global AVG Temperature is faster, specially from 1981 toward to 2013 with increasing significantly.

➔ Both charts show that the earth is hotter and will increase more in the future, it is the reason of strong case of climate change in all countries over the world → it leads to many impacts on economy, especially in Agriculture.

5. Extra question

- Correlation coefficient is the measurement statistic to identify the relationship between 2 variables. There are positive and negative relationship.
- In my opinion, we can estimate trends of my city temperature based on global temperature if:
    ➢ Past data of both databases have strong relationship to each other
    ➢ Measure on the same unit (Fahrenheit or Celsius)
    ➢ Climate change situation happens and impacts same in over the world, same in local country such as El Nino, La Nina, …

6. Review and Lesson

- I made mistake at the beginning of export data. I did not observe well the data tables with 2 different avg_temp columns. Then, after export and I recognized the wrong meaning → Lesson: Observe data carefully and define clearly the definition of each columns in any data table.
- I was thinking about choosing the visualization chart to show the best of comparison. However, I figure out that I should analyze and visualize data from small amount of data (first 10 years) and go compare to big picture of whole years → Lesson: Understanding data and objectives of visualization to approach right story of visualization.

References:

https://stackoverflow.com/questions/174582/how-do-i-rename-a-column-in-a-database-table-using-sql