



SDSC 2102

STATISTICAL METHODS AND DATA ANALYSIS

Lab 1, 2 – Data visualization and analysis



Aims



- ❑ A. Learn descriptive statistics
 - ❑ Calculate mean, median, mode;
 - ❑ Calculate maximum, minimum, range, sample variance, sample standard deviation.
- ❑ B. Learn data representation
 - ❑ Plot the histogram;
 - ❑ Plot the box-whisker.

Part A. Descriptive Statistics

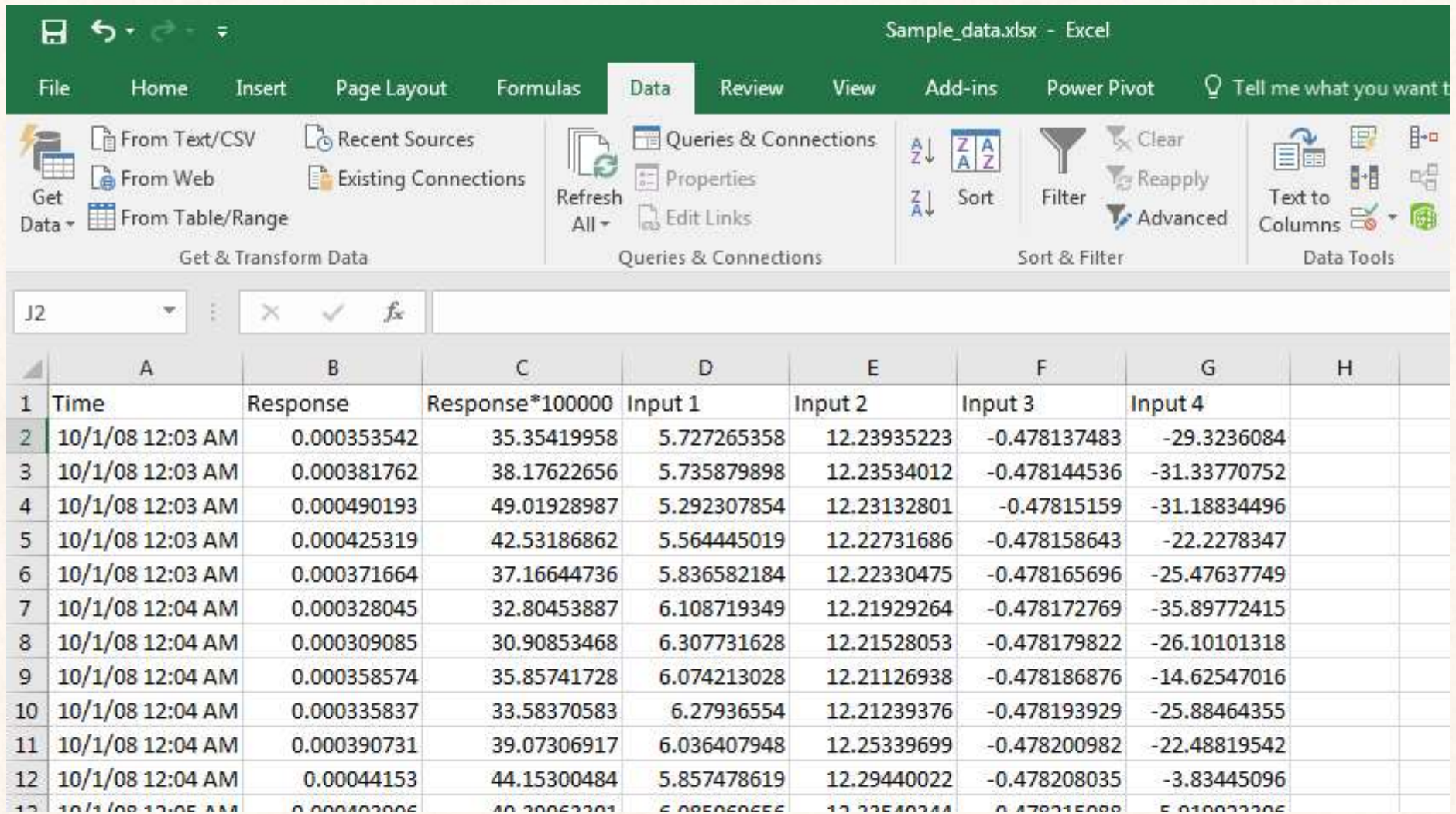
□ A.1 Measures of Central Tendency

- Mean: the sum of the data divided by the number of data points in the data set, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$;
- Median: the middle number when the data set is arranged in ascending (or descending) order.
- Mode: the most frequently occurring number in the data set.

- Example: 1, 2, 2, 3, 5, 7, 9
mean: $(1+2+2+3+5+7+9)/7 = 4.14285$
median: 3
mode: 2

Part A.1 Mean

- Start excel and open the file Sample_data.xlsx

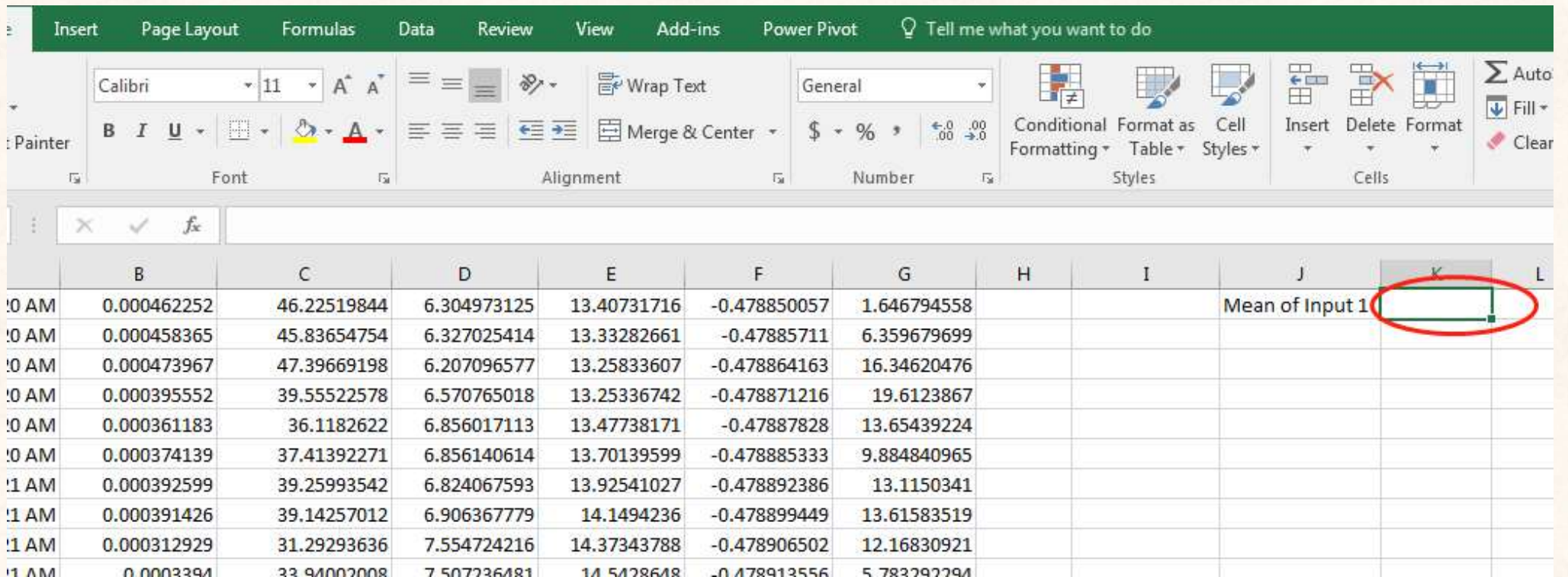


The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The ribbon includes options for 'Get & Transform Data', 'Queries & Connections', 'Sort & Filter', and 'Data Tools'. The data table below has columns A through H, with the first row as headers and subsequent rows containing numerical and timestamp data.

	A	B	C	D	E	F	G	H
1	Time	Response	Response*100000	Input 1	Input 2	Input 3	Input 4	
2	10/1/08 12:03 AM	0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084	
3	10/1/08 12:03 AM	0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752	
4	10/1/08 12:03 AM	0.000490193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496	
5	10/1/08 12:03 AM	0.000425319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347	
6	10/1/08 12:03 AM	0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749	
7	10/1/08 12:04 AM	0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415	
8	10/1/08 12:04 AM	0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318	
9	10/1/08 12:04 AM	0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016	
10	10/1/08 12:04 AM	0.000335837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355	
11	10/1/08 12:04 AM	0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542	
12	10/1/08 12:04 AM	0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096	
13	10/1/08 12:05 AM	0.000402006	40.20062201	5.085060656	12.23540244	-0.478215088	5.010022206	

Part A.1 Mean

- Step 1: place your cursor in the cell labeled Mean

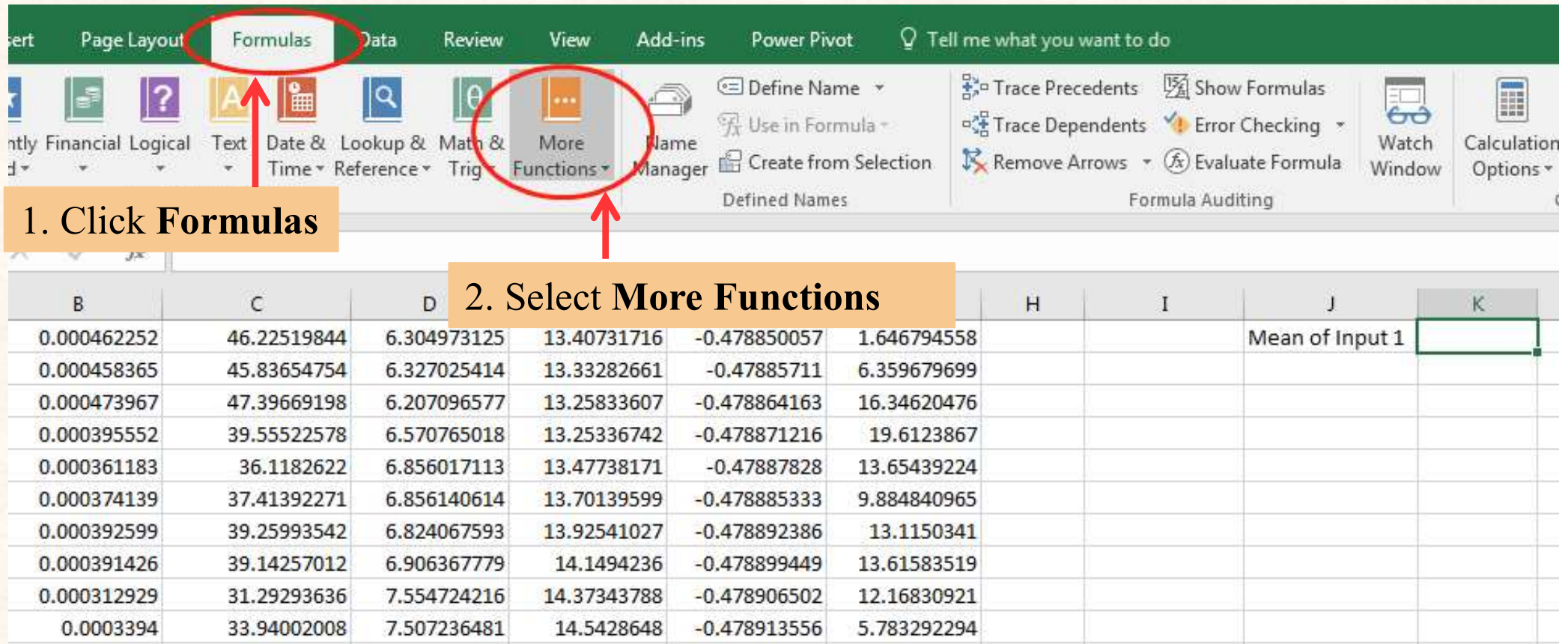


The screenshot shows the Microsoft Excel interface with the following data table:

	B	C	D	E	F	G	H	I	J	K	L
10 AM	0.000462252	46.22519844	6.304973125	13.40731716	-0.478850057	1.646794558			Mean of Input 1		
10 AM	0.000458365	45.83654754	6.327025414	13.33282661	-0.47885711	6.359679699					
10 AM	0.000473967	47.39669198	6.207096577	13.25833607	-0.478864163	16.34620476					
10 AM	0.000395552	39.55522578	6.570765018	13.25336742	-0.478871216	19.6123867					
10 AM	0.000361183	36.1182622	6.856017113	13.47738171	-0.47887828	13.65439224					
10 AM	0.000374139	37.41392271	6.856140614	13.70139599	-0.478885333	9.884840965					
11 AM	0.000392599	39.25993542	6.824067593	13.92541027	-0.478892386	13.1150341					
11 AM	0.000391426	39.14257012	6.906367779	14.1494236	-0.478899449	13.61583519					
11 AM	0.000312929	31.29293636	7.554724216	14.37343788	-0.478906502	12.16830921					
11 AM	0.0003394	33.94007008	7.507736481	14.5478648	-0.478913556	5.782797794					

Part A.1 Mean

- Step 2: select More Functions ... from the Formulas menu
(or click on the paste function - fx - button on the menu)



1. Click **Formulas**

2. Select **More Functions**

B	C	D	H	I	J	K
0.000462252	46.22519844	6.304973125	13.40731716	-0.478850057	1.646794558	Mean of Input 1
0.000458365	45.83654754	6.327025414	13.33282661	-0.47885711	6.359679699	
0.000473967	47.39669198	6.207096577	13.25833607	-0.478864163	16.34620476	
0.000395552	39.55522578	6.570765018	13.25336742	-0.478871216	19.6123867	
0.000361183	36.1182622	6.856017113	13.47738171	-0.47887828	13.65439224	
0.000374139	37.41392271	6.856140614	13.70139599	-0.478885333	9.884840965	
0.000392599	39.25993542	6.824067593	13.92541027	-0.478892386	13.1150341	
0.000391426	39.14257012	6.906367779	14.1494236	-0.478899449	13.61583519	
0.000312929	31.29293636	7.554724216	14.37343788	-0.478906502	12.16830921	
0.0003394	33.94002008	7.507236481	14.5428648	-0.478913556	5.783292294	

Part A.1 Mean

- Step 3: select Function **AVERAGE** under the statistical menu

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The 'Function Library' group is visible, containing 'More Functions' and 'Name Manager'. The 'More Functions' dropdown menu is open, showing a list of function categories: 'Statistical', 'Engineering', 'Cube', 'Information', 'Compatibility', and 'Web'. The 'Statistical' category is highlighted with a red circle and a red arrow pointing to it, with a callout box saying '1. Click Statistical'. The 'Statistical' submenu is also open, showing a list of statistical functions. The 'AVERAGE' function is highlighted with a red arrow pointing to it, with a callout box saying '2. Select AVERAGE'. The background shows a spreadsheet with columns A through D, containing time, response, and calculated values.

	A	B	C	D
1	Time	Response	Response*100000	Input 1
2	10/1/08 12:03 AM	0.000353	35.35	5358
3	10/1/08 12:03 AM	0.000381	38.1	79898
4	10/1/08 12:03 AM	0.000490193	49.01928987	5.292307854
5	10/1/08 12:03 AM	0.000425319	42.53186862	5.564445019
6	10/1/08 12:03 AM	0.000371664	37.16644736	5.836582184
7	10/1/08 12:04 AM	0.000328045	32.80453887	6.108719349
8	10/1/08 12:04 AM	0.000309085	30.90853468	6.307731628
9	10/1/08 12:04 AM	0.000358574	35.85741728	6.074213028
10	10/1/08 12:04 AM	0.000335837	33.58370583	6.27936554

Part A.1 Mean

Function Arguments

AVERAGE

Number1: D2:D60001

Number2: number

Formula result = 7.468938514

Help on this function

OK Cancel

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.1 Mean

<i>f_x</i>	=AVERAGE(D2:D60001)									
	C	D	E	F	G	H	I	J	K	
	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.4689385	
3542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084					
1762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752					
0193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496					
5319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347					
1664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749					
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415					
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318					
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016					
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355					
0731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542					
4153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096					
3906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306					
4855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221					
6776	36.67767011	6.235821022	12.41741085	-0.478230105	1.225527762					

Get the final result of Mean

Get the final result of Mean

Part A.1 Median

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The 'Function Library' group is open, showing the 'Statistical' category. The 'MEDIAN' function is highlighted with a red circle and a red arrow. A text box with the text 'Select MEDIAN' is overlaid on the image.

