Data Analytics using Spreadsheets

Correlation Functions

1 Correlation Function

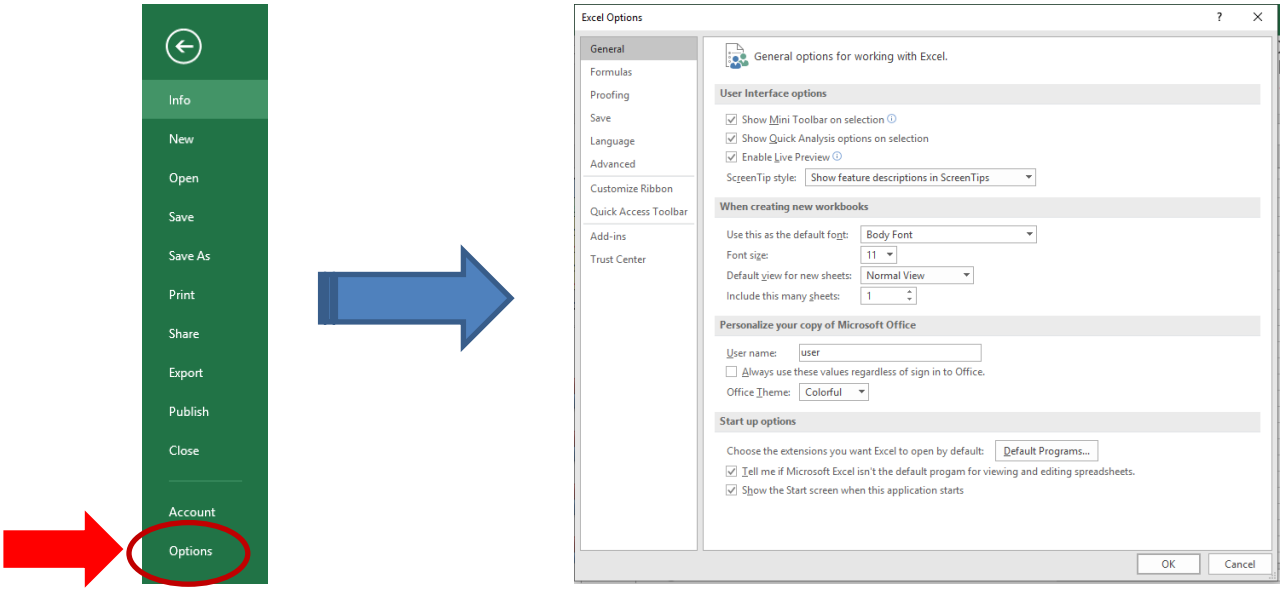
1.1 Data Analysis with Correlation

1. Open spreadsheet **DA_Excel_Correlation.xlsx**

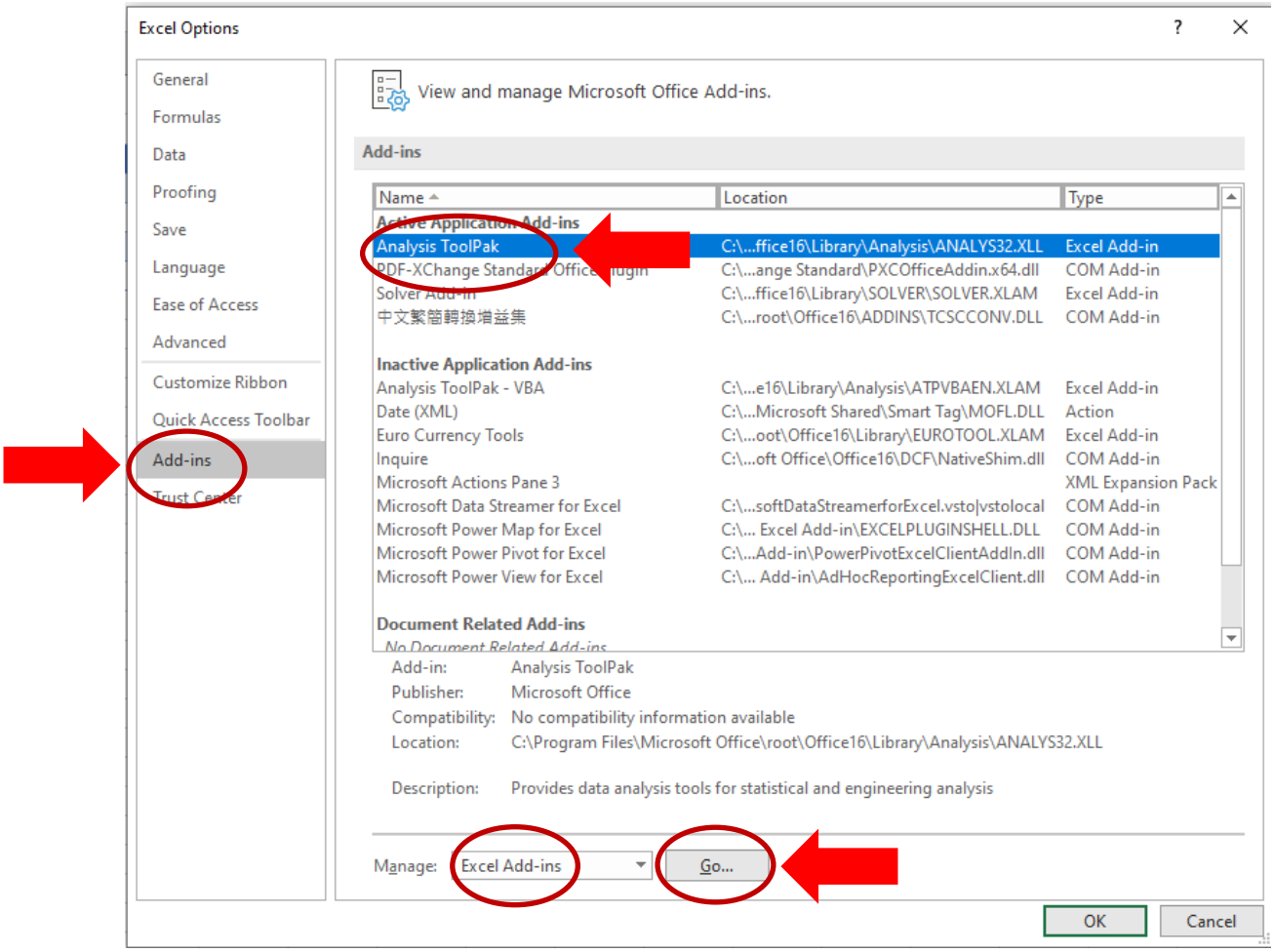
2024 Monthly Weather Summary

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean Air Temperature (°C)	16.1	17	20.8	23.6	28.3	28.6	29.1	28.6	28	25.3	22.9	19.2
Mean Relative Humidity (%)	78	70	76	78	77	80	81	85	78	69	78	76
Total Rainfall (mm)	62.2	5.5	22.7	28.1	57.5	458.8	341.1	615.1	383.3	104.3	73.4	11.9
Total Bright Sunshine (hours)	136.1	108.7	196.2	143.5	236.9	145.2	181.1	116.2	183.3	181.9	123.9	122.2

