Assessment Task 4 – Data Science Practical Assignment

ICT Unit 2

Name: Faisha Putri Indira Surjatin

Student ID No: 16082018

Class: ICT 2-1

Lecturer: Ms. Vani

Date of Submission: 11 May 2017

Table of Contents

Table of Figures and Tables

Chapter 1: Introduction

The case study we were given required us to conduct an analysis on atmospherical data collected in different states around Malaysia in the interest of substantiating or refuting certain hypotheses about the weather and climate in the country. We decided to collect data from Kuala Lumpur, Kota Kinabalu, Johor Bahru and Kangar everyday from 1st January to 30th April 2017 in order to be able to compare weather in different parts of Malaysia over a sizable time period. This analysis will develop our understanding of Malaysia's atmosphere in order to make reliable forecasts. The following hypothesis statements were tested:

1. The average temperature increases over time from January to April.
2. The total precipitation of a city in East Malaysia is higher than the total precipitation of any city in West Malaysia.
3. The higher the average humidity, the higher the average dew point.
4. The lower the latitude of a city, the higher the average UV index.

Chapter 2: Dataset

We used multiple techniques to collect the data required for the analysis. Such techniques included observation, where we closely monitored weather forecasts everyday and looked at the different parameters; reporting, where we referred to established weather reporting sites such as Wunderground and World Weather Online for atmospherical data; and researching, which we resorted to for data that was difficult to acquire. For example, we derived the UV index using a formula and average values calculated from past data (United States Environmental Protection Agency n.d). Research also helped to verify that the sources we used for reporting were reliable as we looked for matching data on multiple websites.

Date	Max temp (°C)	Avg temp (°C)	Min temp (°C)	Max dew point (°C)	Avg dew point (°C)	Min dew point (°C)	Max humidity (%)	Avg humidity (%)	Min humidity (%)	Precipitation (mm)	Pressure (hPa)	Wind speed (km/h)	Cloud (%)	Visibility (km)	Estimated UV index	UV Exposure Category
10/02/2017	34	29	24	24	23	21	89%	73%	36%	0.00	1009	8	12%	10	9	Very High
11/02/2017	33	29	24	24	22	20	85%	66%	39%	2.03	1011	11	11%	10	10	Very High
12/02/2017	33	28	24	22	21	19	74%	62%	36%	0.00	1012	13	32%	10	7	High
13/02/2017	33	28	24	24	22	21	94%	74%	44%	0.25	1013	10	32%	10	7	High
14/02/2017	33	28	24	23	22	21	94%	73%	47%	3.05	1014	8	21%	10	8	Very High
15/02/2017	33	28	24	24	22	21	100%	72%	48%	0.25	1013	6	24%	10	8	Very High
16/02/2017	33	28	24	23	21	19	78%	64%	45%	0.00	1012	11	23%	10	8	Very High
17/02/2017	34	29	24	25	22	20	91%	69%	46%	2.03	1011	11	32%	10	7	High

Table 1: Example of raw data for the city of Kuala Lumpur

Our raw data consisted of various types of data. UV exposure category is an example of ordinal qualitative data. The other parameters are all quantitative. Whilst UV index is discrete, temperature, dew point, humidity, precipitation, pressure, wind speed, cloud cover and visibility are all continuous.

Chapter 3: Data Analysis

1. The average temperature increases over time from January to April.

By analyzing long-term trends in temperature, we can deduce Malaysia's temperal climate, allowing us to easily predict how the temperature each month will change in future years. Proof of this hypothesis would indicate that all four cities follow the same trend of an increasing average temperature over time from January to April.