The 'Formulas' tab ribbon includes the following groups and options:

- Function Library:** Statistical, Engineering, Cube, Information, Compatibility, Web.
- More Functions:** Define Name, Use in Formula, Create from Selection.
- Name Manager:** Define Name, Use in Formula, Create from Selection.
- Formula Auditing:** Trace Precedents, Trace Dependents, Remove Arrows, Evaluate Formula, Watch Window, Calculate Options.

The 'Statistical' function list includes:

- LOGEST
- LOGNORM.DIST
- LOGNORM.INV
- MAX
- MAXA
- MAXIFS
- MEDIAN**
- MIN
- MINA
- MINIFS
- MODE.MULT
- MODE.SNGL
- NEGBINOM.DIST
- NORM.DIST
- NORM.INV
- NORM.S.DIST
- NORM.S.INV

The worksheet data is as follows:

A	B	C	D
Time	Response	Response*100000	Input 1
10/1/08 12:03 AM	0.000353542	35.35419958	5.727265358
10/1/08 12:03 AM	0.000381762	38.17622656	5.735879898
10/1/08 12:03 AM	0.000490193	49.01928987	5.292307854
10/1/08 12:03 AM	0.000425319	42.53186862	5.564445019
10/1/08 12:03 AM	0.000371664	37.16644736	5.836582184
10/1/08 12:04 AM	0.000328045	32.80453887	6.108719349
10/1/08 12:04 AM	0.000309085	30.90853468	6.307731628
10/1/08 12:04 AM	0.000358574	35.85741728	6.074213028
10/1/08 12:04 AM	0.000335837	33.58370583	6.108719349
10/1/08 12:04 AM	0.000390731	39.07306911	6.108719349
10/1/08 12:04 AM	0.00044153	44.15300484	6.108719349
10/1/08 12:05 AM	0.000403906	40.39063301	6.085069656
10/1/08 12:05 AM	0.000384855	38.48553584	6.155450344
10/1/08 12:05 AM	0.000366776	36.67757911	6.225831032
10/1/08 12:05 AM	0.00034419	34.41899034	6.342935562
10/1/08 12:05 AM	0.000304045	30.40450058	6.747737408

The 'Formula Auditing' group on the right shows the following data:

I	J	K
	Mean of Input 1	7.468939
	Median of Input 1	

Part A.1 Median

The image shows an Excel spreadsheet with the following data:

B	C	D	E	F	G	H	I	J	K	L
Response	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468939	
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	=MEDIAN(D1:D60001)	
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752					
0.000490193										
0.000425319										
0.000371664										
0.000328045										
0.000309085										
0.000358574										
0.000335837										
0.000390731										
0.00044153										
0.000403906										
0.000384855										
0.000366776										
0.00034419										
0.000294946										
0.000302216										
0.000328515										
0.000362973										
0.000412096	41.20962422	6.027213097	12.66343021	-0.478271544	-4					

The **MEDIAN** function dialog box is open, showing the range **D1:D60001** entered in the **Number1** field. The **OK** button is circled in red.

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.1 Median

✕ ✓ <i>fx</i> =MEDIAN(D1:D60001)									
B	C	D	E	F	G	H	I	J	K
Response	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468939
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.3665221
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752				
0.000490193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
0.000425319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
0.000335837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
0.000403906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
0.000384855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
0.000366776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

Get the final result of Median

Part A.1 Mode

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The 'Statistical' function category is expanded, and the 'MODE.SNGL' function is highlighted. An orange callout box with a red arrow points to 'MODE.SNGL' with the text 'Select MODE.SNGL'.

Function Library

B	C	D
Response	Response*100000	Input 1
0.000353542	35.35419958	5.727265358
0.000381762	38.17622656	5.735879898
0.000490193	49.01928987	5.292307854
0.000425319	42.53186862	5.564445019
0.000371664	37.16644736	5.836582184
0.000328045	32.80453887	6.108719349
0.000309085	30.90853468	6.307731628
0.000358574	35.85741728	6.074213028
0.000335837	33.58370583	6.27936554
0.000390731	39.07306917	6.036407948
0.00044153	44.15300484	5.857478619
0.000403906	40.39063301	6.085069656
0.000384855	38.48553584	6.155450344
0.000366776	36.67757911	6.225831032
0.000344118	34.41999024	6.242925562

Statistical Functions

- KURT
- LARGE
- LINEST
- LOGEST
- LOGNORM.DIST
- LOGNORM.INV
- MAX
- MAXA
- MAXIFS
- MEDIAN
- MIN
- MINA
- MINIFS
- MODE.MULT
- MODE.SNGL**
- NEGBINOM.DIST

Formula Auditing

I	J	K
	Mean of Input 1	7.4689385
	Median of Input 1	7.3665223
	Mode of Input 1	

Part A.1 Mode

The image shows an Excel spreadsheet with the following data:

C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	D2:D6000)
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				

The function arguments dialog box for MODE.SNGL is open, showing the following arguments:

- Number1: D2:D60001
- Number2: (empty)

The dialog box also displays the formula result as 8.994737625 and the OK button is circled in red.

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.1 Mode

=MODE.SNGL(D2:D60001)								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496				
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

Get the final result of Mode

Part A. Descriptive Statistics (cont'd)

□ A.2 Measures of Variability

- Maximum and Minimum: the largest number and the smallest number in the data set
- Range: maximum minus minimum
- Sample Variance: the sum of each data points from the mean and divided by the number of data points minus one

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Sample Standard Deviation: the positive square root of the sample variance

$$\sigma = \sqrt{s^2}$$

Part A.2 Maximum & Minimum

The screenshot shows the Microsoft Excel interface. The 'Formulas' tab is active, and the 'Function Library' group is expanded. The 'Statistical' category is selected, and the 'MAX' function is highlighted with a red circle. A red arrow points from a yellow callout box labeled 'Select MAX' to the 'MAX' function. The background shows a spreadsheet with columns C, D, and E, and rows 42 through 76. The 'MAX' function is being applied to the data in column C.

	C	D	E
	Response*100000	Input 1	Input 2
42	35.35419958	5.727265358	12.23935223
62	38.17622656	5.735879898	12.23534012
93	49.01928987	5.292307854	12.23132801
19	42.53186862	5.564445019	12.22731686
64	37.16644736	5.836582184	12.22330475
45	32.80453887	6.108719349	12.21929264
85	30.90853468	6.307731628	12.2152
74	35.85741728	6.074213028	12.2112
37	33.58370583	6.27936554	12.21239376
31	39.07306917	6.036407948	12.25339699
53	44.15300484	5.857478619	12.29440022
06	40.39063301	6.085069656	12.33540344
55	38.48553584	6.155450344	12.37640667
76	36.67757911	6.225831032	12.41741085

Part A.2 Maximum & Minimum

The screenshot shows an Excel spreadsheet with columns C through K. Column C contains the formula `=MAX(D2:D60001)`. The spreadsheet data includes columns for 'Input 1' through 'Input 4', 'Mean of Input 1', 'Median of Input 1', and 'Standard Deviation of Input 1'. A 'Function Arguments' dialog box for the MAX function is open, showing 'Number1' as 'D2:D60001' and 'Number2' as empty. A red circle highlights the 'Number1' field, and a red arrow points to it. The dialog box also shows the formula result as 21.00825119 and an 'OK' button circled in red.

Function Arguments

MAX

Number1: D2:D60001 = {5.72726535797119;5.73587989807128}

Number2: = number

Returns the largest value in a set of values

Number1: number1,number2,... are 1 to 255 numbers, empty cells, logical values, or text numbers for which you want the maximum.

Formula result = 21.00825119

[Help on this function](#)

OK Cancel

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.2 Maximum & Minimum

=MAX(D2:D60001)								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347				
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				

Get the final result of Maximum

Part A.2 Maximum & Minimum

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The 'More Functions' dropdown menu is open, displaying a list of statistical functions. The 'MIN' function is highlighted with a red circle, and a red arrow points to it with the text 'Select MIN'.

Function Library

C	D	E	F	G
Response*100000	Input 1	Input 2	Input 3	Input 4
35.35419958	5.727265358	12.23935223	-0.47815159	-31.11
38.17622656	5.735879898	12.23534012	-0.478158643	-22.7
49.01928987	5.292307854	12.23132801	-0.478165696	-25.4
42.53186862	5.564445019	12.22731686	-0.478172769	-35.8
37.16644736	5.836582184	12.22330475	-0.478179822	-26.1
32.80453887	6.108719349	12.21929264	-0.478200982	-22.4
30.90853468	6.307731628	12.21528053	-0.478208035	-3.8
35.85741728	6.074213028	12.21126938	-0.478215088	5.91
33.58370583	6.27936554	12.21239376	-0.478222142	16.8
39.07306917	6.036407948	12.25339699		
44.15300484	5.857478619	12.29440022		
40.39063301	6.085069656	12.33540344		
38.48553584	6.155450344	12.37640667		

Formulas

More Functions

- Statistical
- Engineering
- Cube
- Information
- Compatibility
- Web

MIN

Select MIN

Part A.2 Maximum & Minimum

The screenshot shows an Excel spreadsheet with columns C through K. Column C contains the formula `=MIN(D2:D60001)`. The spreadsheet data includes input values and calculated statistics like Mean and Median. Overlaid on the spreadsheet is the 'Function Arguments' dialog box for the MIN function. The 'Number1' field is set to 'D2:D60001' and is circled in red. A red arrow points from an orange text box to this field. The 'Number2' field is empty. The dialog box also shows a list of arguments, the formula result, and 'OK' and 'Cancel' buttons at the bottom, with the 'OK' button also circled in red.

Function Arguments

MIN

Number1: D2:D60001 {5.72726535797119;5.73587989807128}

Number2: = number

Returns the smallest number in the array of numbers.

Formula result = 0.400000483

[Help on this function](#)

OK Cancel

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.2 Maximum & Minimum

✕ ✓ <i>f_x</i> =MIN(D2:D60001)								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749				
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				

Get the final result of Minimum

Part A.2 Range

Insert Page Layout Formulas Data Review View Power Pivot Tell me what you want to do								
Function Library								
Defined Names Formula Auditing								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-11.51452984				
39.07306917	6.036407948	12.25339699	-0.478201082	-8.40349918				
44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				
38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221				
36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763				

place your cursor in the cell labeled Range

Part A.2 Range

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The formula bar displays the formula `=K4-K5`, which is circled in red. A red arrow points from the formula bar to cell K4 in the worksheet. The worksheet contains data for 'Input 1' and 'Input 2' across multiple rows. The 'Range of Input 1' is calculated in cell K4 using the formula `=K4-K5`.

Input “= maximum – minimum” and then press Enter

	C	D	E	F	G	H	I	J	K
	Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
2	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
2	38.17622656	5.725870888	12.23524012	-0.478144536	-21.32770752			Mode of Input 1	0.400016367
3								Maximum of Input 1	21.00825119
9	42.55180802	5.504443012	12.22751080	-0.478138043	-22.2278347			Minimum of Input 1	0.400000483
4	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	=K4-K5
5	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415				
5	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
4	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
7	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
1	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				
3	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096				
6	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306				

Part A.2 Range

InsertPage LayoutFormulasDataReviewViewPower PivotTell me what you want to do

Recently Used

Financial

Logical

Text

Date & Time

Lookup & Reference

Math & Trig

More Functions

Function Library

Name Manager

Define Name

Use in Formula

Create from Selection

Defined Names

Trace Precedents

Trace Dependents

Remove Arrows

Show Formulas

Error Checker

Evaluate Formula

Formula Auditing

Get the final result of Range

Part A.2 Sample Variance

The screenshot displays the Microsoft Excel interface with the 'Formulas' tab selected. The 'Function Library' pane on the left shows various function categories. The 'Statistical' category is expanded, and the 'VAR.S' function is highlighted with a red circle. A red arrow points from an orange box labeled 'Select VAR.S' to the 'VAR.S' function. The background spreadsheet shows columns C, D, and E with numerical data.

C	D	E
Response*100000	Input 1	Input 2
35.35419958	5.727265358	12.23935223
38.17622656	5.735879898	12.23534012
49.01928987	5.292307854	12.23132801
42.53186862	5.564445019	12.22731686
37.16644736	5.836582184	12.22330475
32.80453887	6.108719349	12.21929264
30.90853468	6.307731628	12.21528053
35.85741728	6.074213028	12.21126938
33.58370583	6.27936554	12.21239376
39.07306917	6.036407948	12.25339699
44.15300484	5.857478619	12.29440022
40.39063301	6.085069656	12.33540344
38.48553584	6.155450344	12.37640667
36.67757911	6.225831032	12.41741085
34.41899034	6.342935562	12.45841408

Part A.2 Sample Variance

Function Arguments

VAR.S

Number1: D2:D60001

Number2:

Estimates variance based on a sample of a population.

Formula result = 8.054780762

[Help on this function](#)

OK Cancel

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.2 Sample Variance

=VAR.S(D2:D60001)								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	20.60825071
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			Variance of Input 1	8.054780762
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318				
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				

Get the final result of sample variance

Part A.2 Standard Deviation

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The 'Function Library' pane on the left shows the 'Statistical' category expanded. The 'STDEV.S' function is highlighted with a red circle and a red arrow. A text box with the text 'Select STDEV.S' is overlaid on the image.

Formulas Tab - Function Library

- Recently Used
- Financial
- Logical
- Text
- Date & Time
- Lookup & Reference
- Math & Trig
- More Functions
- Name Manager
- Define Name
- Use in Formula
- Create from Selection
- Trace Precedents
- Trace Dependents
- Remove Arrows
- Show Formulas
- Error Checking
- Evaluate Formula

Statistical Functions

- RANK.EQ
- RSQ
- SKEW
- SKEW.P
- SLOPE
- SMALL
- STANDARDIZE
- STDEV.P
- STDEV.S**
- STDEVA
- STDEVPA
- STEYX
- T.DIST
- T.DIST.2T
- T.DIST.RT
- T.INV

Worksheet Data

	C	D	E	F	G
1	Response*100000	Input 1	Input 2	Input 3	Input 4
2	35.35419958	5.727265358	12.23935223	-0.47815159	-31.18
3	38.17622656	5.735879898	12.23534012	-0.478158643	-22.5
4	49.01928987	5.292307854	12.23132801	-0.478165696	-25.4
5	42.53186862	5.564445019	12.22731686	-0.478172769	-35.8
6	37.16644736	5.836582184	12.22330475	-0.478179822	-26.1
7	32.80453887	6.108719349	12.21929264	-0.478186876	-14.6
8	30.90853468	6.307731628	12.21528053	1	5.8
9	35.85741728	6.074213028	12.21126938	1	2.4
10	33.58370583	6.27936554	12.20725823	-0.478208035	-3.8
11	39.07306917	6.036407948	12.20324708	-0.478215088	5.91
12	44.15300484	5.857478619	12.19923593	-0.478222142	16.8
13	40.39063301	6.085069656	12.19522478	-0.478229195	1.22
14	38.48553584	6.155450344	12.19121363		
15	36.67757911	6.225831032	12.18720248		

Formula Auditing

	K
	7.468938514
1	7.366522312
	0.400016367
out 1	21.00825119
out 1	0.400000483
	20.60825071
ut 1	8.054780762
ion of Input 1	

Part A.2 Standard Deviation

The screenshot shows an Excel spreadsheet with columns C through K. Column C contains values like 35.35419958, 38.17622656, etc. Columns D through J contain data for 'Input 1' through 'Input 4'. Column K contains statistical results like 'Mean of Input 1', 'Median of Input 1', etc. A formula bar at the top shows '=STDEV.S(D2:D60001)'. A 'Function Arguments' dialog box for the STDEV.S function is open, showing 'Number1' as 'D2:D60001' and 'Number2' as empty. A red arrow points to the 'Number1' field. The 'OK' button at the bottom right of the dialog box is circled in red. An orange text box with a red arrow pointing to the 'Number1' field contains the following text:

A window as show here will pop out.

