UDACITY: DATA ANALYSIS NANODEGREE
PROJECT1
EXPLORING WEATHER TRENDS

Isra Khamis Bazzar
October 2, 2018

Outline:

Step1: Extracting Data

Tools: SQL
Resources: Database provided by Udacity through the lesson
Output: CSV files

Step2: Calculating Moving Averages

Tools: Microsoft Excel
Input: CSV files from previous step
Output: New columns with averages to plot

Step3: Drawing Line Charts

Tools: Microsoft Excel
Input: Calculated averages from previous step
Output: Line Chart to study and compare

Conclusion: Observations and comparisons
Further Conclusion:

- Correlation coefficient
- Adding more cities to comparison

Details and explanations Below

Step1: Extracting The Data

Extracting Global Data

Sql Statement:

*SELECT * FROM global_data*

SQL Result Screenshot:

Input		HISTORY ∨	MENU ∨
SCHEMA ↻	1 SELECT * FROM global_data		
city_data ∨			
city_list ∨			
global_data ∨			
	Success!		EVALUATE

Output 266 results   ↓ Download CSV

year	avg_temp
1750	8.72
1751	7.98
1752	5.78
1753	8.39
1754	8.47
1755	8.36
1756	8.85
1757	9.02

Extracting My City(AbuDhabi) Data

Sql Statement:

SELECT year, avg_temp FROM city_data

WHERE city = 'Abu Dhabi'

Result Screenshot

Input		HISTORY ∨	MENU ∨
SCHEMA ↻	1 SELECT year, avg_temp FROM city_data		
city_data ∨	2 WHERE city = 'Abu Dhabi'		
city_list ∨			
global_data ∨			
	Success!		EVALUATE

Output   171 results   ⬇ Download CSV

year	avg_temp
1843	26.04
1844	26.26
1845	
1846	
1847	
1848	25.83
1849	26.01
1850	25.69

Extracting Data for more Cities(Random Choices around the globe)

SQL Statements

Istanbul:
SELECT year, avg_temp FROM city_data
WHERE city = 'Istanbul'

Rabat:
SELECT year, avg_temp FROM city_data
WHERE city = 'Rabat'

Mecca:
SELECT year, avg_temp FROM city_data
WHERE city = 'Mecca'

Paris:
SELECT year, avg_temp FROM city_data
WHERE city = 'Paris'

Step2: Open Data using Excel and Calculating Moving Averages
After extracting Data I saved results as CSV files and opened them using Microsoft Excel,
Then for more readability I put global data and data for my residence country(Abu Dhabi) in the same sheet

Calculating Moving Averages
I chose to calculate moving average by decade (for each 10 years) starting from year 1855 since there is no regular data before that for Abu Dhabi in the database, below is a snapshot from table with weather data for both Global and AbuDhabi and calculated moving average.

*   ** To calculate Moving average for each year the value of moving average (10-Y MA) will be the average for weather data from last 10 years including year of destination.*

Year	Global_avg_ter	10-years MA for global weather	AbuDhabi	10-years MA for AbuDhabi	Difference
1870	8.2	8.129	26.24	25.793	17.664
1871	8.12	8.156	26.17	25.9	17.744
1872	8.19	8.219	26.16	25.982	17.763
1873	8.35	8.243	26.42	26.262	18.019
1874	8.43	8.288	26.3	26.292	18.004
1875	7.86	8.256	26.13	26.279	18.023
1876	8.08	8.235	26.19	26.289	18.054
1877	8.54	8.245	26.59	26.304	18.059
1878	8.83	8.303	26.59	26.33	18.027
1879	8.17	8.277	26.09	26.288	18.011
1880	8.12	8.269	26.2	26.284	18.015
1881	8.27	8.284	26.75	26.342	18.058
1882	8.13	8.278	26.05	26.331	18.053
1883	7.98	8.241	26.3	26.319	18.078
1884	7.77	8.175	26.15	26.304	18.129
1885	7.92	8.181	26.1	26.301	18.12
1886	7.95	8.168	26.38	26.32	18.152
1887	7.91	8.105	26.08	26.269	18.164
1888	8.09	8.031	26.3	26.24	18.209
1889	8.32	8.046	26.61	26.292	18.246
1890	7.97	8.031	26.2	26.292	18.261