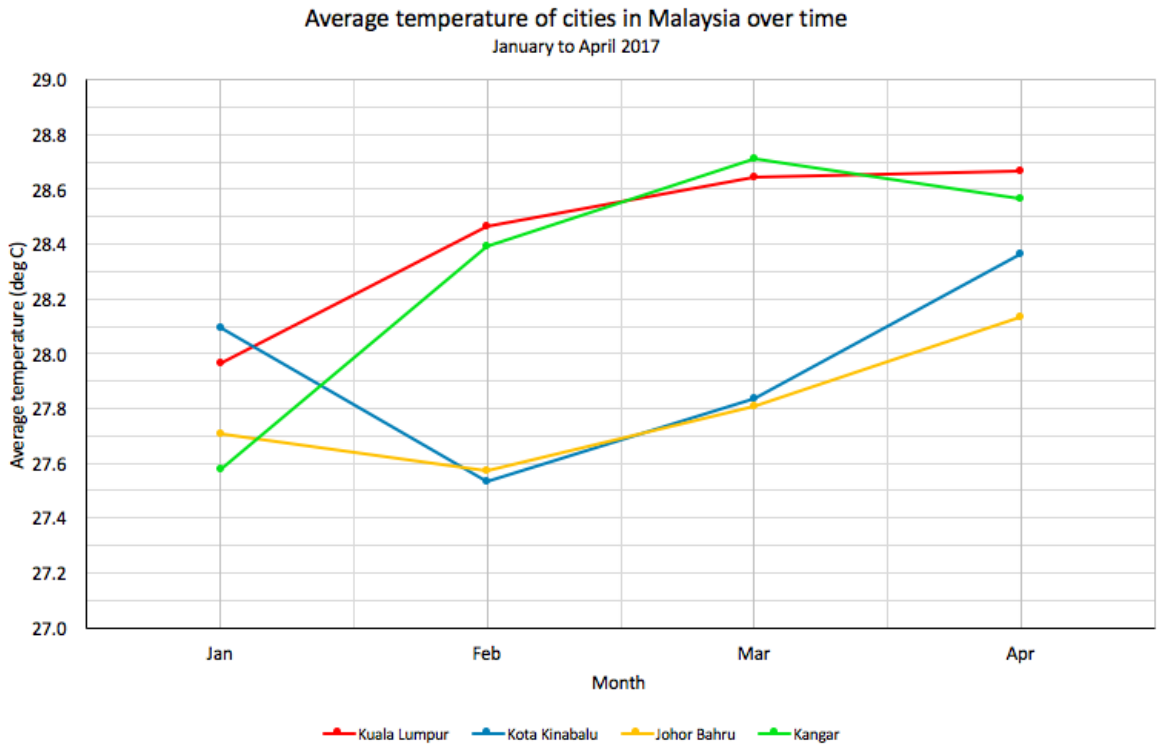


Figure 1: Line chart of average temperature of cities in Malaysia over time

Kuala Lumpur shows an increase in average temperature from January to April, however for the cities Kota Kinabalu and Johor Bahru there is a dip in February before the average temperature increases. On the other hand, Kangar's average temperature increases from January through to March, then decreases in April. Therefore, although it holds true for Kuala Lumpur, the temperature trends in Kota Kinabalu, Johor Bahru and Kangar disprove this hypothesis.

Although Malaysia is located on the Northern hemisphere – and should therefore generally experience an increase in temperature from January to April – the graph shows a variation in temperature for each city. This may be because the country's closeness to the equator means that the different cities may not experience the exact same seasonal weather pattern.

2. The total precipitation of a city in East Malaysia is higher than the total precipitation of any city in West Malaysia.

Investigating the precipitation will improve our understanding of the various monsoon seasons in Malaysia that take place in different regions at different times of the year. If this hypothesis were to be proven within the scope of our data collection, Kota Kinabalu would have the lowest total precipitation against the three other cities we collected data from.