1. Input the selected column in the first cell
2. Then click ok

Part A.2 Standard Deviation

=STDEV.S(D2:D60001)								
C	D	E	F	G	H	I	J	K
Response*100000	Input 1	Input 2	Input 3	Input 4			Mean of Input 1	7.468938514
35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084			Median of Input 1	7.366522312
38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752			Mode of Input 1	0.400016367
49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			Maximum of Input 1	21.00825119
42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			Minimum of Input 1	0.400000483
37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			Range of Input 1	20.60825071
32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			Variance of Input 1	8.054780762
30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Standard deviation of Input 1	2.838094565
35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016				
33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355				
39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542				

Get the final result of standard deviation

Part B Data Representation

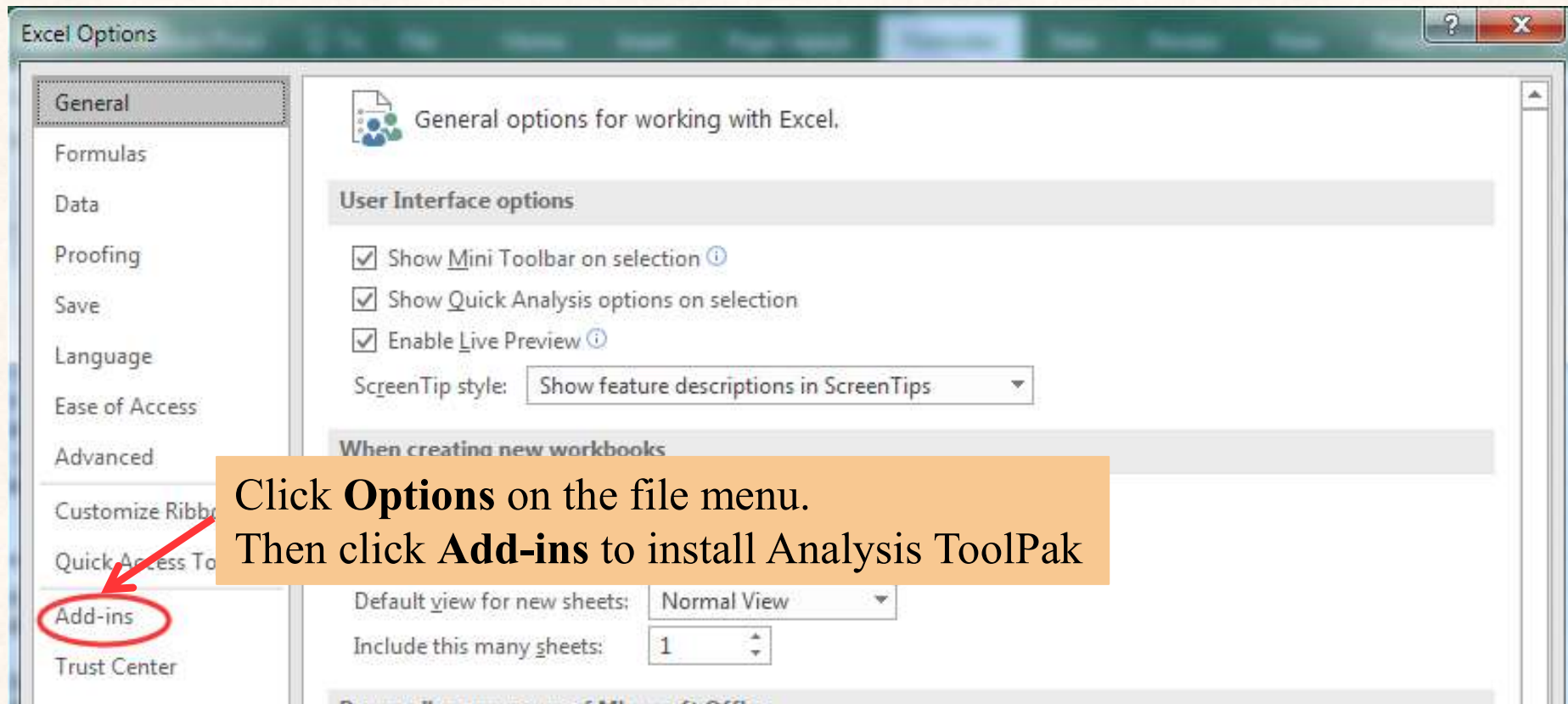
■ B.1 Histograms (Frequency Histograms)

A histogram is one method of graphically representing a set of data. Graphs help to provide a sense of shape to the data points, which may provide some insight to the analysis about how the data is distributed.

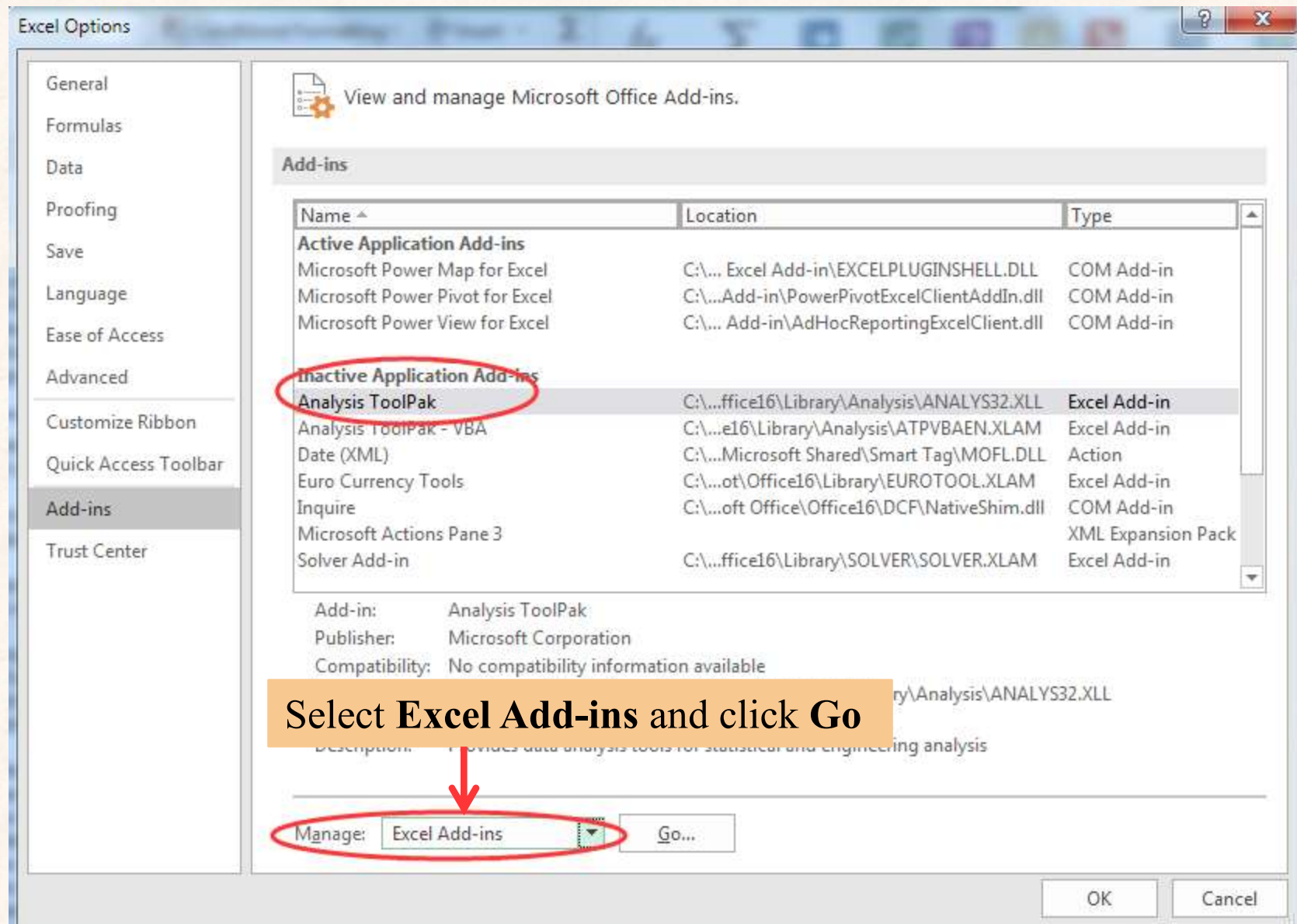
- Create a Histogram using Data Analysis ToolPak
- Create a Histogram using Frequency function

Part B.1 Create a Histogram using Data Analysis ToolPak

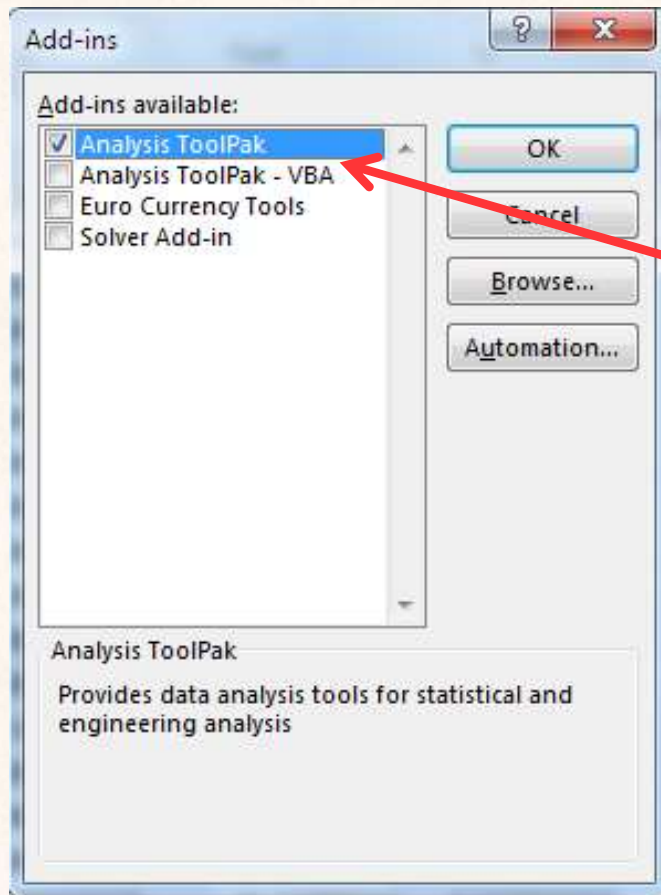
- Step 1: make sure the Analysis ToolPak Add-in is ready to use before you use the Histogram tool



Part B.1 Create a Histogram using Data Analysis ToolPak

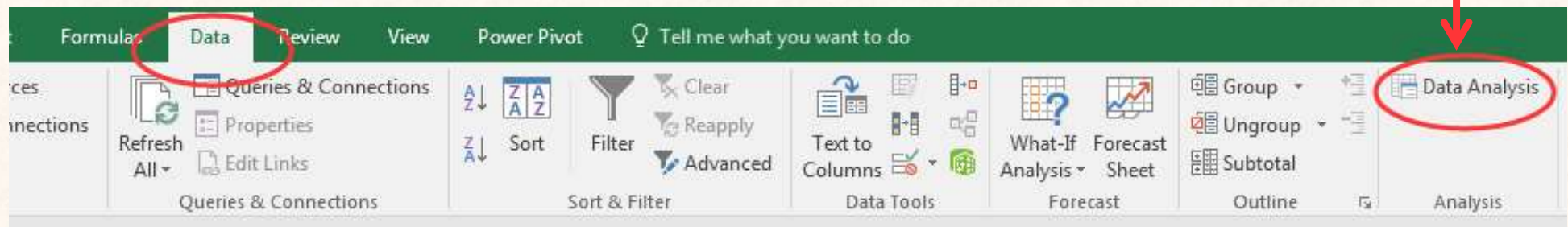


Part B.1 Create a Histogram using Data Analysis ToolPak



1. In the Add-ins dialog box, select **Analysis ToolPak** and click Ok.

2. Click **Data** in the panel and check whether Analysis ToolPak has been installed



Part B.1 Create a Histogram using Data Analysis ToolPak

- Step 2: create a histogram using Input 1. Firstly, we need to create the data intervals in which we want to find the data frequency. There are called bins.

B	C	D	E	F	G	H	I
Response	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins
0.000353542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6
0.000490182	49.01822887	5.733307854	12.23133801	-0.47815150	-31.18834496		9
0.00042531	42.531				278347		12
0.00037160	37.160				637749		15
0.00032804	32.804				772415		18
0.00030908	30.908				101318		
0.00035857	35.857				547016		
0.00033583	33.583				464355		
0.000390731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		
0.00044153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		
0.000403906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		
0.000384855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221		
0.000366776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763		
0.00034419	34.41899034	6.342935562	12.45841408	-0.478236258	-3.713405132		
0.000294946	29.49460058	6.747737408	12.4994173	-0.478243311	-1.201598406		
0.000302216	30.22158733	6.595600128	12.54042053	-0.478250374	-0.233026966		
0.000328515	32.85147654	6.443462849	12.58142376	-0.478257428	-8.387624741		
0.000362973	36.29734264	6.260338306	12.62242699	-0.478264491	-4		
0.000412096	41.20962422	6.027213097	12.66343021	-0.478271544	-4		

With the data Input1, you need to specify these bins separately in an additional column as shown on the right. The intervals can be determined based on the range of Input 1.

Part B.1 Create a Histogram using Data Analysis ToolPak

- Step 3: select Histogram in the Data dialog box to create a histogram based on the data, Input 1 and created bins.

1. Click **Data** in the panel

2. Click **Data Analysis**

3. Select **Histogram**, then click **OK**

	C	D	E	F	G	H	I	J
	onse*100000	Input 1	Input 2	Input 3	Input 4		Bins	
35.	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3	
38.					770752		6	
49.					834496		9	
42.					278347		12	
37.					637749		15	
32.					772415		18	
30.					101318			
35.								
33.								
39.					819542			
44.					445096			

Part B.1 Create a Histogram using Data Analysis ToolPak

The screenshot shows the 'Histogram' dialog box in Excel. The background is a spreadsheet with columns C through I. Column C is labeled 'Response*100000', column D is 'Input 1', column E is 'Input 2', column F is 'Input 3', column G is 'Input 4', and column I is 'Bins'. The dialog box has the following fields and options:

- Input Range:** \$D\$2:\$D\$60001 (circled in red)
- Bin Range:** \$I\$2:\$I\$6 (circled in red)
- Output options:**
 - ☒ **Output Range:** \$K\$1 (circled in red)
 - ☐ New Worksheet Ply:
 - ☐ New Workbook
 - ☐ Pareto (sorted histogram)
 - ☐ Cumulative Percentage
 - ☒ **Chart Output** (circled in red)

Red arrows point from the numbered instructions to the corresponding fields in the dialog box.