2. Click the **File** button in the top menu bar. In the drop-down list, open the **Options** dialog box.

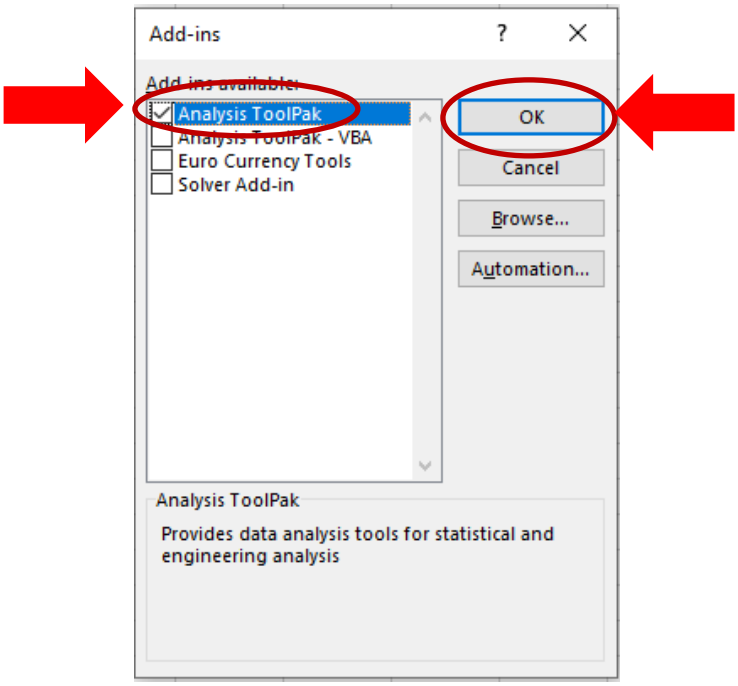


3. Open the **Add-ins** option. Select **Analysis ToolPak** *(it may be under Inactive Application Add-ins)*, and make sure that it is **'Excel Add-ins'** for the **Manage** command, then click Go.

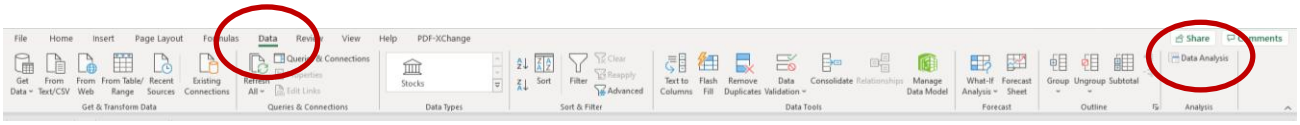


For MAC Users
Click **Tools** (on the top menu bar) -> **Excel Add-ins Tools** -> **Analysis ToolPak**

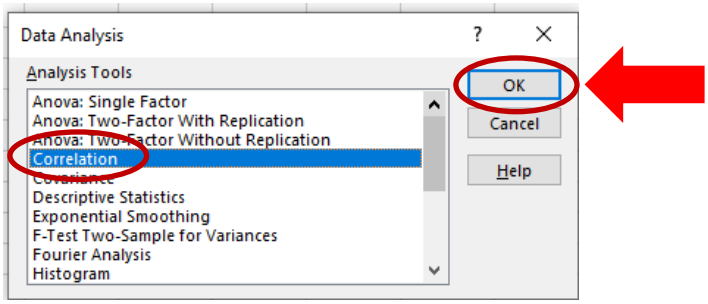
4. In the **Add-ins** dialog box, select **'Analysis ToolPak'** and click OK.



5. Notice that in the **Data** ribbon menu, the **Data Analysis** button is added in the **Analysis** group.

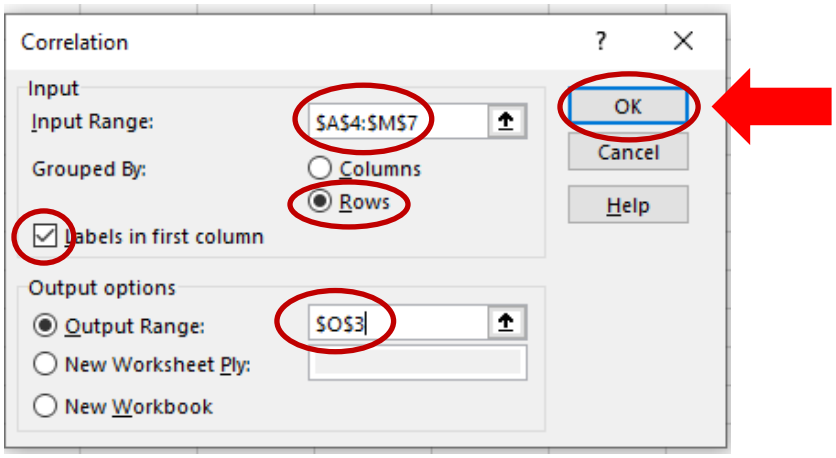


6. Click the **Data Analysis** button and open the Data Analysis window. Select **Correlation** and click OK.



7. In the Correlation window,
   - Set **input range** to A4:M7.
   - Select **Rows** in Grouped By.
   - Select **Labels in first column**.
   - Set **output range** to O3.