FIGURE 1: MOVING AVERAGE SNAPSHOT

Step3: Line Charts

Now, to draw line charts I used Insert Line chart feature with table of data in Excel

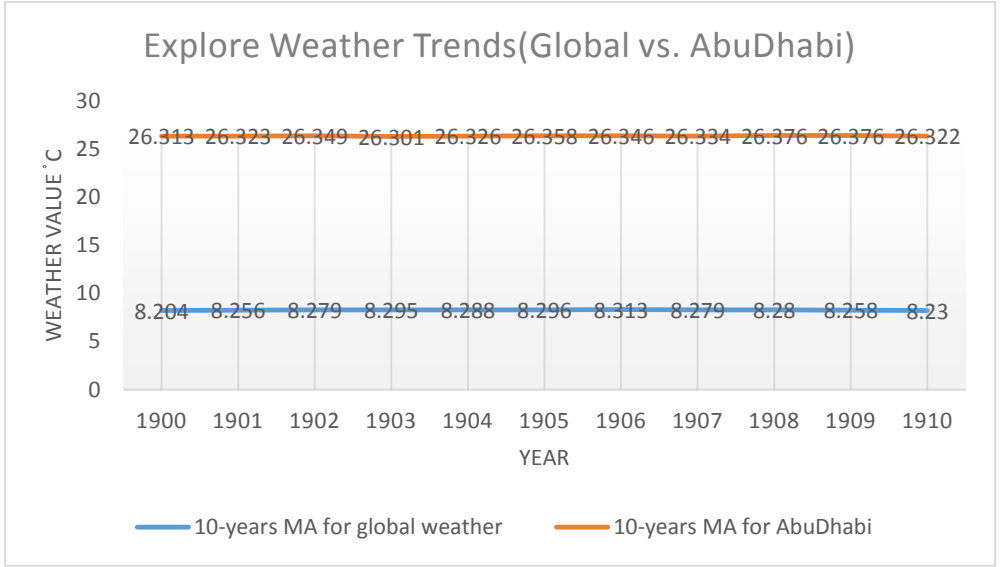Chart 1: For all