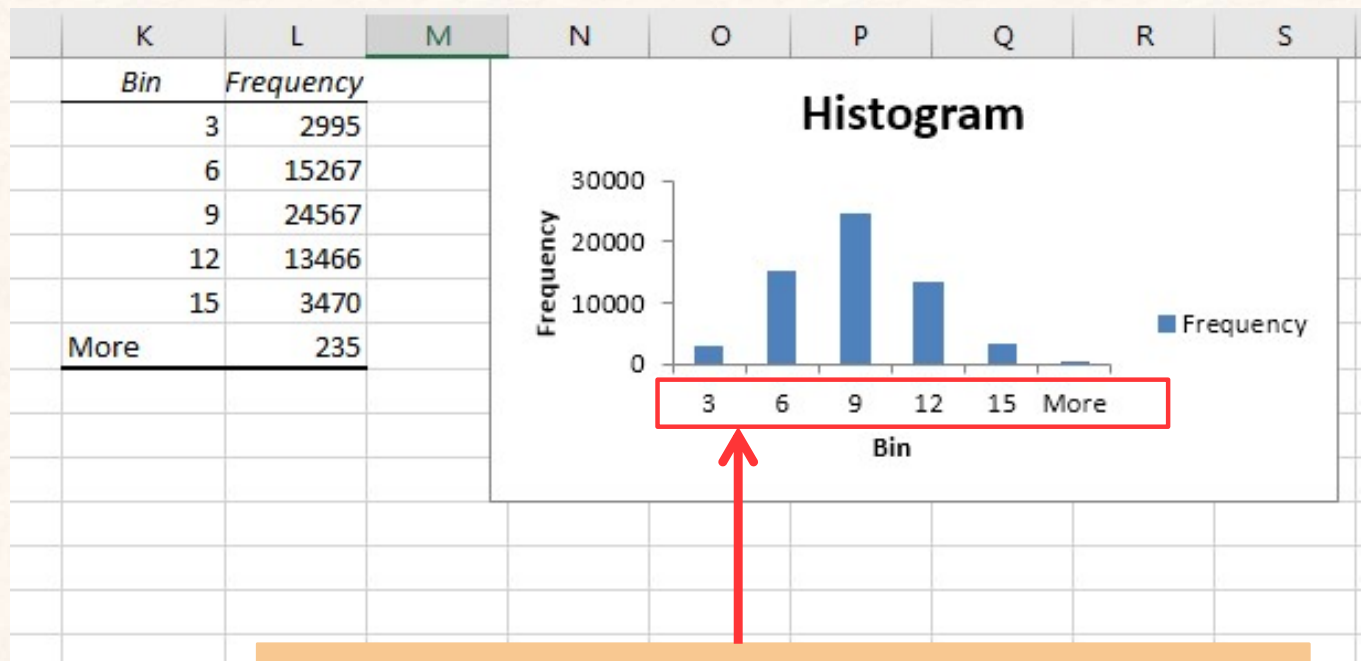
1. Select the **Input Range** (cell D2:D60001)

2. Select the **Bin Range** (cell I2:I6)

3. Specify the **Output Range** if you want to get the Histogram in the same worksheet. Otherwise, choose New Worksheet/Workbook option to get it in a separate worksheet/workbook

4. Select **Chart Output**. Then click OK

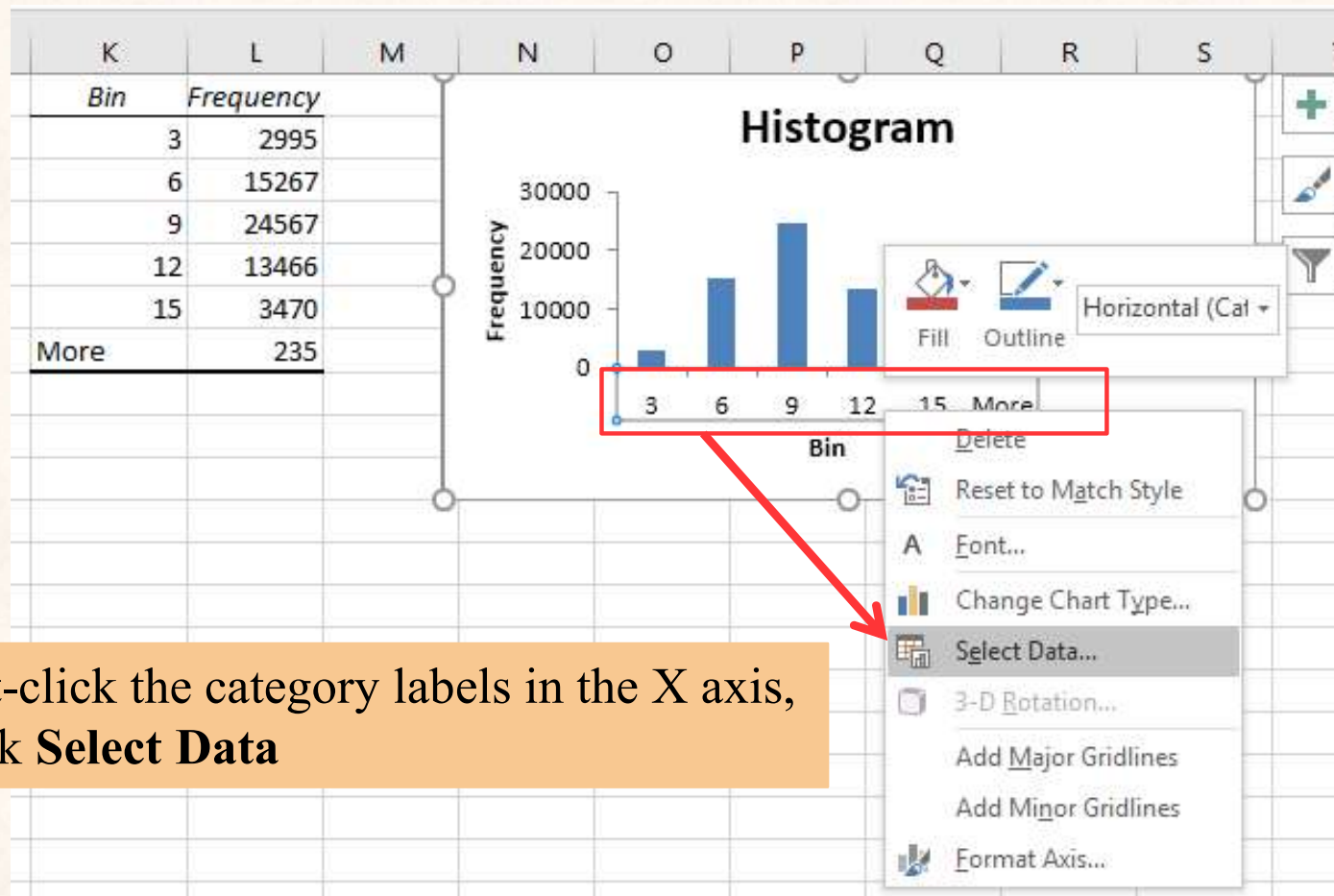
Part B.1 Create a Histogram using Data Analysis ToolPak



With the Analysis ToolPak, Excel adds the horizontal axis labels based on the bin numbers that you specify. How to display ranges instead of bin numbers?

Part B.1 Create a Histogram using Data Analysis ToolPak

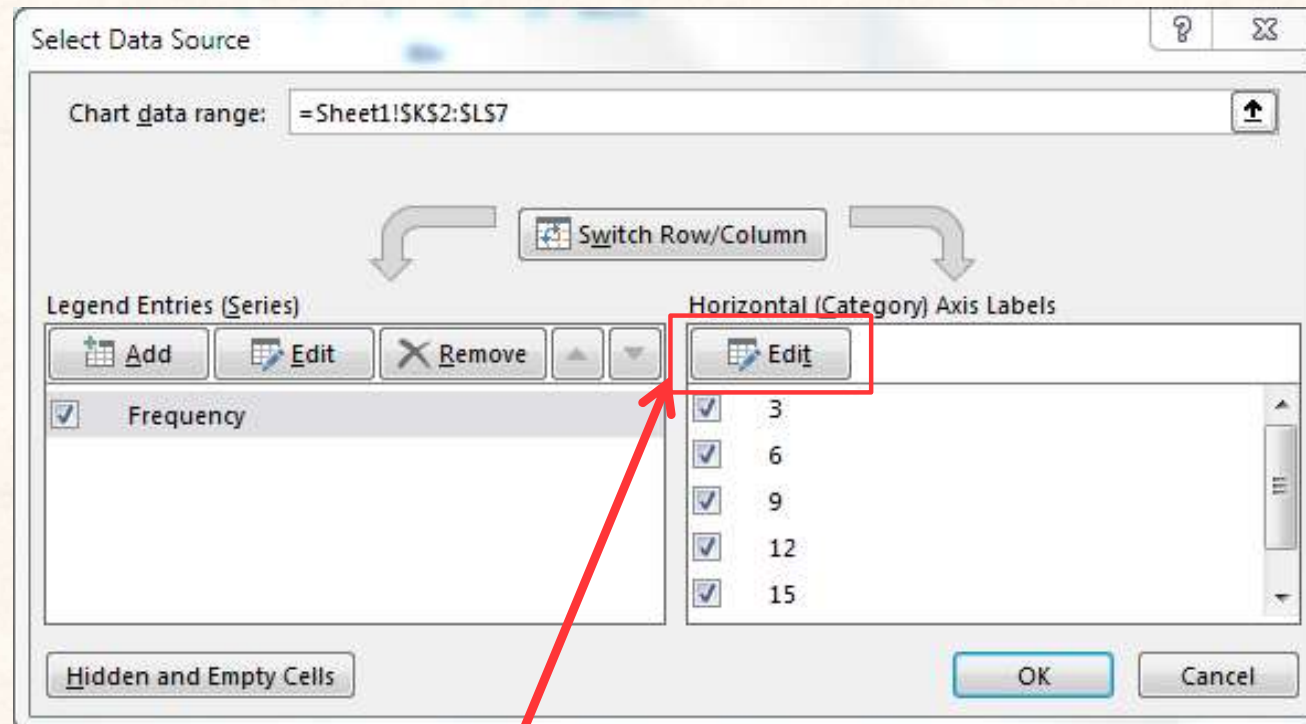
- Change the axis labels on an Excel histogram chart



1. Right-click the category labels in the X axis, and click **Select Data**

Part B.1 Create a Histogram using Data Analysis ToolPak

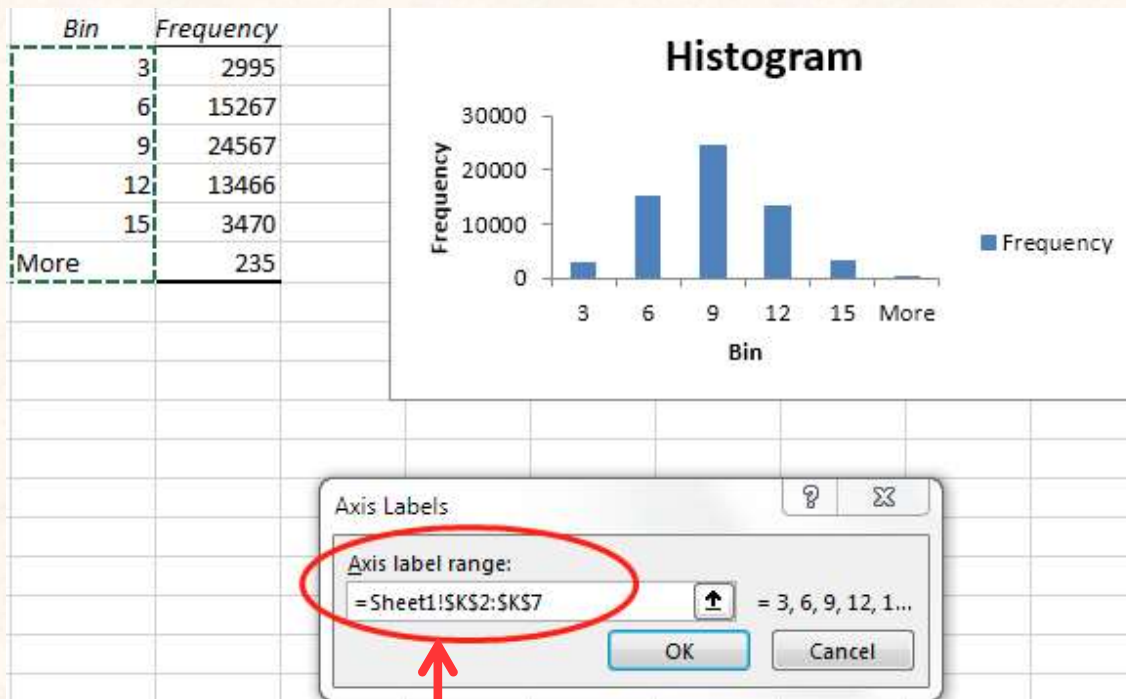
- Change the axis labels on an Excel histogram chart



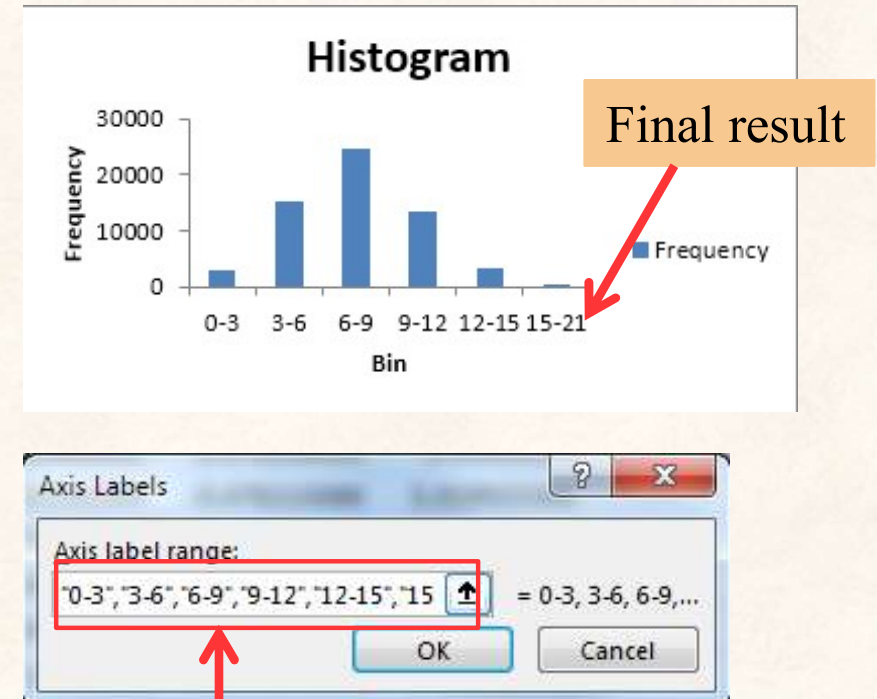
2. On the right-hand side pane, under Horizontal (category) Axis Labels, click the **Edit** button

Part B.1 Create a Histogram using Data Analysis ToolPak

- Change the axis labels on an Excel histogram chart



3. Enter the labels you want to display, separated by commas. If you are entering the intervals, enclose them in double quotes.



4. After entering the intervals, click **OK**

Part B.1 Create a Histogram using formulas

❑ FREQUENCY function:

It's the other way to create a histogram in Excel. The biggest advantage of using FREQUENCY function is that you won't have to re-do your histogram with each change in the input data. The histogram will **update automatically** as soon as you edit, add new or delete existing input values.

➤ Step 1: select your source data and create the bin numbers in a new column.

B	C	D	E	F	G	H	I
Response	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins
0.000353542	35.35419538	5.727265358	12.23935223	-0.478137483	-29.3236084		3
0.000381762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6
		5.292307854	12.23534012				9
		5.564445019	12.23534012				12
0.000371664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		15
0.000328045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415		18
0.000309085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318		
0.000358574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		

Select source data Input 1

Create the bin numbers

1. It's recommended to enter one more Frequency formula than the number of bins. The extra cell is required to display the count of values above the highest bin. For the sake of clarity, you can label it *More* (but don't include that *More* cell in your bins_array)

2. Select a range of adjacent cells where you want to output the frequencies (The Frequency function should be entered as a **multi-cell array formula**).

Part B.1 Create a Histogram using formulas

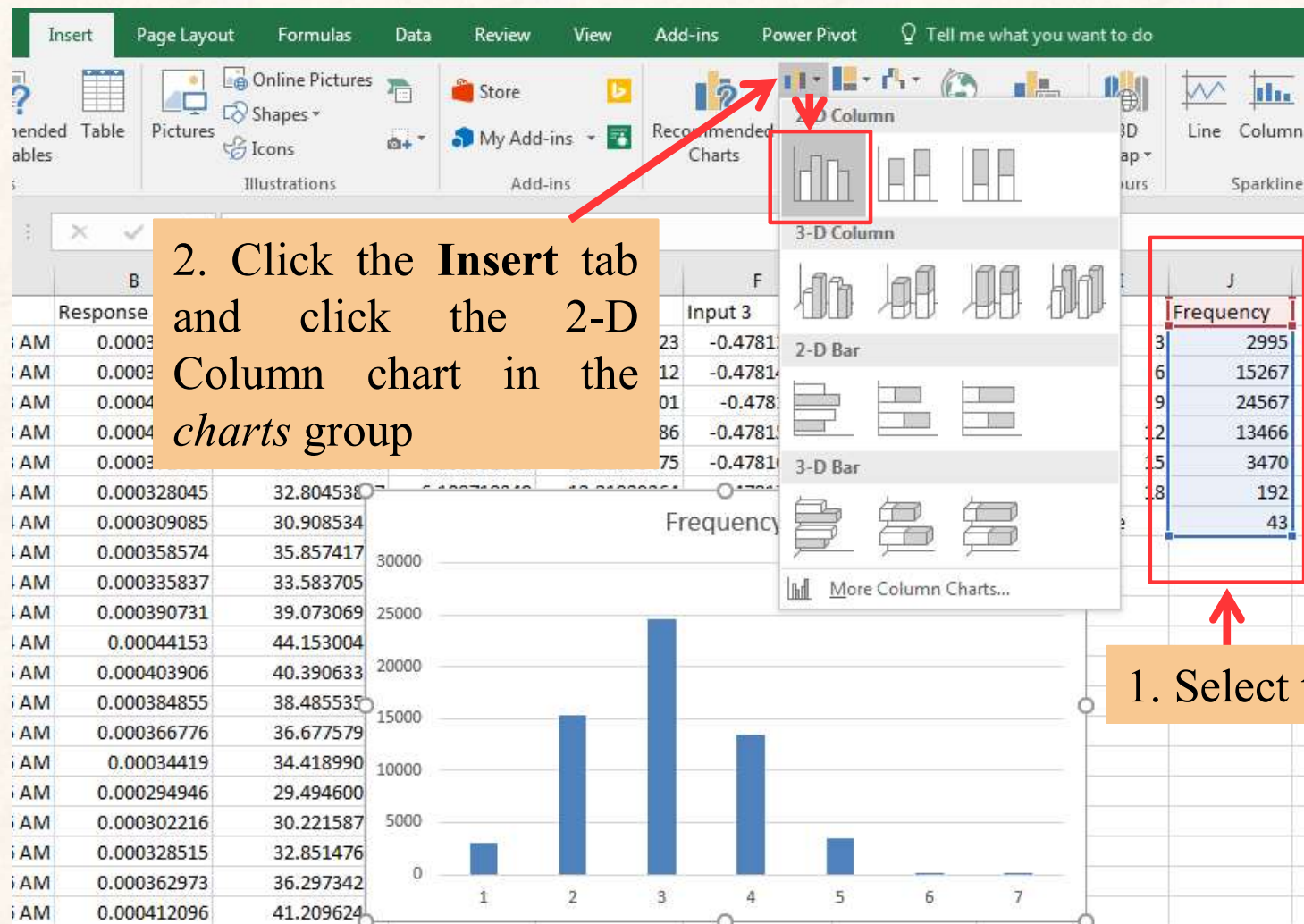
Get the results of Frequency. The Excel FREQUENCY function returns values that are greater than a previous bin and less than or equal to a given bin. The last Frequency formula returns the number of values greater than the highest bin.