   Then click OK.



8. A correlation table is created at cell O3, with the **Correlation Coefficients** for each pair of data attributes. Which pair of data has the highest correlation and which pair has the lowest?

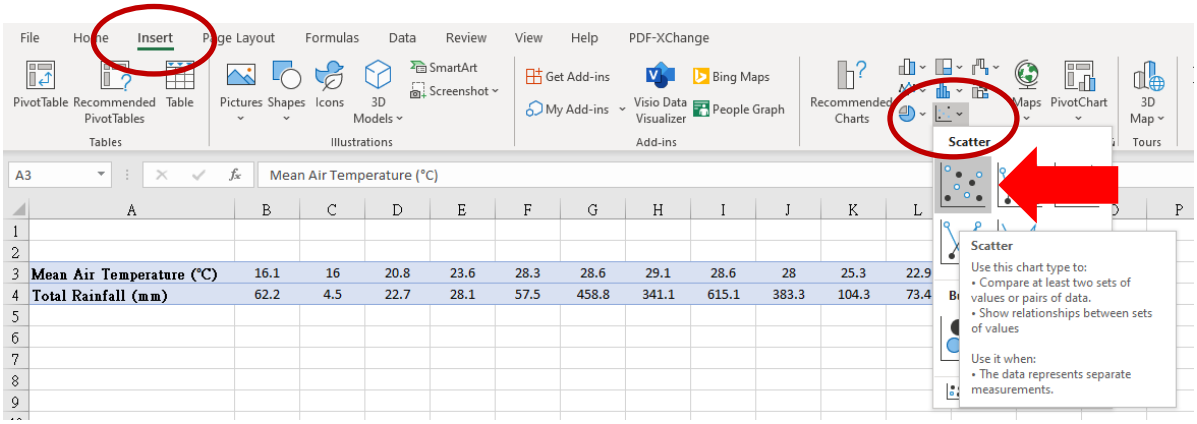
	Mean Air Temperature (°C)	Mean Relative Humidity (%)	Total Rainfall (mm)	Total Bright Sunshine (hours)
Mean Air Temperature (°C)	1			
Mean Relative Humidity (%)	0.542686895	1		
Total Rainfall (mm)	0.713850334	0.68560194	1	
Total Bright Sunshine (hours)	0.47094329	-0.061207987	-0.079919421	1

1.2 Scatter Chart with Trendline

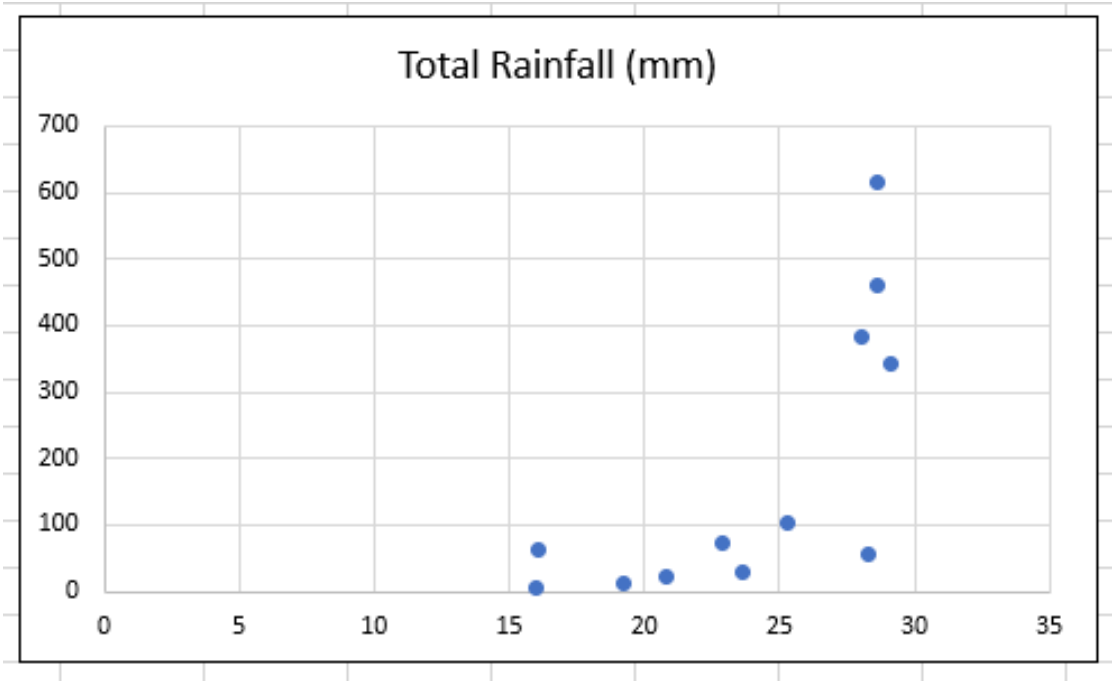
1. Copy the two rows – **Mean Air Temperature** and **Total Rainfall** – and paste them into a new worksheet.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	**Mean Air Temperature (°C)**	16.1	16	20.8	23.6	28.3	28.6	29.1	28.6	28	25.3	22.9	19.2
4	**Total Rainfall (mm)**	62.2	4.5	22.7	28.1	57.5	458.8	341.1	615.1	383.3	104.3	73.4	11.9
5													