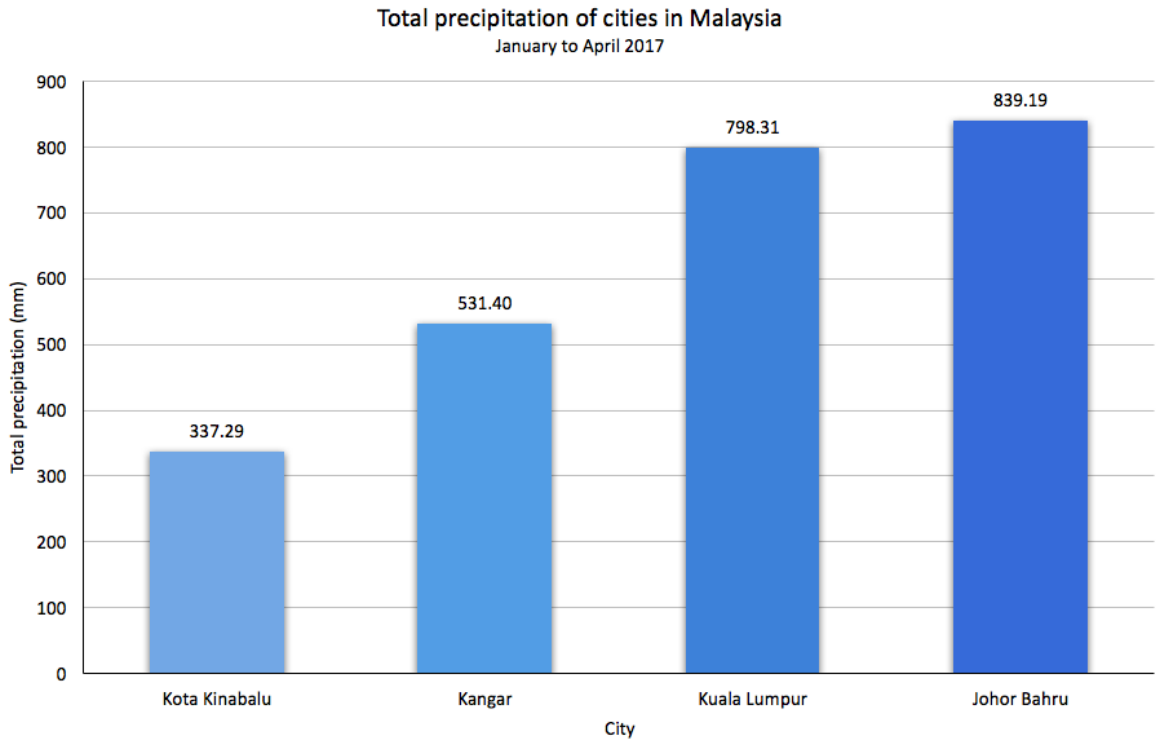


Figure 2: Bar chart of total precipitation of cities in Malaysia

The chart shows that Kota Kinabalu had the least amount of precipitation in comparison to Kuala Lumpur, Johor Bahru and Kangar, which are all cities in West Malaysia. Hence, this hypothesis was substantiated.

This is because from October to March the country faces a Northeast Monsoon which mostly affects the east coast of peninsular Malaysia, hence Johor Bahru exhibiting the most rainfall. This monsoon, however, has no effect on the island of Borneo, therefore Kota Kinabalu and the rest of Sabah experience little precipitation during these months (Wonderful Malaysia n.d).

3. The higher the average humidity, the higher the average dew point.

Identifying how humidity affects the dew point - if there is a relationship between the two - enables us to reliably forecast one factor based on the other. This hypothesis was tested by plotting the average dew point against the average humidity on the same day.

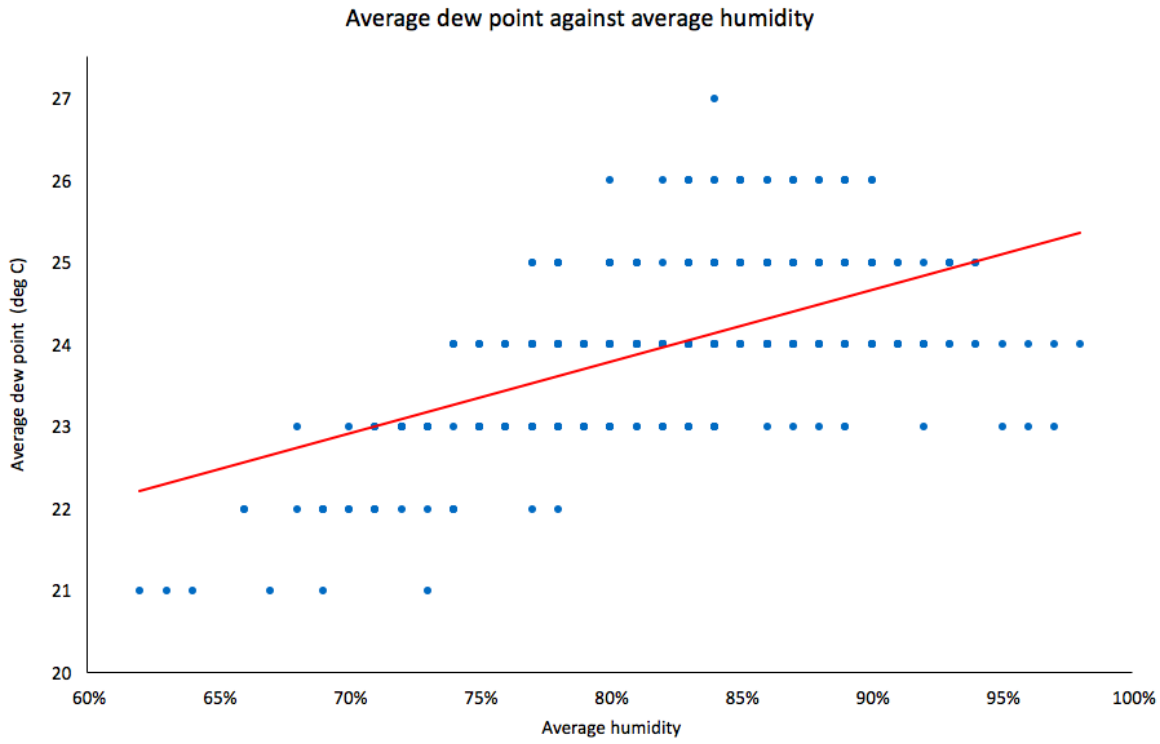


Figure 3: Scatter plot of average dew point against average humidity

The scatter graph generated from the data shows there is a weak positive correlation between the average humidity and the average dew point, therefore this hypothesis was substantiated.