years from 1870- 2013 (since data starts from 1861; 10-Years Moving Average values will start from 1870)
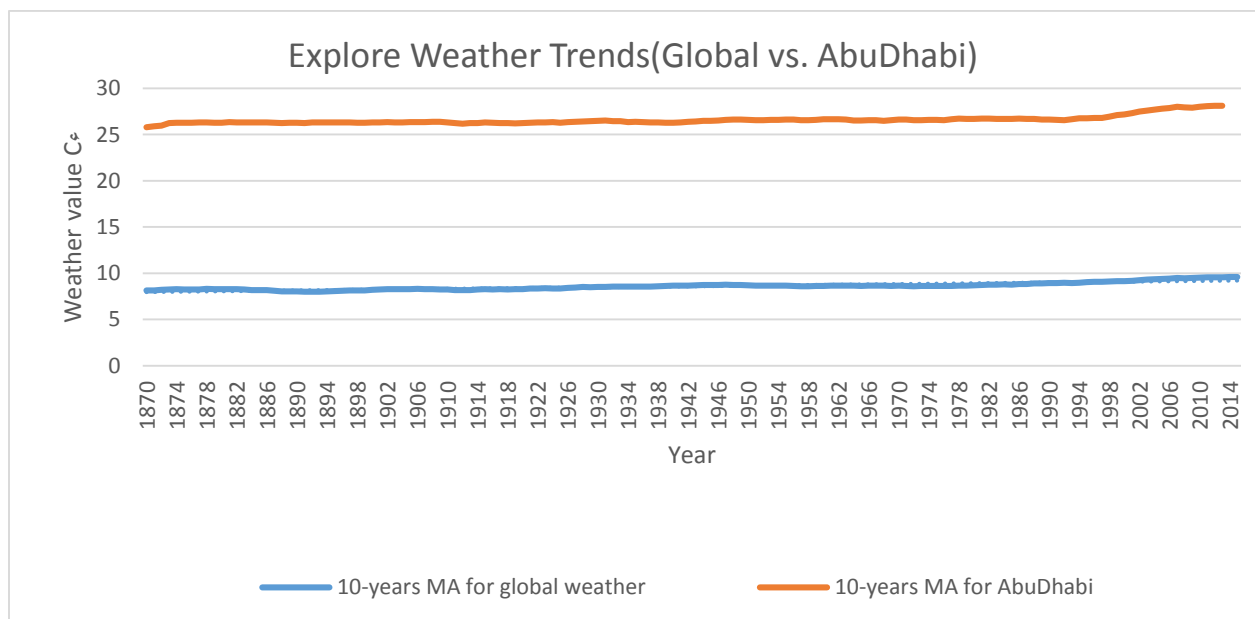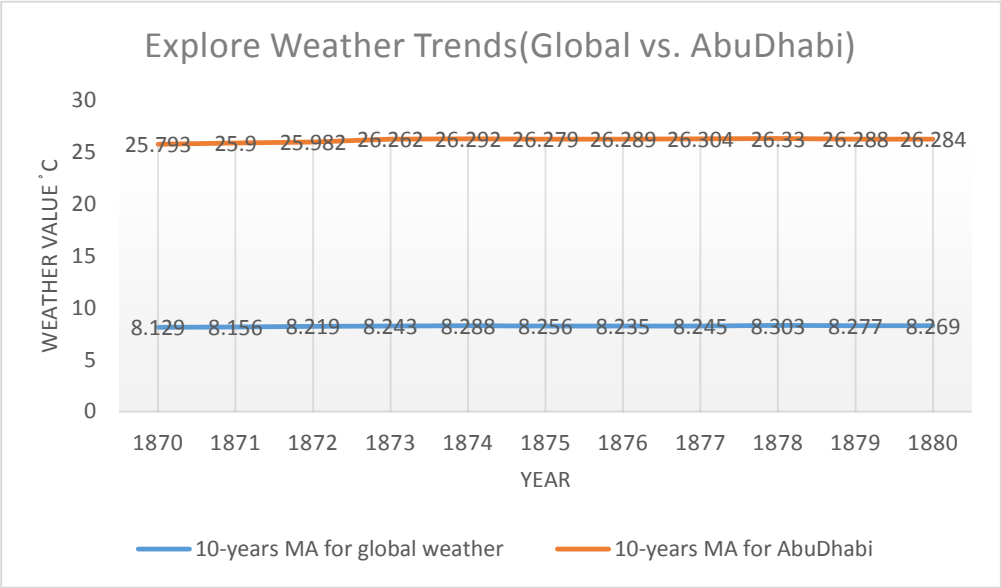
Conclusion:

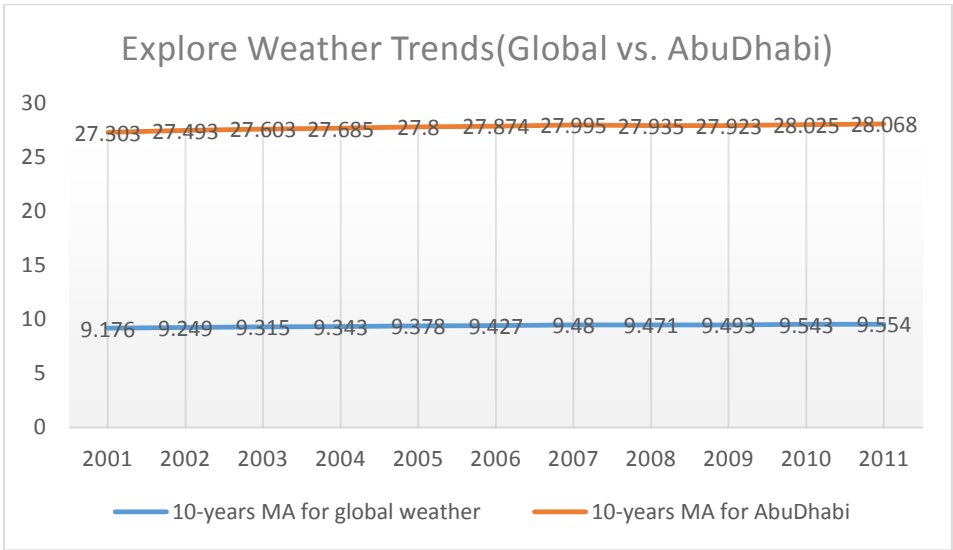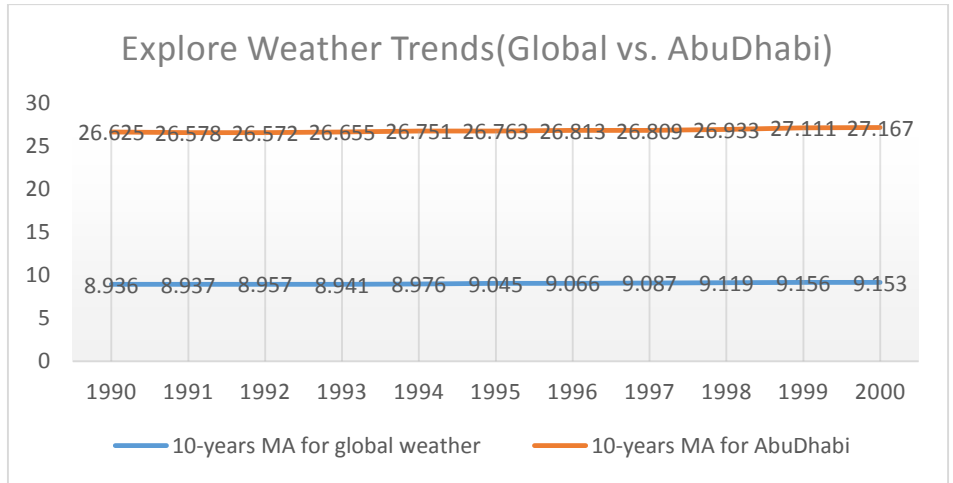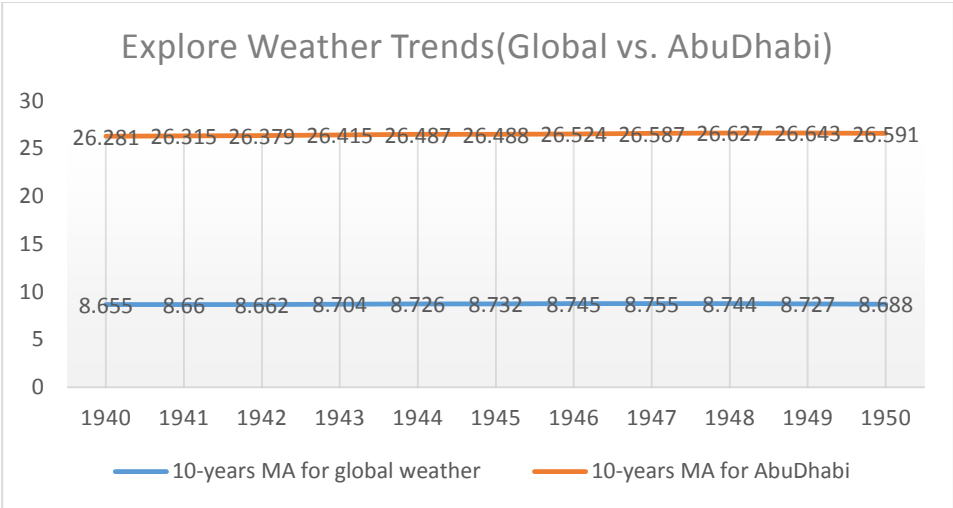By looking at this chart we can notice some points:

- For global line(blue line) values are mainly based around 9 degrees starting from around 8 degrees reaching around 10 degrees by the end of the chart, while for Abu Dhabi line(orange line) values are mainly around 26 degrees, starting from around 25.7 degrees reaching around 28 degrees for last years
- For both lines: all years the line is almost horizontal (slope ~0); which means values over the years are almost the same with small differences (tenths of degree), but it's also noticeable at the right end of the chart (for last years) that the lines is getting little bit up after year 1995, which can be interpreted as: temperature is rising faster year after year, (I think this is possibly caused by Global Warming!)

- Again by looking since both lines are almost horizontal, and even when getting up at the last both lines are getting up together, we can guess the difference between values for global and Abu Dhabi over the years is almost the same, So at a specific year we can guess one value given the other; below is a small table to show the difference between two lines at different chosen years;

Year	10-Years MA For global	10-Years MA for Abu Dhabi	Difference
1880	8.269	26.284	18.015
1940	8.655	26.281	17.626
1970	8.647	26.621	17.974
2000	9.153	27.167	18.014
2010	9.543	28.025	18.482

For more accurate value I used excel table, I calculated the difference between Abu Dhabi value and global value for all years (1871-2013), then I took the average, the result = 17.998 ~= 18; So to estimate Abu Dhabi Average weather value at any year, *we add 18 degrees* to the global average value at that year.

To have closer look at the values and differences between global values and Abu Dhabi values; I have inserted 5 more line charts each for different decade;

Explore Weather Trends(Global vs. AbuDhabi)

Chart 1: Line chart showing 10-years moving average weather values from 1870 to 1880.

AbuDhabi (orange line): 25.793, 25.9, 25.982, 26.262, 26.292, 26.279, 26.289, 26.304, 26.33, 26.288, 26.284

Global weather (blue line): 8.129, 8.156, 8.219, 8.243, 8.288, 8.256, 8.235, 8.245, 8.303, 8.277, 8.269

WEATHER VALUE °C (y-axis: 0–30)
YEAR (x-axis: 1870–1880)

Legend: 10-years MA for global weather — 10-years MA for AbuDhabi

Explore Weather Trends(Global vs. AbuDhabi)

Chart 2: Line chart showing 10-years moving average weather values from 1900 to 1910.

AbuDhabi (orange line): 26.313, 26.323, 26.349, 26.301, 26.326, 26.358, 26.346, 26.334, 26.376, 26.376, 26.322

Global weather (blue line): 8.204, 8.256, 8.279, 8.295, 8.288, 8.296, 8.313, 8.279, 8.28, 8.258, 8.23

WEATHER VALUE °C (y-axis: 0–30)
YEAR (x-axis: 1900–1910)

Legend: 10-years MA for global weather — 10-years MA for AbuDhabi

Explore Weather Trends(Global vs. AbuDhabi)

26.281 26.315 26.379 26.415 26.487 26.488 26.524 26.587 26.627 26.643 26.591

8.655 8.66 8.662 8.704 8.726 8.732 8.745 8.755 8.744 8.727 8.688

1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950

— 10-years MA for global weather   — 10-years MA for AbuDhabi

Explore Weather Trends(Global vs. AbuDhabi)

26.625 26.578 26.572 26.655 26.751 26.763 26.813 26.809 26.933 27.111 27.167

8.936 8.937 8.957 8.941 8.976 9.045 9.066 9.087 9.119 9.156 9.153

1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000

— 10-years MA for global weather   — 10-years MA for AbuDhabi

Explore Weather Trends(Global vs. AbuDhabi)