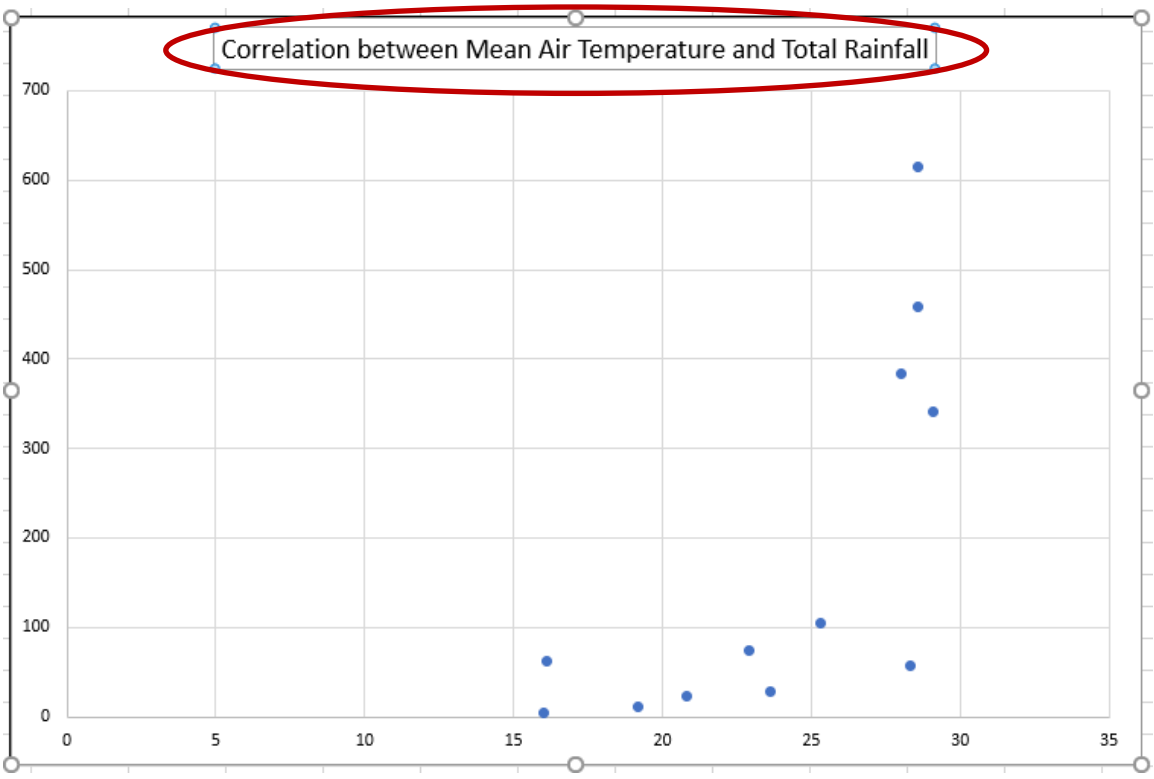
2. Select the two rows of data. Open the **Insert** ribbon menu, then select **Scatter** chart.