This is due to the fact that the dew point is the temperature at which, given pressure is constant, the air cannot hold any more water vapour, thus the relative humidity is 100% (National Weather Service n.d). This results in a positive correlation between average dew point and humidity. The weak correlation can be attributed to the fact that the pressure in the actual environment is changing.

4. The lower the latitude of a city, the higher the average UV index.

UV index is important to humans – the higher the UV index, the more damage can be done to our skin from ultraviolet radiation. Establishing a relationship between latitude and UV index can help us estimate the UV index based the city's distance from the equator.

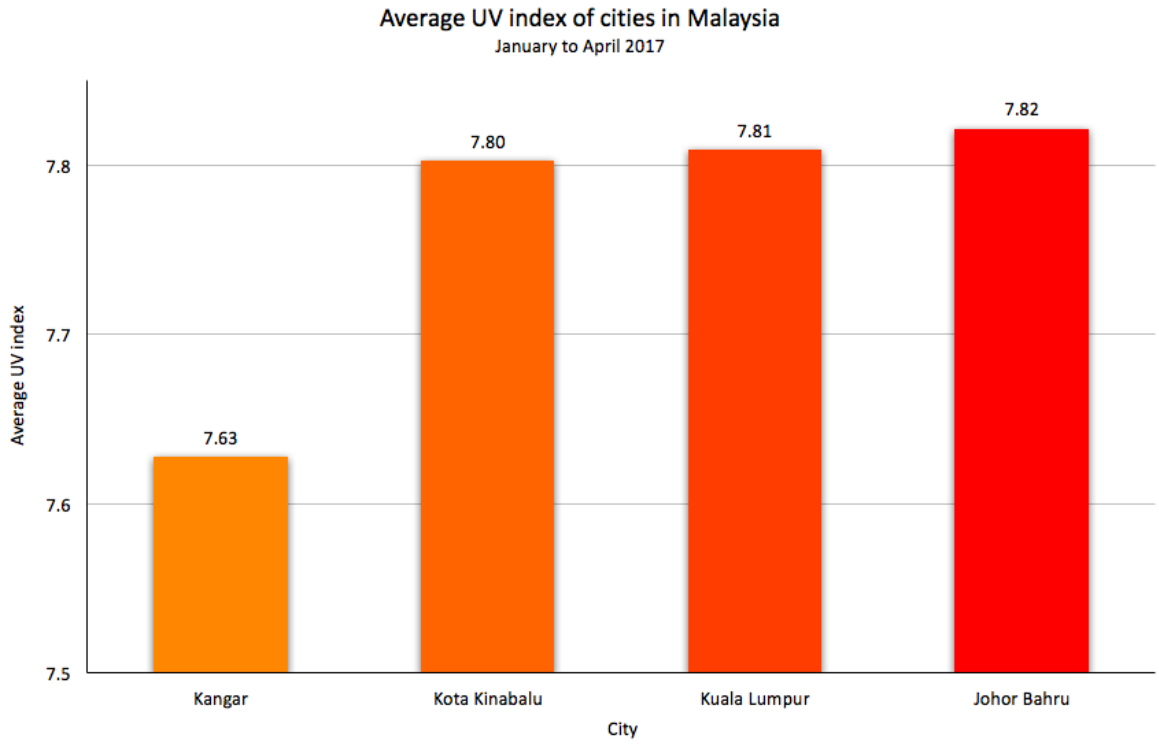


Figure 4: Bar chart of average UV index of cities in Malaysia

Although all the cities are shown to have a high average UV index, the chart shows that the order of cities in ascending average UV index is Kangar, Kota Kinabalu, Kuala Lumpur and Johor Bahru. This is also the order of descending latitude, meaning Johor Bahru is the closest to the equator. Thus, the hypothesis was proven to be true.

This can be explained by the fact that around the March equinox, the Sun is shining directly on the equator (Time and Date n.d). Therefore, the closer the city is to the equator, the higher the intensity of UV radiation.

Chapter 4: Conclusion

In conclusion, through analysing the data with the tools made available through Microsoft Excel, three out of four hypotheses about the weather and climate in Malaysia were substantiated, whilst one was refuted. In future, if this analysis were to be continued, we would recommend using a larger sample of the country and collecting data for at least a year to improve the accuracy and reliability of the findings. The findings were presented in an infographic made using Piktochart (see appendix B).