27.303 27.493 27.603 27.685 27.8 27.874 27.995 27.935 27.923 28.025 28.068

9.176 9.249 9.315 9.343 9.378 9.427 9.48 9.471 9.493 9.543 9.554

2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011

— 10-years MA for global weather   — 10-years MA for AbuDhabi

Further Conclusion
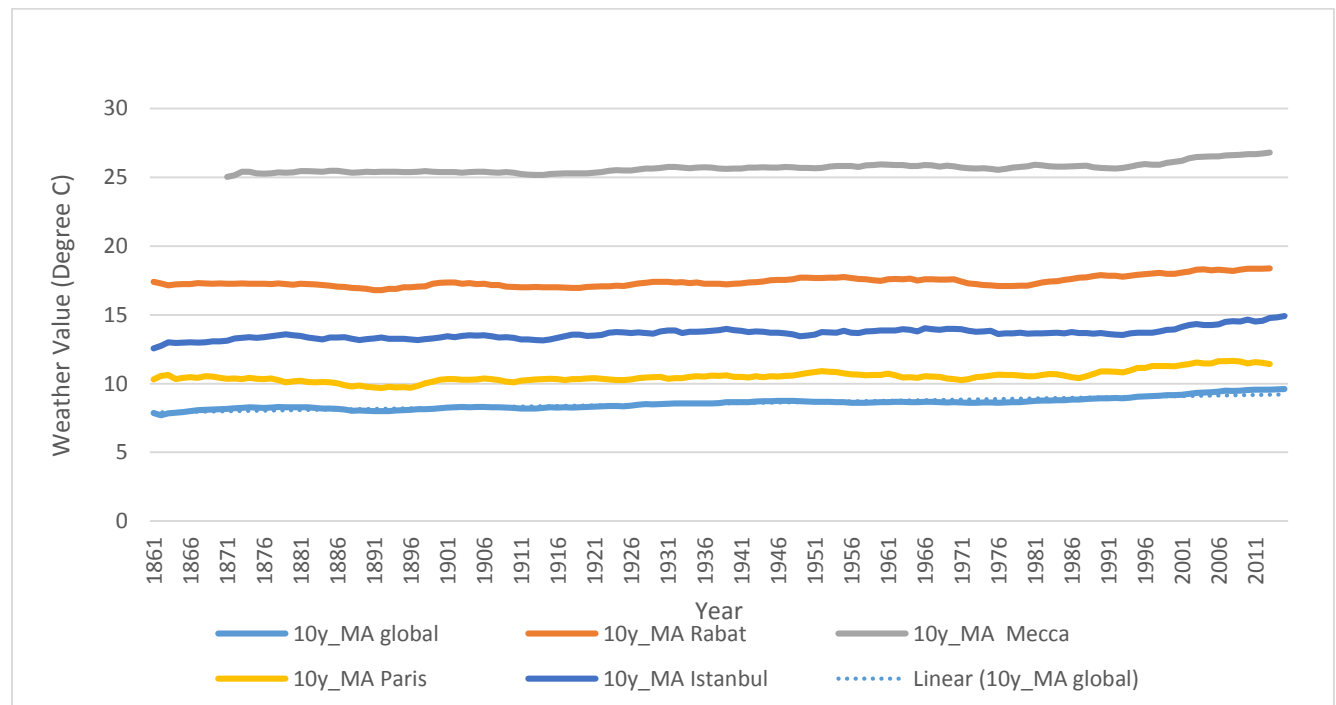
Correlation Coefficient:

Definition: "Correlation between sets of data is a measure of how well they are related. The most common measure of correlation in stats is the Pearson Correlation. The full name is the **Pearson Product Moment Correlation (PPMC)**. It shows the linear relationship between two sets of data. In simple terms, it answers the question, *Can I draw a line graph to represent the data?* Two letters are used to represent the Pearson correlation: Greek letter rho (ρ) for a population and the letter "r" for a sample."

Reference: https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/#Excel

For this analysis, to calculate the Pearson correlation coefficient between the two lines(Global weather average, Abu Dhabi weather average), I used a function in Microsoft excel called (CORREL), it takes two data sets and gives a number result, For our data the result = 0.897551, which means we have very strong positive linear relationship between the two datasets.

Comparing More Cities

For Further comparison I Added more cities data compared to global data, then calculated 10-years calculated moving average for each city and represented this in a linear chart with all lines

THE END