	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4		Bins	Frequency
2	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		3	2995
2	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		6	15267
3	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		9	24567
9	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		12	13466
4	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		15	3470
5	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415		18	192
5	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318		More	43
4	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016			
7	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355			
1	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542			
3	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096			
6	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306			
5	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			
6	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763			

H	I	J
Interval	Bins	Frequency
0-3	3	2995
3-6	6	15267
6-9	9	24567
9-12	12	13466
12-15	15	3470
15-18	18	192
Over 18	More	43

Part B.1 Create a Histogram using formulas

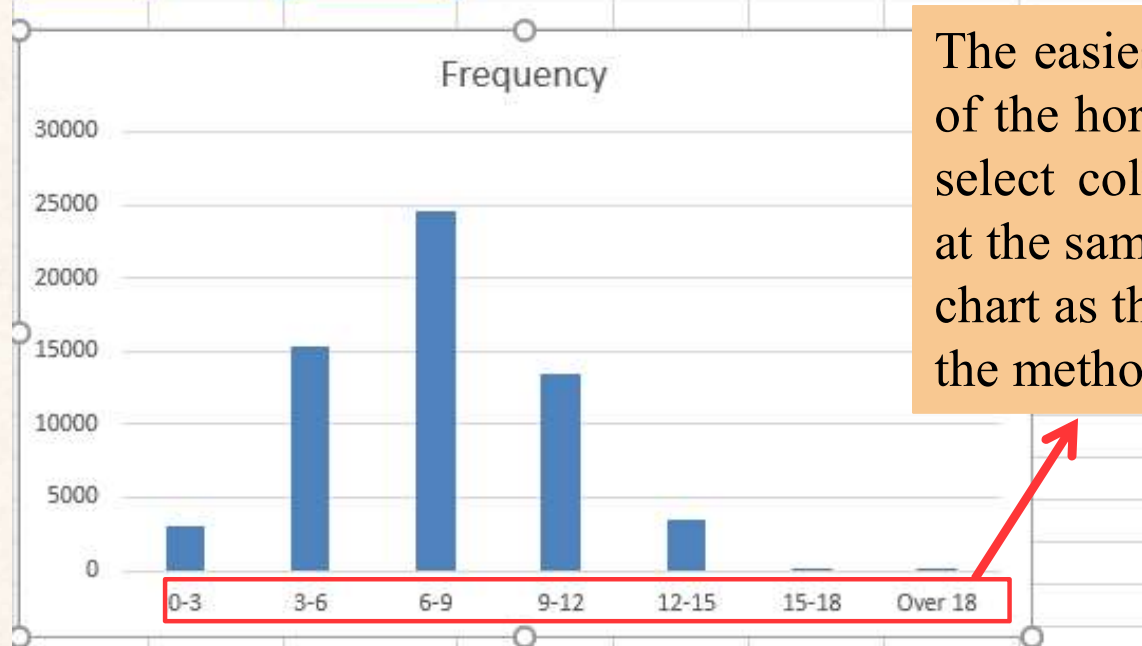
- Step 3: make a histogram based on the summary data



Part B.1 Create a Histogram using formulas

- Step 3: make a histogram based on the summary data

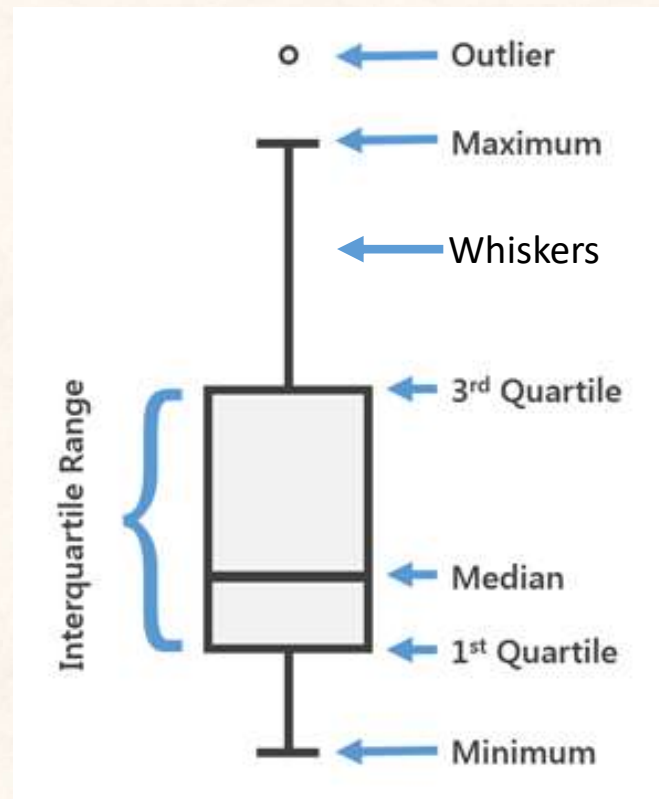
H	I	J	K	L	M	N	O
Interval	Bins	Frequency					
0-3	3	2995					
3-6	6	15267					
6-9	9	24567					
9-12	12	13466					
12-15	15	3470					
15-18	18	192					
Over 18	More	43					



The easiest way to change the default labels of the horizontal axis with bin numbers is to select columns of **Interval** and **Frequency** at the same time. Then click the 2-D Column chart as the previous slide. Also, you can use the method in P40.

Part B.2 Box Plot (Box-whisker Plot)

- ❑ In a box plot, numerical data is divided into quartiles, and a box is drawn between the first and third quartiles, with an additional line drawn along the second quartile to mark the median. The minimum and maximum outside the first and third quartiles are depicted with lines/whiskers.



Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

Value	Formula
Minimum value	MIN(cell range)
First quartile	QUARTILE.INC(cell range, 1)
Median value	QUARTILE.INC(cell range, 2)
Third quartile	QUARTILE.INC(cell range, 3)
Maximum value	MAX(cell range)

How to calculate the five values? You can have try it by yourself with the method we have learned in Part A.

Part B.2 Box Plot (Box-whisker Plot)

➤ Step 1: Calculate the quartile values

Since the formula is given, the easiest way to calculate the value is to use the formula in the formula bar directly.

	Response*100000	Input 1	Input 2	Input 3	Input 4		Input 1
542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084	Minimum	=MIN(D2:D60001)
562	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752	Q1	
593	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496	Median	
619	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347	Q3	
664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749	Maximum	
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415		
8085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318		
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		

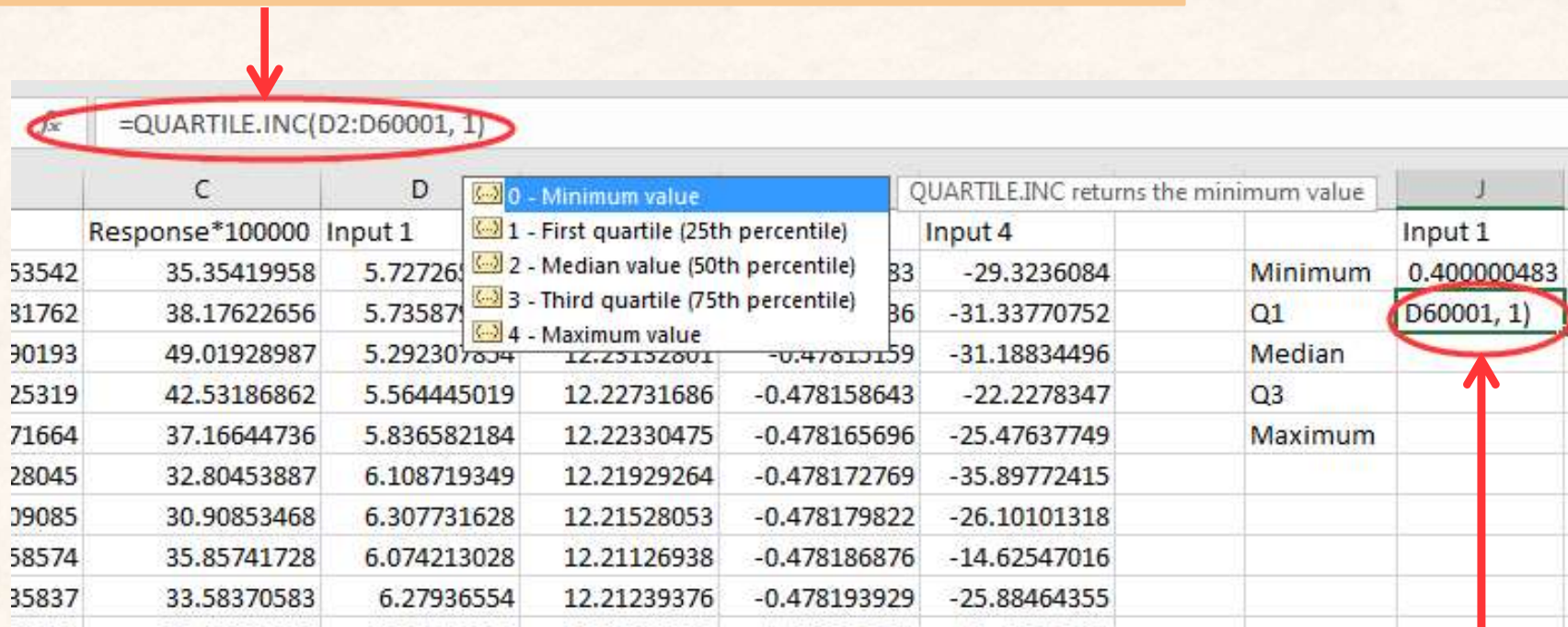
1. place your cursor in the cell labeled Minimum

2. Type “=MIN(cell range)” and press Enter

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type “=QUARTILE.INC(cell range, 1)” and press Enter



	C	D		J
	Response*100000	Input 1		Input 1
53542	35.35419958	5.727265		
81762	38.17622656	5.735879		
90193	49.01928987	5.292307834		
25319	42.53186862	5.564445019		
71664	37.16644736	5.836582184		
28045	32.80453887	6.108719349		
09085	30.90853468	6.307731628		
58574	35.85741728	6.074213028		
35837	33.58370583	6.27936554		

1. place your cursor in the cell labeled Q1

- Step 1: Calculate the quartile values

2. Type “=QUARTILE.INC(cell range, 2)” and press Enter

The screenshot shows an Excel spreadsheet with the following data:

	Response*100000	Input 1	0 - Minimum value	1 - First quartile (25th percentile)	2 - Median value (50th percentile)	3 - Third quartile (75th percentile)	4 - Maximum value
3542	35.35419958	5.727265					
31762	38.17622656	5.735879					
30193	49.01928987	5.292307					
25319	42.53186862	5.564445					
71664	37.16644736	5.836582					
28045	32.80453887	6.108719					
39085	30.90853468	6.307731					
38574	35.85741728	6.074213					
35837	33.58370583	6.279365					
30731	39.07306917	6.036407					
34153	44.15300484	5.857478					
33906	40.39063301	6.085069					
34855	38.48553584	6.155450					

The formula bar shows: `=QUARTILE.INC(D2:D60001, 2)`

The dropdown menu for the 'quart' argument shows the following options:

- 0 - Minimum value
- 1 - First quartile (25th percentile)
- 2 - Median value (50th percentile)
- 3 - Third quartile (75th percentile)
- 4 - Maximum value

The cell J2 contains 'Input 1'. The cell J3 contains '0.400000483'. The cell J4 contains '5.586860299'. The cell J5 contains 'D60001, 2)'. A red arrow points to the formula bar, and another red arrow points to the cell J5.

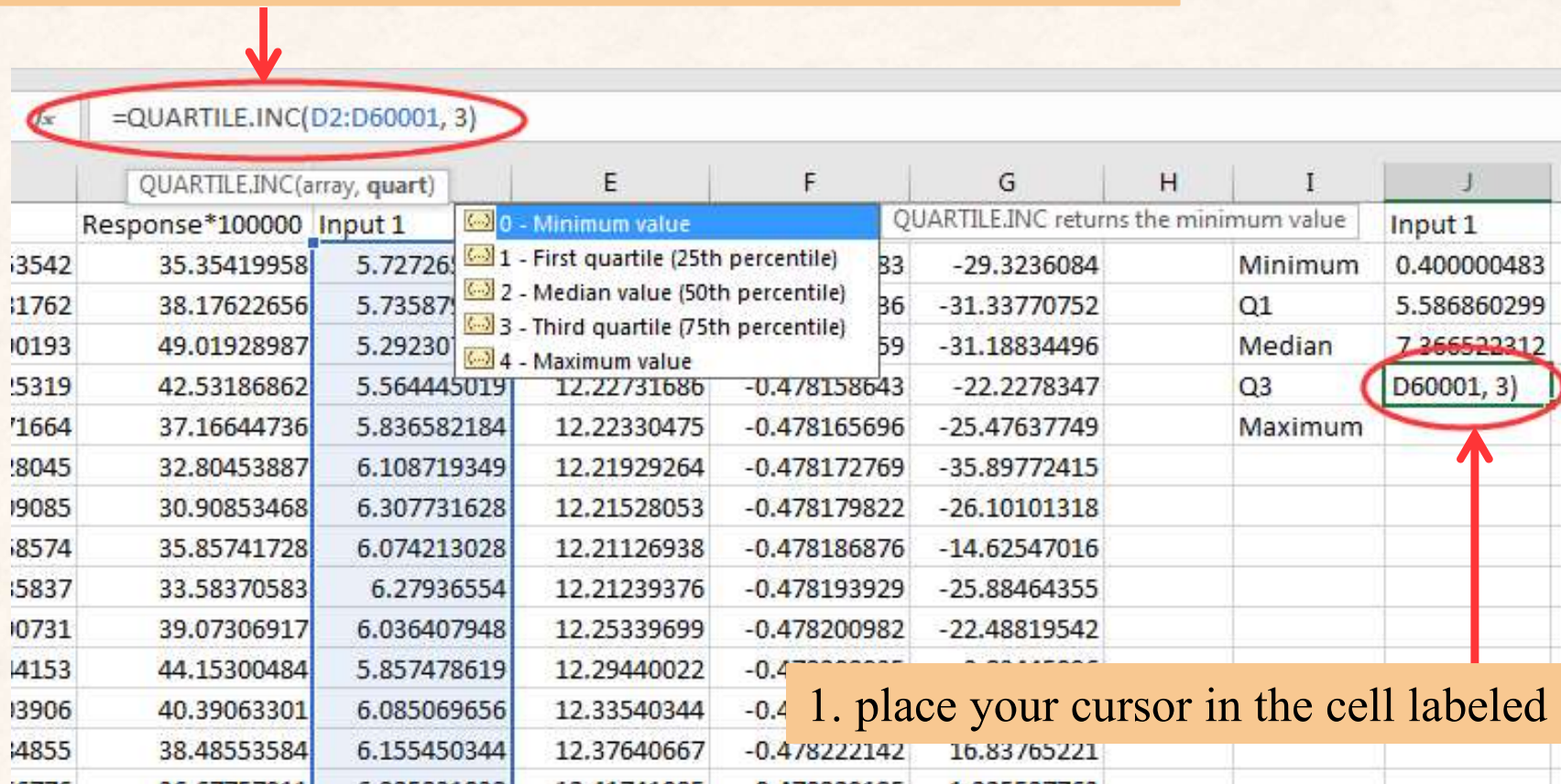
1. place your cursor in the cell labeled