References

National Weather Service n.d, *Dew Point vs Humidity*, viewed 8 May 2017, http://www.weather.gov/arx/why_dewpoint_vs_humidity

Time and Date n.d, *March Equinox*, viewed 8 May 2017, https://www.timeanddate.com/calendar/march-equinox.html

United States Environmental Protection Agency n.d, *Calculating the UV Index*, viewed 8 May 2017, https://www.epa.gov/sunsafety/calculating-uv-index-0

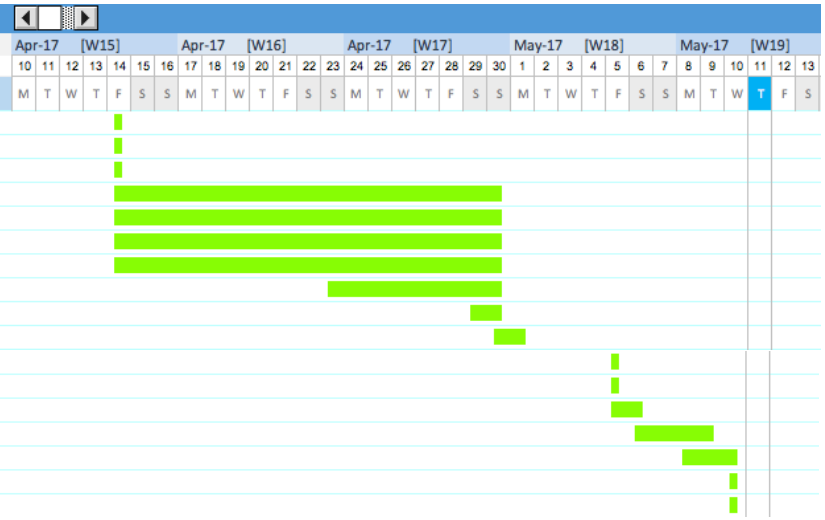
Wonderful Malaysia n.d, *Climate and monsoon in Malaysia*, viewed 8 May 2017, http://www.wonderfulmalaysia.com/malaysia-climate-and-monsoon.htm