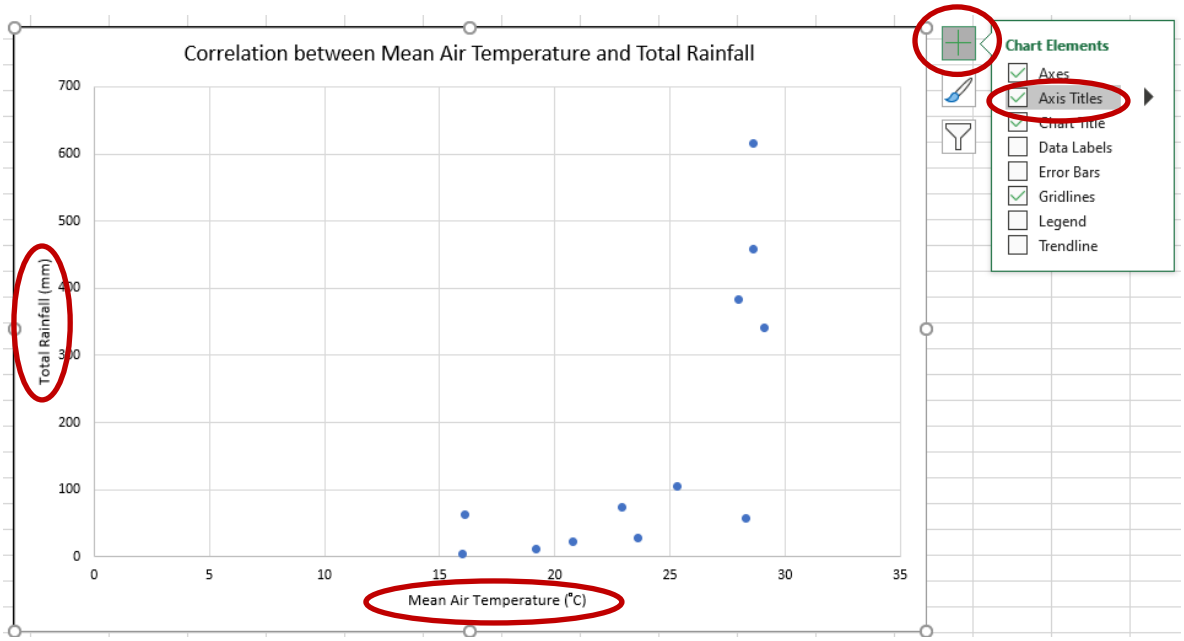
3. A **Scatter** chart is created as follows. However the header is incorrect, and there are no labels on the axis.



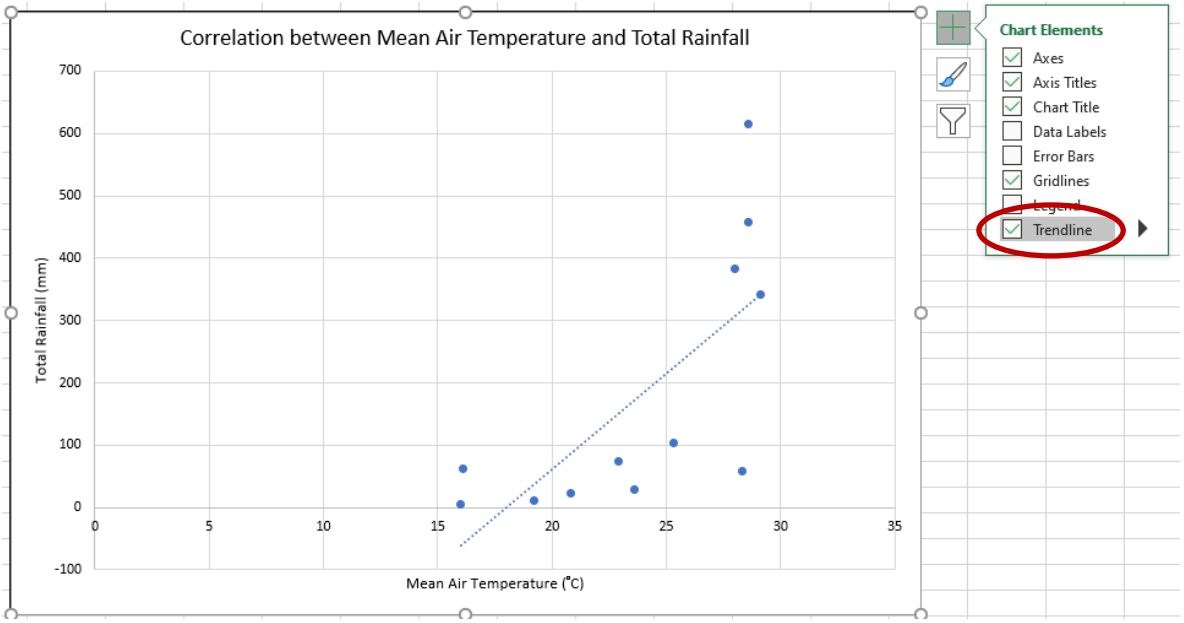
4. Click on the header, and change it as follows.