1. place your cursor in the cell labeled Median

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type “=QUARTILE.INC(cell range, 3)” and press Enter

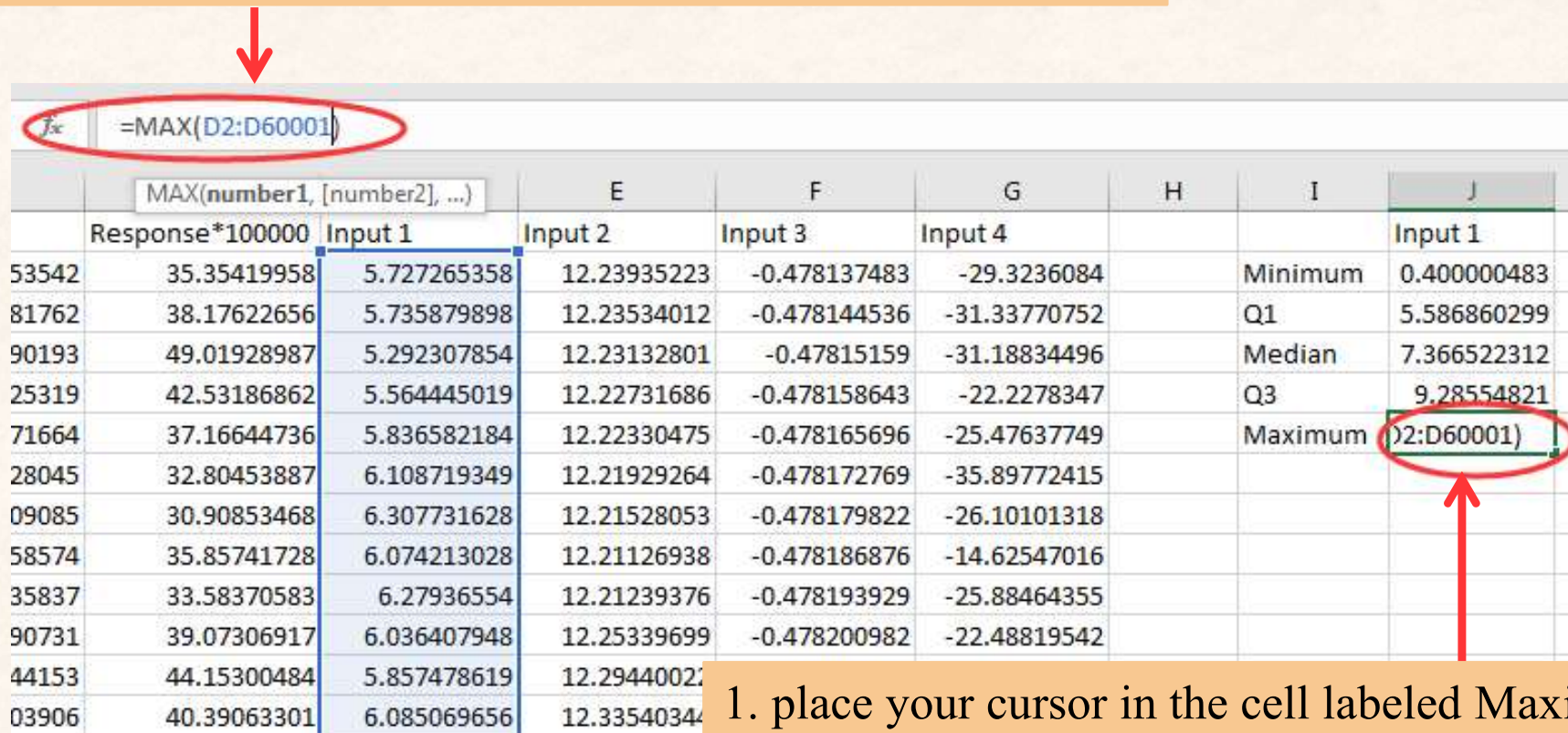


	QUARTILE.INC(array, quart)	E	F	G	H	I	J
	Response*100000	Input 1	0 - Minimum value	QUARTILE.INC returns the minimum value			Input 1
3542	35.35419958	5.72726	1 - First quartile (25th percentile)	33	-29.3236084	Minimum	0.400000483
1762	38.17622656	5.73587	2 - Median value (50th percentile)	36	-31.33770752	Q1	5.586860299
0193	49.01928987	5.29230	3 - Third quartile (75th percentile)	59	-31.18834496	Median	7.266523312
5319	42.53186862	5.564445019	4 - Maximum value		-22.2278347	Q3	D60001, 3)
1664	37.16644736	5.836582184			-25.47637749	Maximum	
8045	32.80453887	6.108719349			-35.89772415		
9085	30.90853468	6.307731628			-26.10101318		
8574	35.85741728	6.074213028			-14.62547016		
5837	33.58370583	6.27936554			-25.88464355		
0731	39.07306917	6.036407948			-22.48819542		
4153	44.15300484	5.857478619					
3906	40.39063301	6.085069656					
4855	38.48553584	6.155450344					

Part B.2 Box Plot (Box-whisker Plot)

- Step 1: Calculate the quartile values

2. Type “=Maximum(cell range)” and press Enter



		MAX(number1, [number2], ...)	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
53542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		Minimum	0.400000483
81762	38.17622656	5.735879898	12.23534012	-0.478144536	-31.33770752		Q1	5.586860299
90193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
25319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	=MAX(D2:D60001)
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
09085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			
58574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016			
35837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355			
90731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542			
44153	44.15300484	5.857478619	12.2944002					
03906	40.39063301	6.085069656	12.3354034					

1. place your cursor in the cell labeled Maximum

Part B.2 Box Plot (Box-whisker Plot)

➤ Step 2: Calculate quartile differences as the following

- First quartile – Minimum
- Median – first quartile
- Third quartile – mean
- Maximum – third quartile

Jx		=J2						
	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
3542	35.35419958	5.727265358	12.23935223	-0.478137483	-29.3236084		Minimum	0.400000483
1762	36.17622656	5.725870888	12.22524012	-0.478144536	-31.33770752		Q1	5.586860299
01	36.17622656	5.725870888	12.22524012	-0.478144536	-31.18834496		Median	7.366522312
5315	42.53180802	5.504443015	12.22751080	-0.478138043	-22.2278347		Q3	9.28554821
1664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
8045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	=J2
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	
07							Median-Q1	
41							Q3-Median	
3906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	

2. Type “Minimum” and press Enter

1. To create the dataset of the chat, paste Minimum firstly.

Part B.2 Box Plot (Box-whisker Plot)

- Step 2: Calculate quartile differences as the following

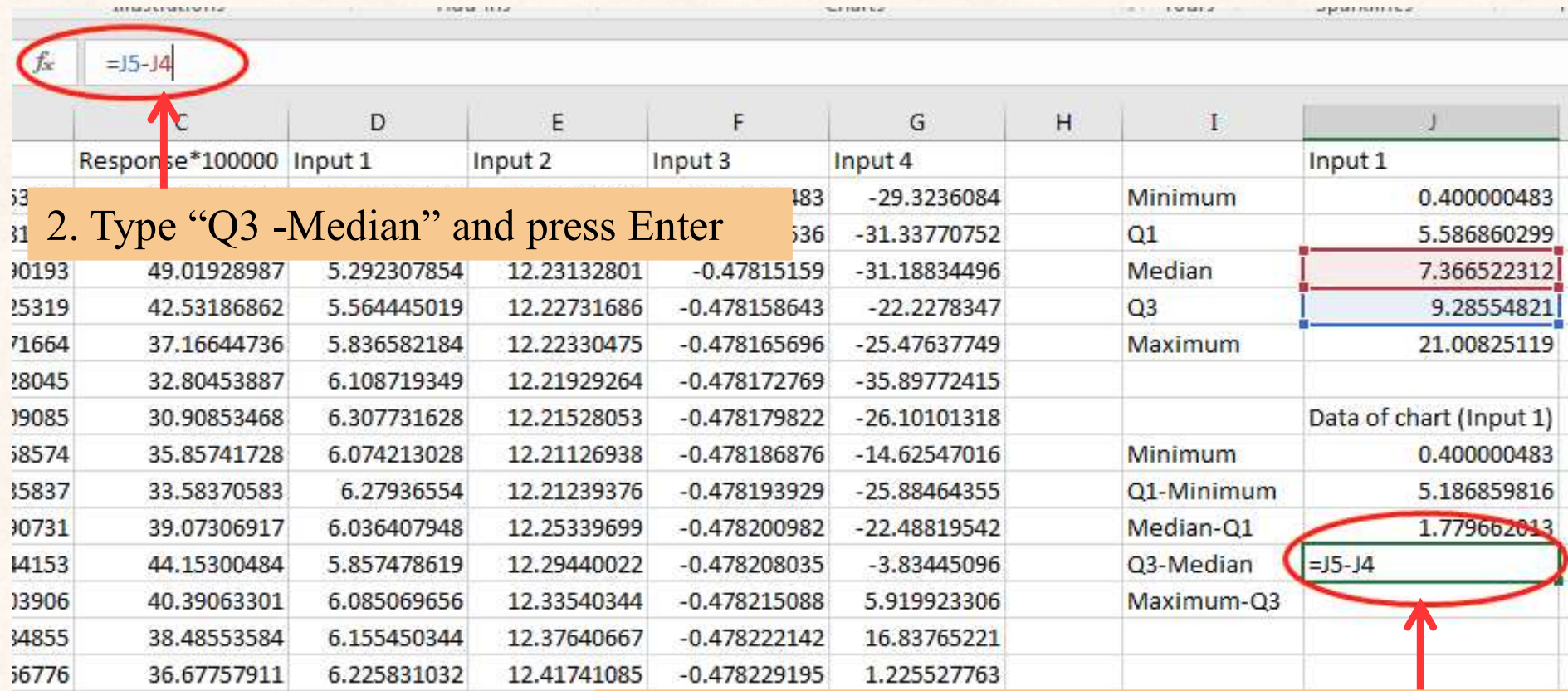
	f_x	=J3-J2							
		C	D	E	F	G	H	I	J
		Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
					483	-29.3236084		Minimum	0.400000483
					536	-31.33770752		Q1	5.586860299
								Median	7.366522312
								Q3	9.28554821
								Maximum	21.00825119
0193		49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496			
5319		42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347			
1664		37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749			
8045		32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
9085		30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
8574		35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
5837		33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	=J3-J2
0731		39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	
4153		44.15						Median	
3906		40.39						Maximum-Q3	
4855		38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			

2. Type “Q1-Minimum” and press Enter

1. place your cursor in the cell labeled Q1 -Minimum

Part B.2 Box Plot (Box-whisker Plot)

- Step 2: Calculate quartile differences



2. Type "Q3 -Median" and press Enter

	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
3							Minimum	0.400000483
1							Q1	5.586860299
0193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
5319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
9085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
8574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
5837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	5.186859816
0731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	1.779662013
4153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		Q3-Median	=J5-J4
3906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	
4855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			
6776	36.67757911	6.225831032	12.41741085	-0.478229195	1.225527763			

1. place your cursor in the cell labeled Q3 -Median

Part B.2 Box Plot (Box-whisker Plot)

- Step 2: Calculate quartile differences

fx		=J6-J5						
	C	D	E	F	G	H	I	J
	Response*100000	Input 1	Input 2	Input 3	Input 4			Input 1
				137483	-29.3236084		Minimum	0.400000483
				144536	-31.33770752		Q1	5.586860299
90193	49.01928987	5.292307854	12.23132801	-0.47815159	-31.18834496		Median	7.366522312
25319	42.53186862	5.564445019	12.22731686	-0.478158643	-22.2278347		Q3	9.28554821
71664	37.16644736	5.836582184	12.22330475	-0.478165696	-25.47637749		Maximum	21.00825119
28045	32.80453887	6.108719349	12.21929264	-0.478172769	-35.89772415			
09085	30.90853468	6.307731628	12.21528053	-0.478179822	-26.10101318			Data of chart (Input 1)
58574	35.85741728	6.074213028	12.21126938	-0.478186876	-14.62547016		Minimum	0.400000483
35837	33.58370583	6.27936554	12.21239376	-0.478193929	-25.88464355		Q1-Minimum	5.186859816
90731	39.07306917	6.036407948	12.25339699	-0.478200982	-22.48819542		Median-Q1	1.779662013
44153	44.15300484	5.857478619	12.29440022	-0.478208035	-3.83445096		Q3-Median	1.919025898
03906	40.39063301	6.085069656	12.33540344	-0.478215088	5.919923306		Maximum-Q3	=J6-J5
84855	38.48553584	6.155450344	12.37640667	-0.478222142	16.83765221			

2. Type “Maximum-Q3” and press Enter

1. place your cursor in the cell labeled Maximum-Q3

Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart

2. Click Insert

3. Select Insert Column chart and then click stacked column

1. Select the data to create a chart

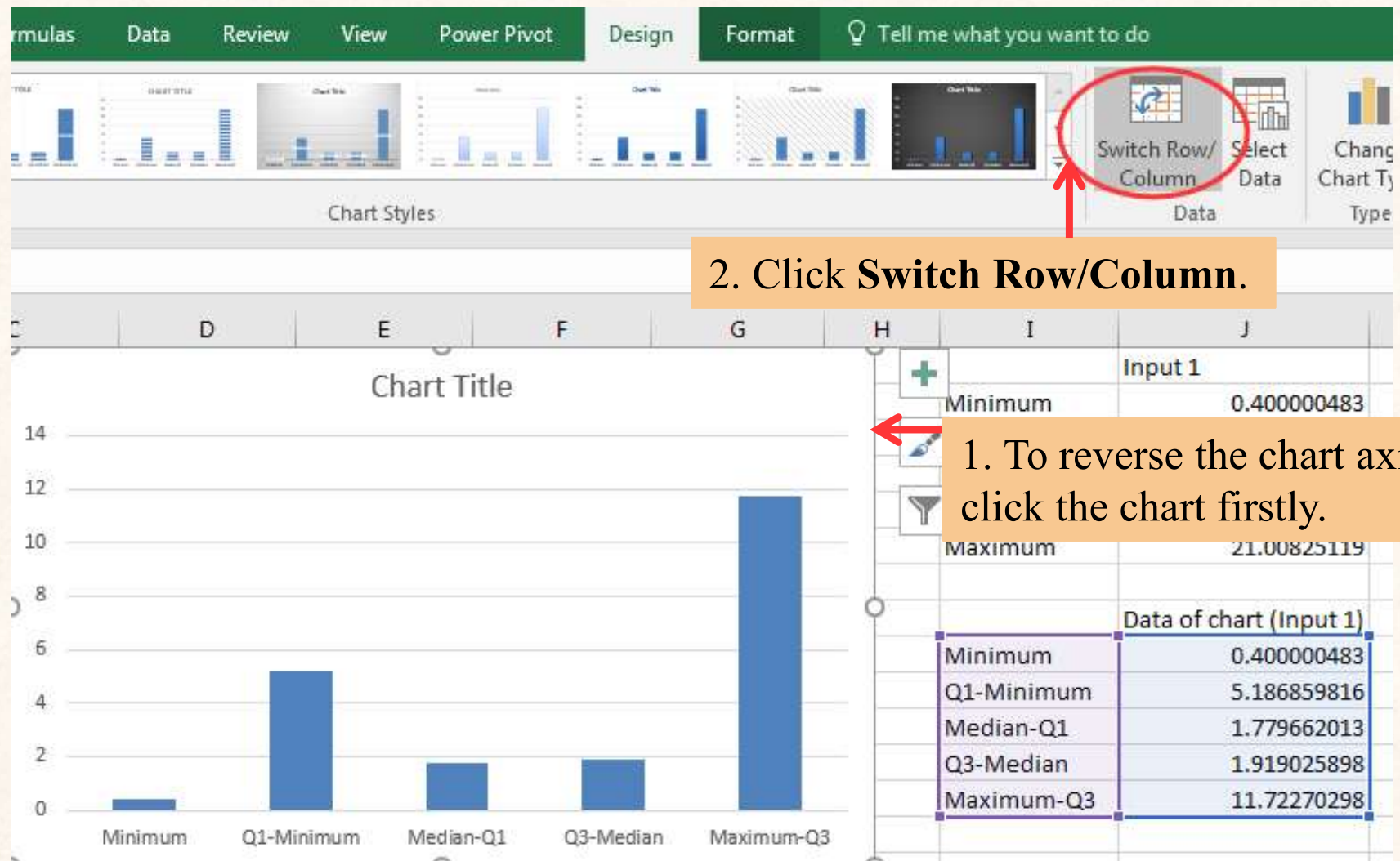
	Response					
13 AM	0.00035354					
13 AM	0.00038176					
13 AM	0.00049019					
13 AM	0.000425319	42.53186862	5.564445019	12.22731686	-0.4781	
13 AM	0.000371664	37.16644736	5.836582184	12.22330475	-0.4781	
14 AM	0.000328045	32.80453887	6.108719349	12.21929264	-0.4781	
14 AM	0.000309085	30.90853468	6.307731628	12.21528053	-0.4781	
14 AM	0.000358574	35.85741728	6.074213028	12.21126938	-0.4781	
14 AM	0.000335837	33.58370583	6.27936554	12.21239376	-0.4781	
14 AM	0.000390731	39.07306917	6.036407948	12.25339699	-0.4782	-22.48819542
14 AM	0.00044153	44.15300484	5.857478619	12.29440022	-0.4782	-3.83445096
15 AM	0.000403906	40.39063301	6.085069656	12.33540344	-0.4782	5.919923306
15 AM	0.000384855	38.48553584	6.155450244	12.37640667	-0.4782	16.82765221
15 AM	0.000366776	36.67757911	6.2258			