Appendix A: Gantt chart

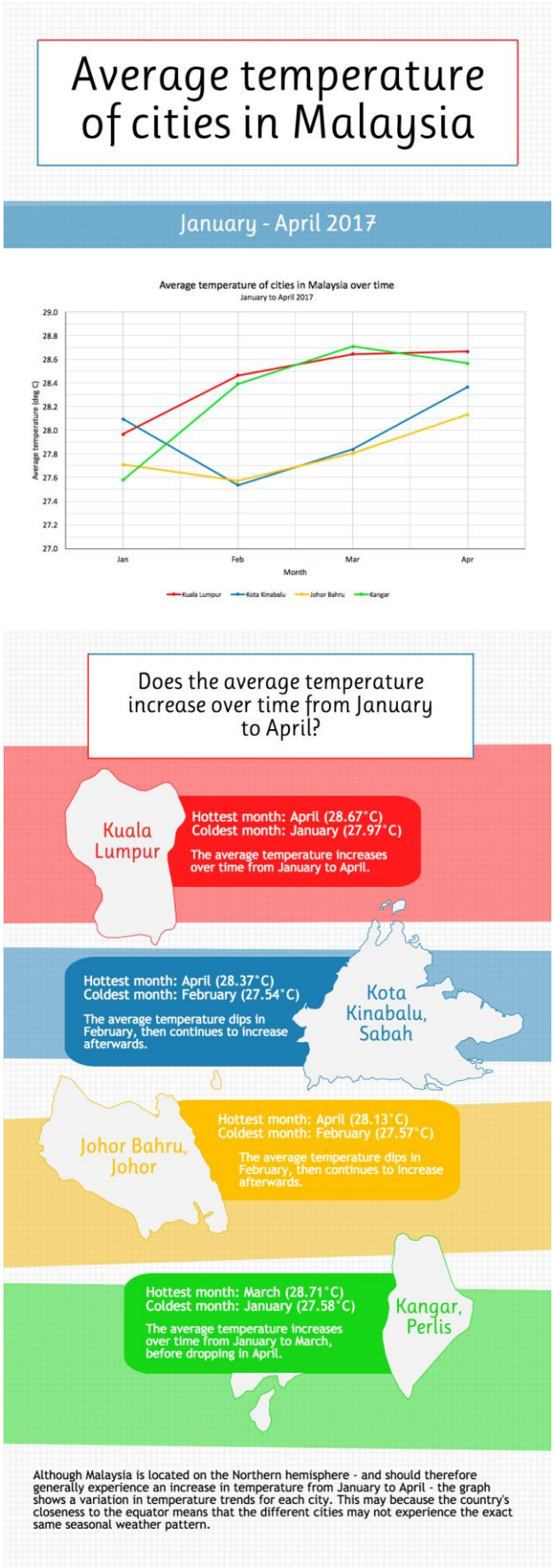
Malaysia Weather Analysis

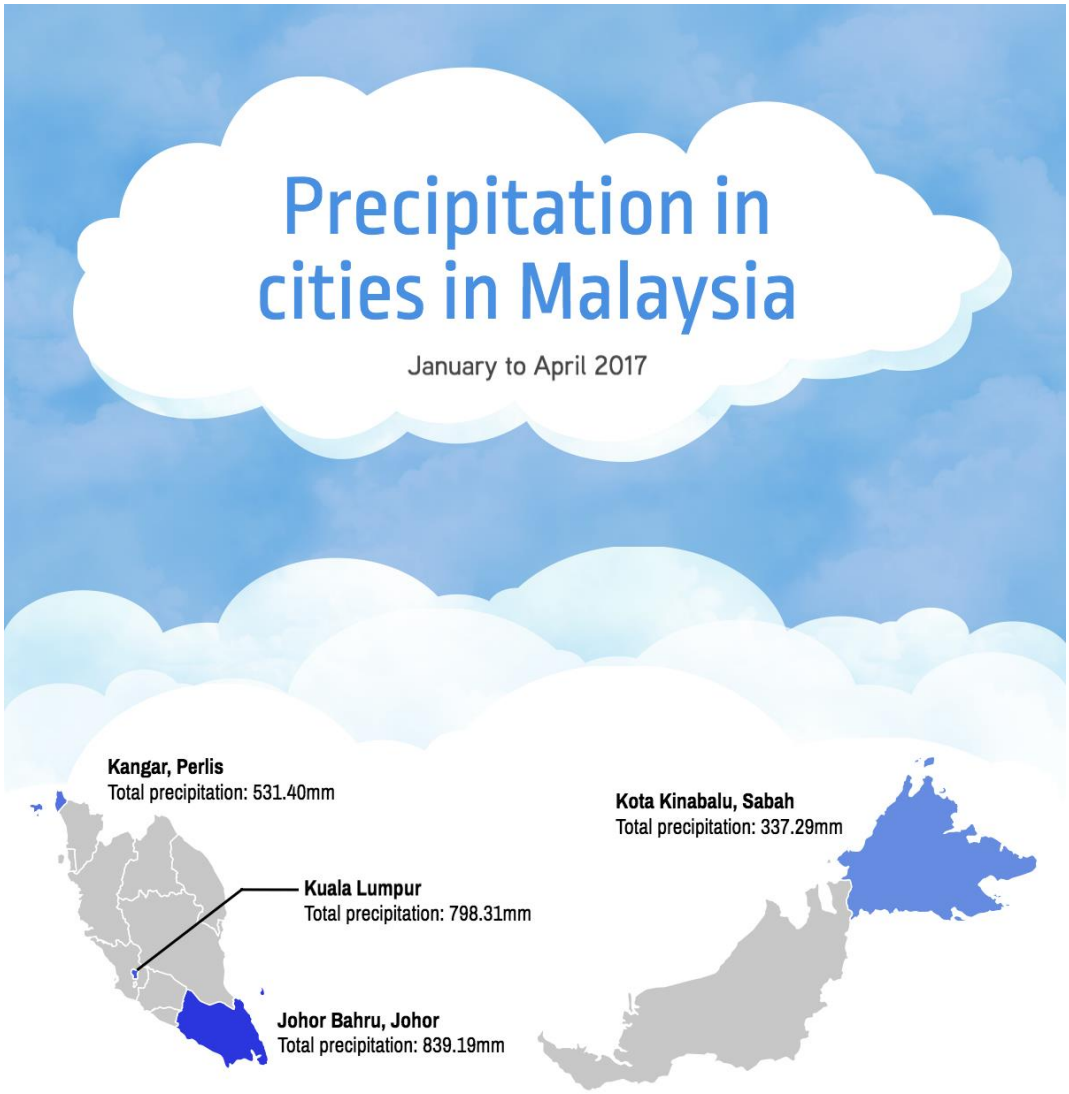
Project Lead: Faisha Surjatin

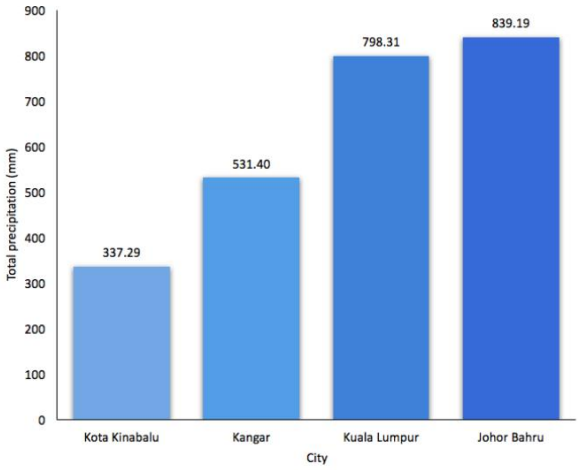
WBS	Task Name	Resource Names	Start	Finish	Duration	Percent Complete
1	Task segregation - Faisha, Izad		Fri 14/Apr/17	Fri 14/Apr/17	1	100%
2	Develop Gantt chart - Faisha, Izad	Microsoft Excel	Fri 14/Apr/17	Fri 14/Apr/17	1	100%
3	Design workbook - Faisha, Izad	Microsoft Excel	Fri 14/Apr/17	Fri 14/Apr/17	1	100%
4	Collect KL data - Faisha	Google Chrome, Microsoft	Fri 14/Apr/17	Sun 30/Apr/17	11	100%
5	Collect Kangar data - Faisha	Google Chrome, Microsoft	Fri 14/Apr/17	Sun 30/Apr/17	11	100%
6	Collect Kota Kinabalu data - Izad	Google Chrome, Microsoft	Fri 14/Apr/17	Sun 30/Apr/17	11	100%
7	Collect Johor Bahru data - Izad	Google Chrome, Microsoft	Fri 14/Apr/17	Sun 30/Apr/17	11	100%
8	Individual data analysis - Faisha	Microsoft Excel	Sun 23/Apr/17	Sun 30/Apr/17	5	100%