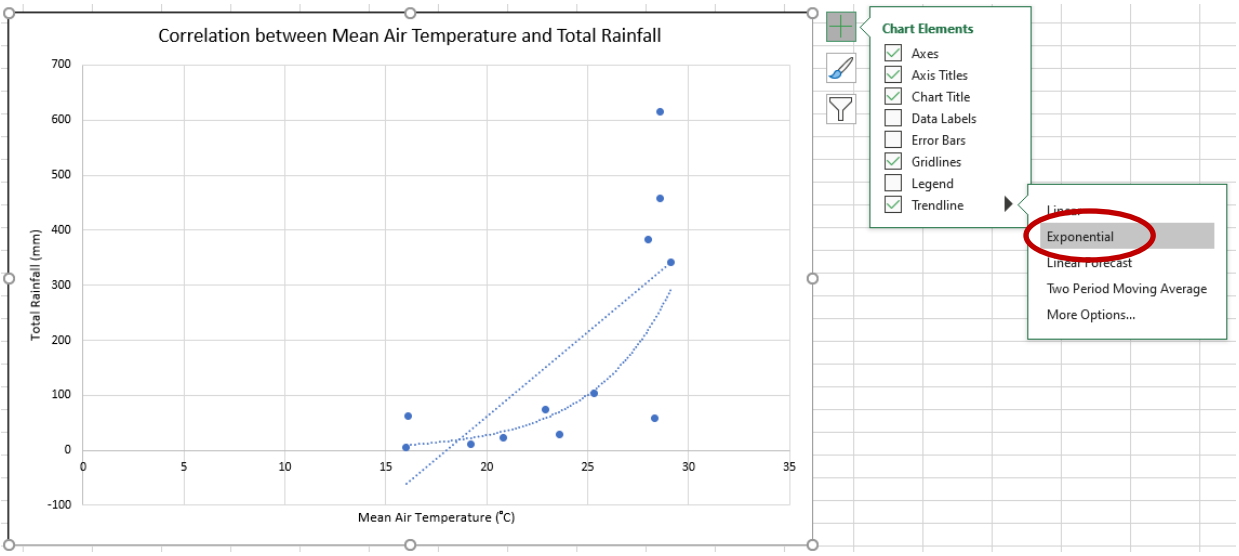
5. Click once on the chart, and select the **Chart Elements** icon. Then select **Axis Titles,** and input the two axis titles as follows.



6. Select **Trendline** under Chart Elements. Notice that a linear correlation trendline is fitted into the chart. This trendline matches the Correlation Coefficient calculated in step 8 of section 2.1 on p.8.



7. Open the dropdown menu of **Trendline**, and select **Exponential**. Another trendline is fitted into the chart. Which line is the more suitable one?



2 Lab Exercises (for submission)

1. Open the spreadsheet **DA_Excel_Correlation.xlsx**. Add a scatter chart with one linear trendline to show the correlation between Total Bright Sunshine and Mean Relative Humidity.

Submit the Excel file when finished.