9	Complete testing - Faisha, Izad	Microsoft Excel	Sat 29/Apr/17	Sun 30/Apr/17	1	100%
10	Create infographic - Faisha	Piktochart	Sun 30/Apr/17	Mon 01/May/17	1	100%
1	Format report - Faisha	Microsoft Word	Fri 05/May/17	Fri 05/May/17	1	100%
2	Write introduction - Faisha	Microsoft Word	Fri 05/May/17	Fri 05/May/17	1	100%
3	Write dataset - Faisha	Microsoft Word	Fri 05/May/17	Sat 06/May/17	1	100%
4	Write data analysis - Faisha	Microsoft Word	Sat 06/May/17	Tue 09/May/17	2	100%
5	Write conclusion - Faisha	Microsoft Word	Mon 08/May/17	Wed 10/May/17	3	100%
6	Add references - Faisha	Microsoft Word	Wed 10/May/17	Wed 10/May/17	1	100%
7	Add appendix - Faisha	Microsoft Word	Wed 10/May/17	Wed 10/May/17	1	100%



Appendix B: Infographic







Wettest city from January to April:

Johor Bahru

Driest city from January to April:

Kota Kinabalu

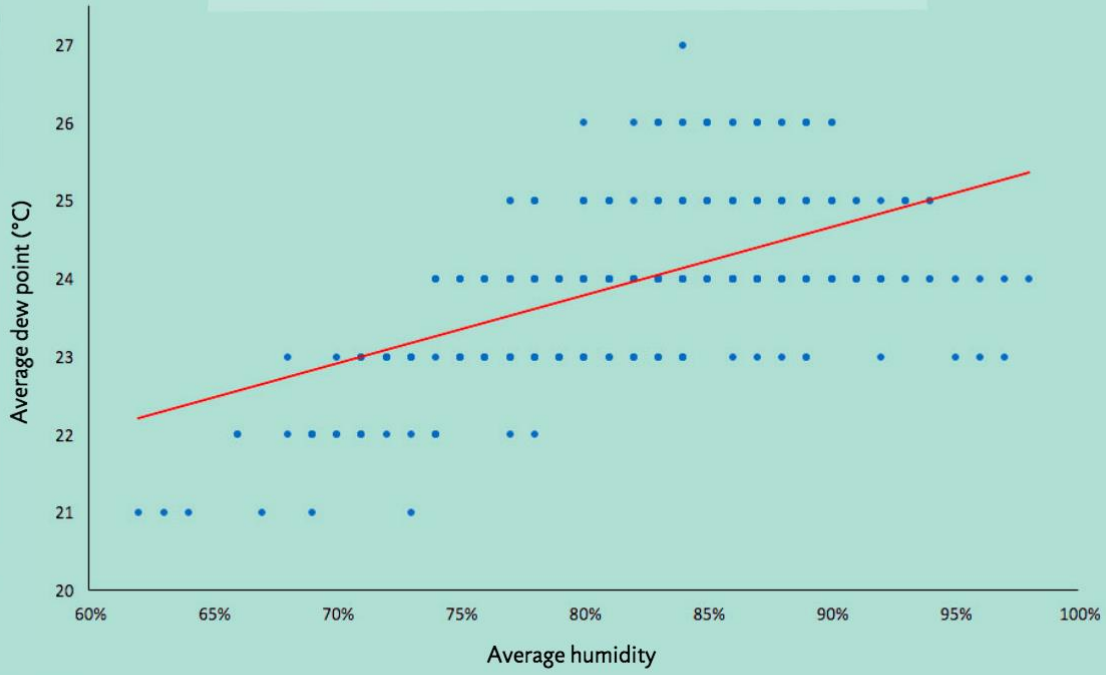
The total precipitation in Kota Kinabalu from January to April is lower than the total precipitation in Kuala Lumpur, Johor Bahru and Kangar.

This is because from January to April Kota Kinabalu experiences a dry season, whilst cities in West Malaysia are experiencing monsoon season.

Humidity and Dew Point

How does humidity affect the average dew point?

Average dew point against average humidity*



*Dataset comprises of humidity and dew point readings in Kuala Lumpur, Kota Kinabalu, Johor Bahru and Kangar from 1st January to 30th April 2017.

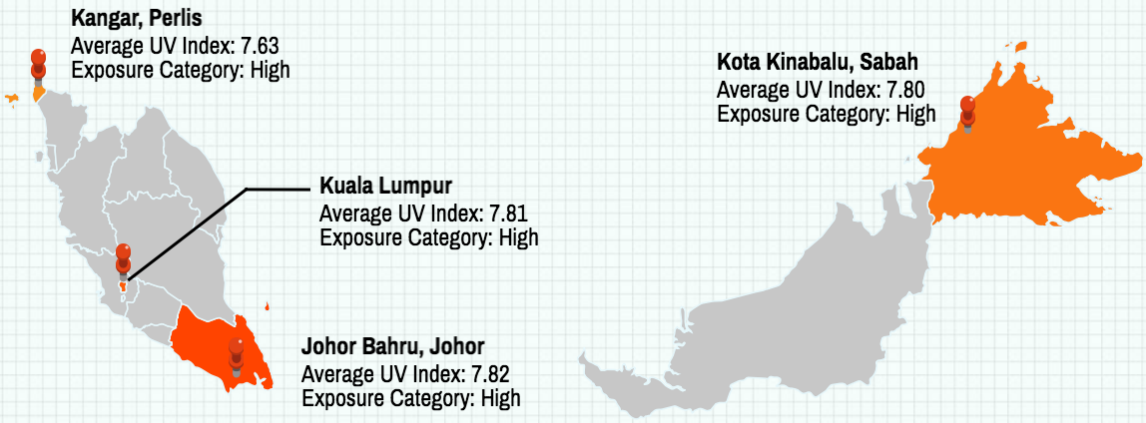
There is a weak **positive** correlation between the average humidity and average dew point.

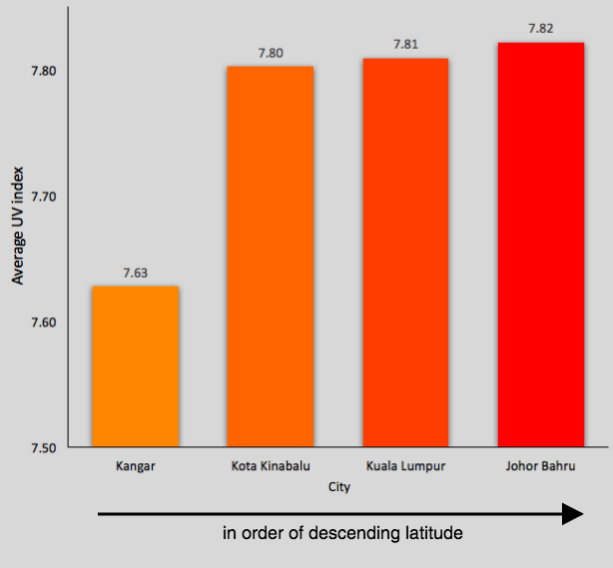
This can be attributed to the fact that other factors such as pressure affect the average dew point.

UV index of cities in Malaysia

January to April 2017

The UV index indicates the intensity of ultraviolet radiation from the Sun reaching the Earth's surface on a scale of 0 to 11. The higher the UV index, the more damage can be done to our skin from UV radiation.





An Observation - UV index and the Equator

As the latitude (the distance from the equator) decreases, the UV index increases.

This is because around the March equinox, the Sun is shining directly on the equator. Therefore, the closer the city is to the equator, the higher the intensity of UV radiation.