	Input 1
mum	0.400000483
	5.586860299
ian	7.366522312
	9.28554821
mum	21.00825119

	Data of chart (Input 1)
mum	0.400000483
Minimum	5.186859816
Median-Q1	1.779662013
Q3-Median	1.919025898
Maximum-Q3	11.72270298

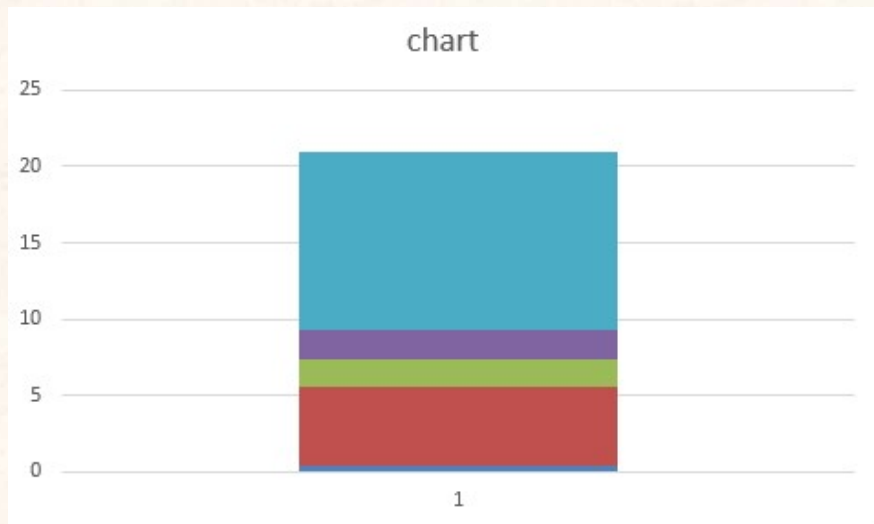
Part B.2 Box Plot (Box-whisker Plot)

- Step 3: Create a stacked column chart

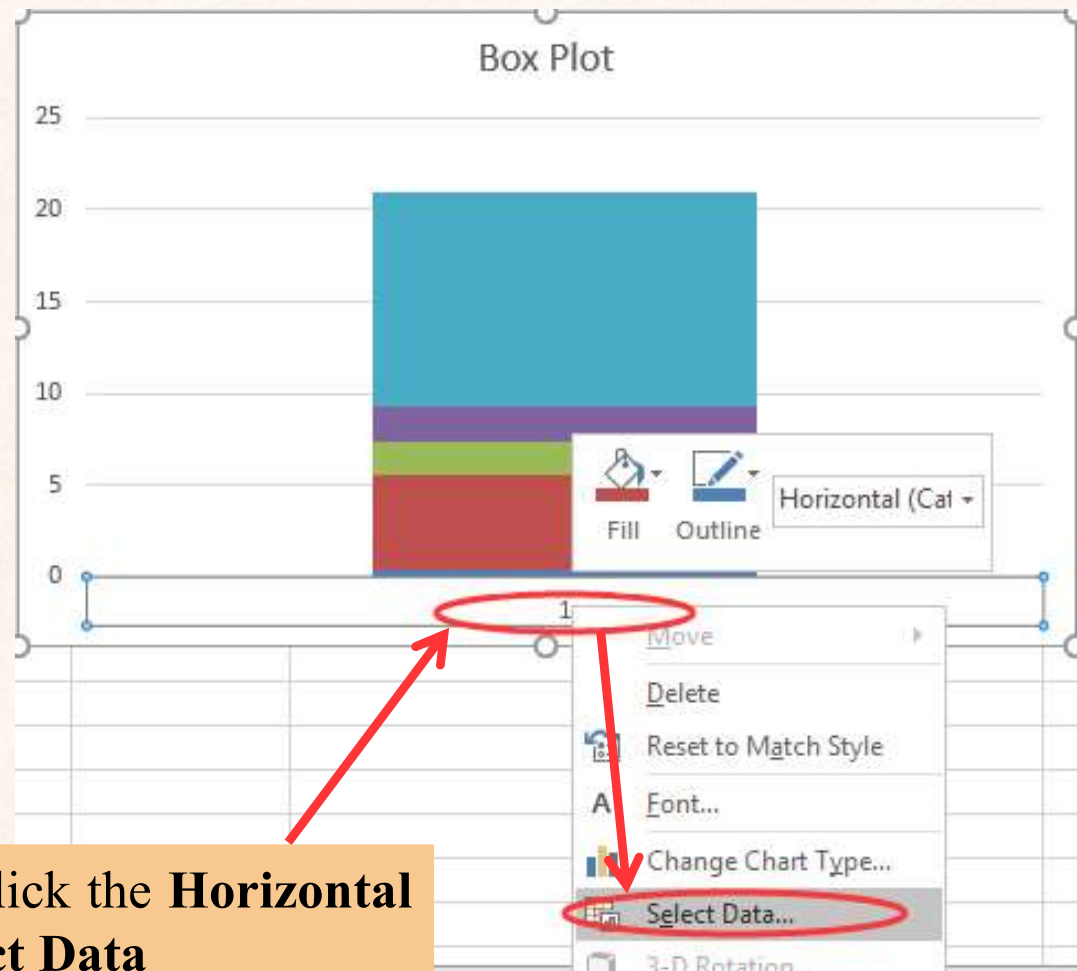


Part B.2 Box Plot (Box-whisker Plot)

➤ Step 3: Create a stacked column chart



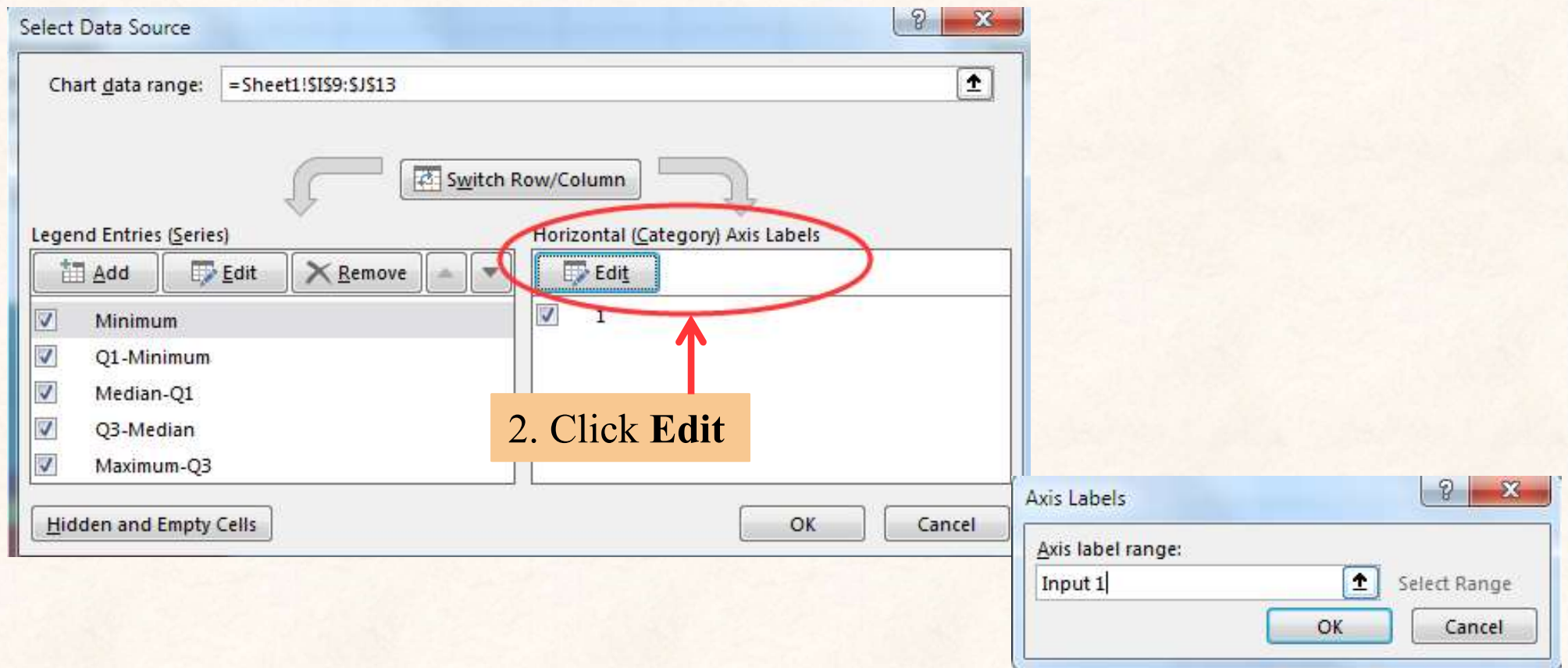
Original stacked column chart



1. To rename your columns, right click the **Horizontal (Category) axis labels** side -> **Select Data**

Part B.2 Box Plot (Box-whisker Plot)

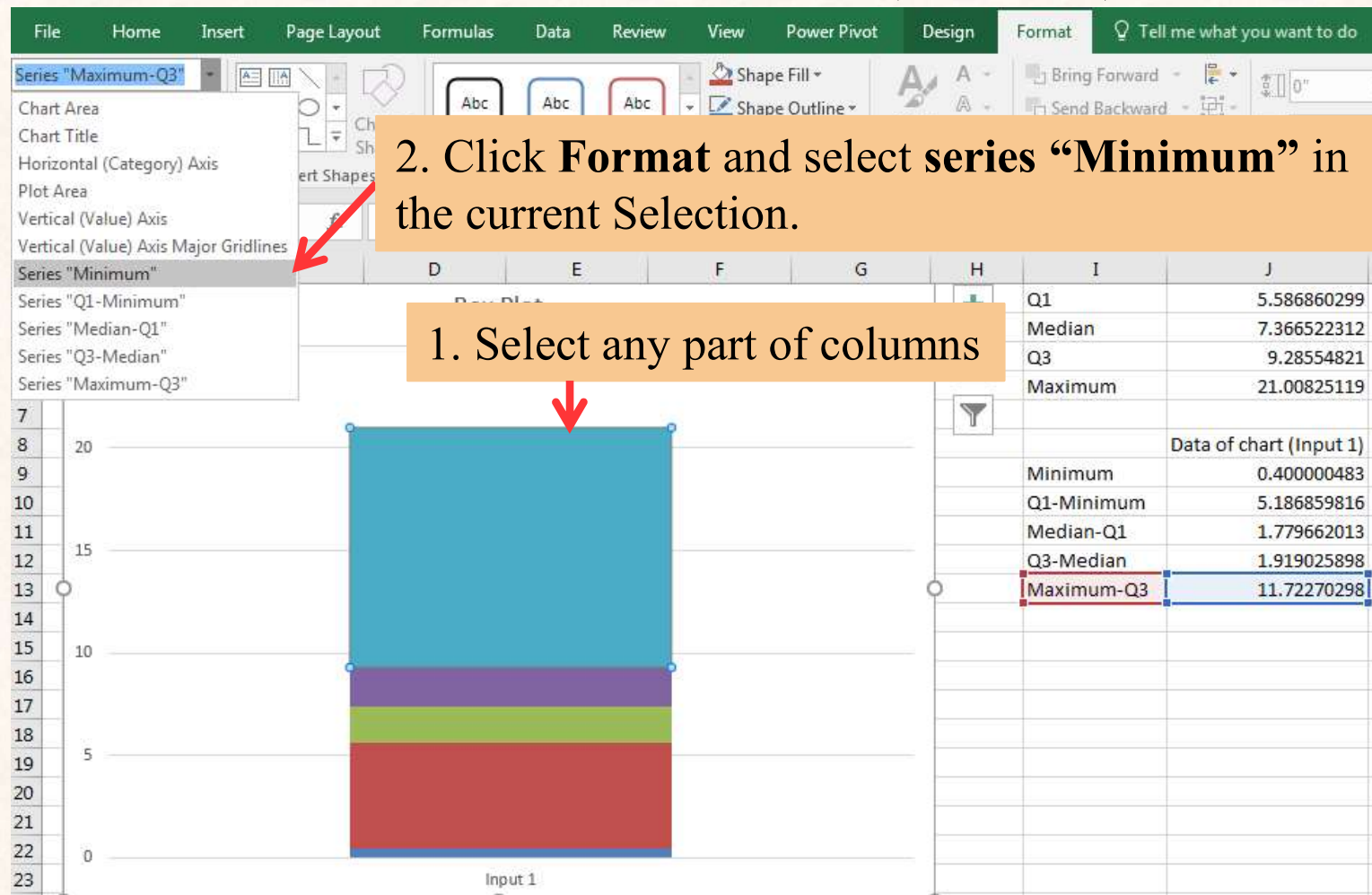
- Step 3: Create a stacked column chart



3. A new window pops out. Change the name as you want and click **OK**.

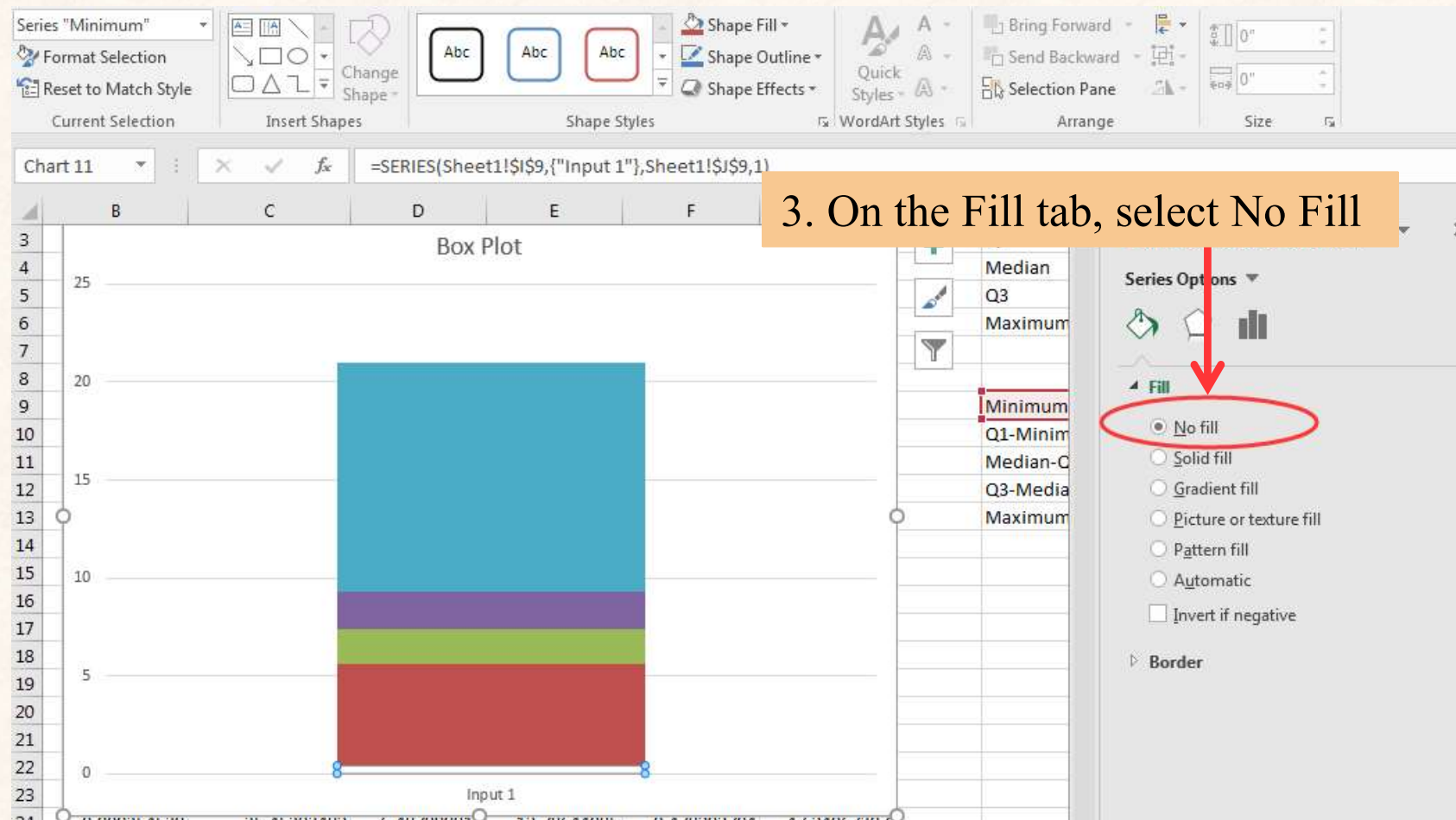
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.1 Hidden the bottom data series (minimum)



Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.1 Hidden the bottom data series (minimum)

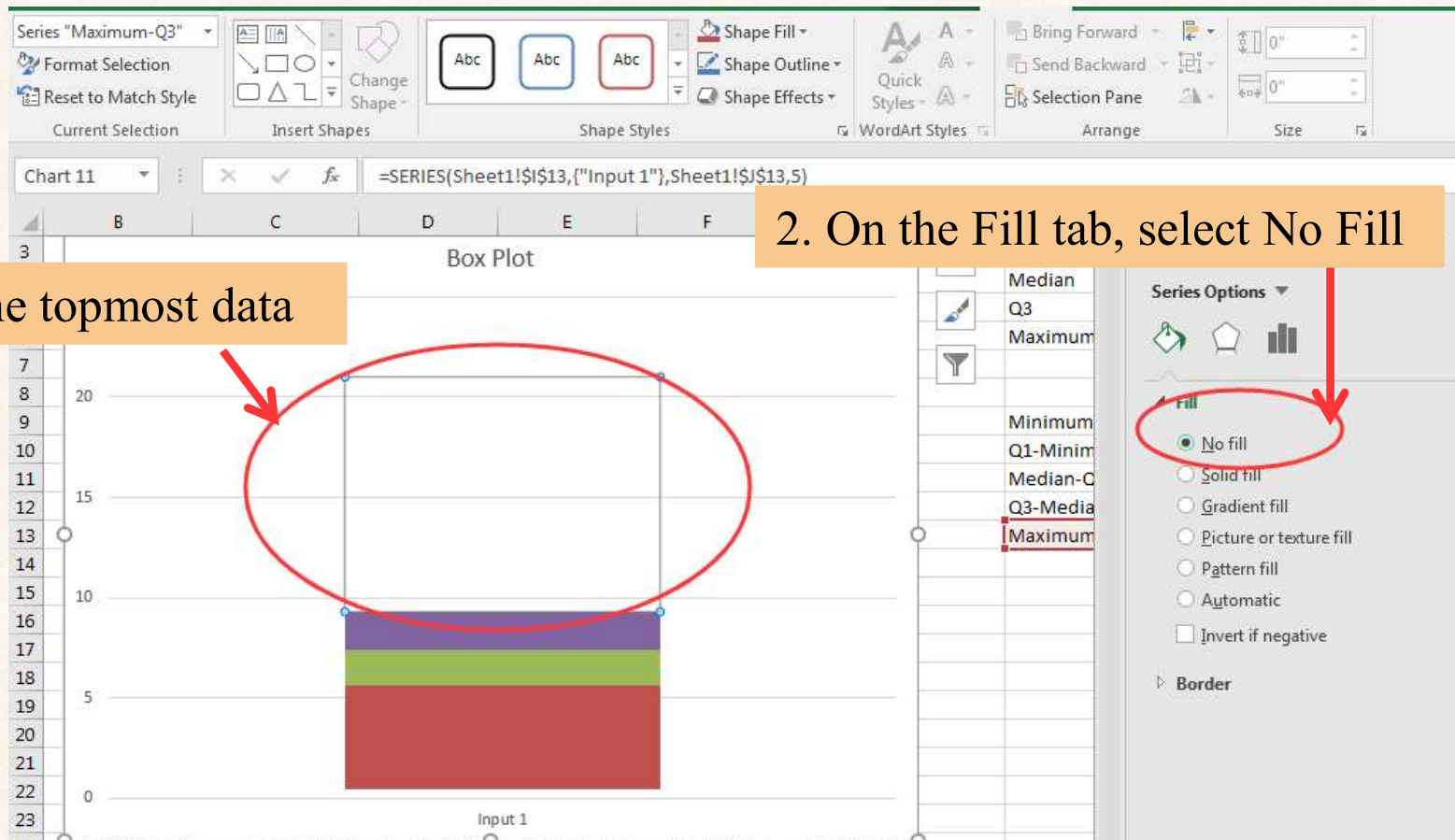


Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.2 Create whiskers for the box plot

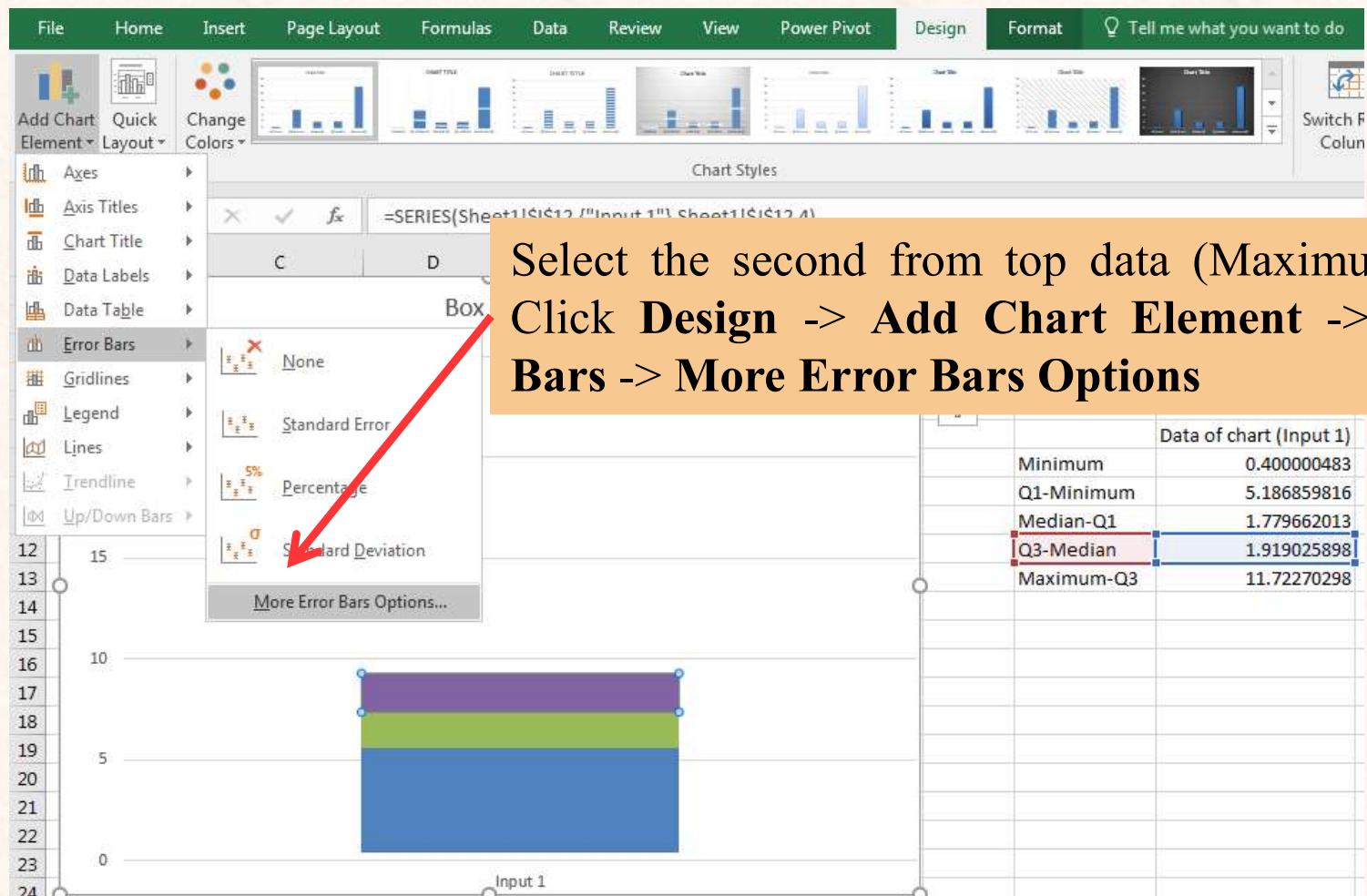
1. Click the topmost data

2. On the Fill tab, select No Fill



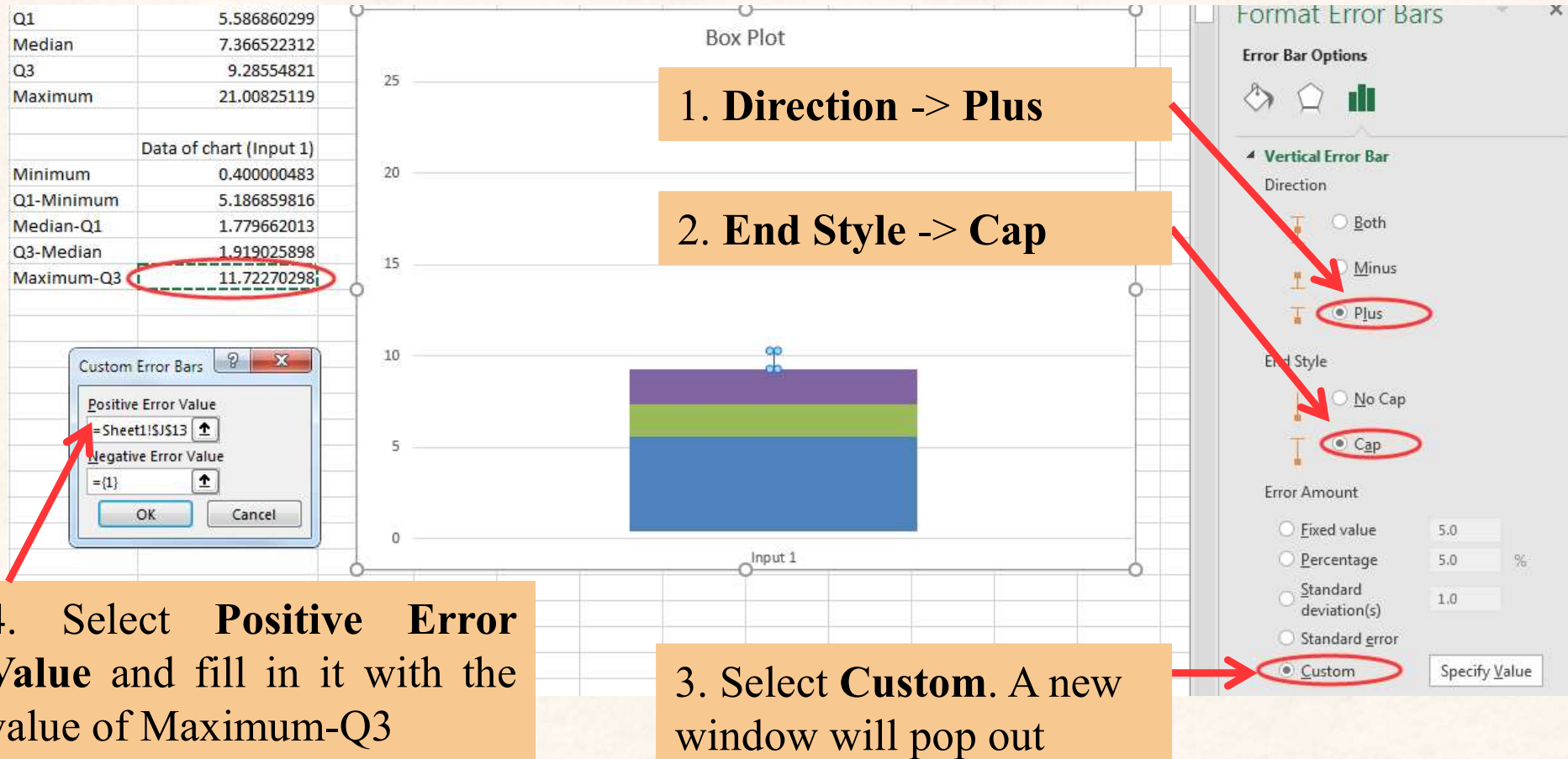
Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.2 Create whiskers for the box plot



Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.2 Create whiskers for the box plot



Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.2 Create whiskers for the box plot

1. Repeat the previous steps to select **No Fill** format for the second-from-bottom data (Q1-Minimum).

Minimum	0.400000483
Q1-Minimum	5.186859816
Median-Q1	1.779662013
Q3-Median	1.919025898
Maximum-Q3	11.72270298

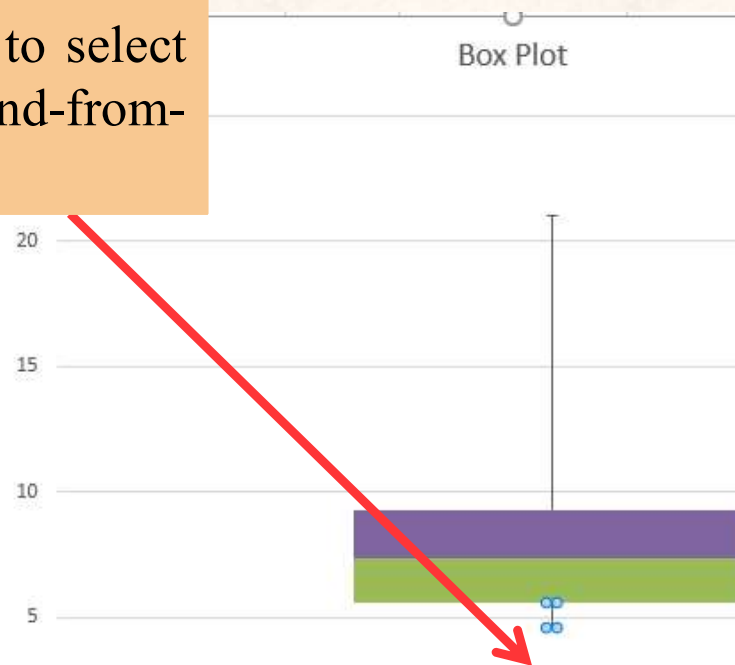
Custom Error Bars

Positive Error Value
={1}

Negative Error Value
=Sheet1!\$J\$10

OK Cancel

5. Select **Negative Error Value** and fill in it with the value of Q1-Minimum



4. Select **Custom**. A new window will pop out

2. In **Error Bar Options**, **Direction -> Minus**

Error Bar Options



Vertical Error Bar

Direction

☐ Both

☒ Minus

☐ Plus

☐ No Cap

☒ Cap

Error Amount

☐ Fixed value 0.1

☐ Percentage 5.0 %

☐ Standard deviation(s) 1.0

☐ Standard error

☒ Custom

Specify Value

Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.2 Create whiskers for the box plot

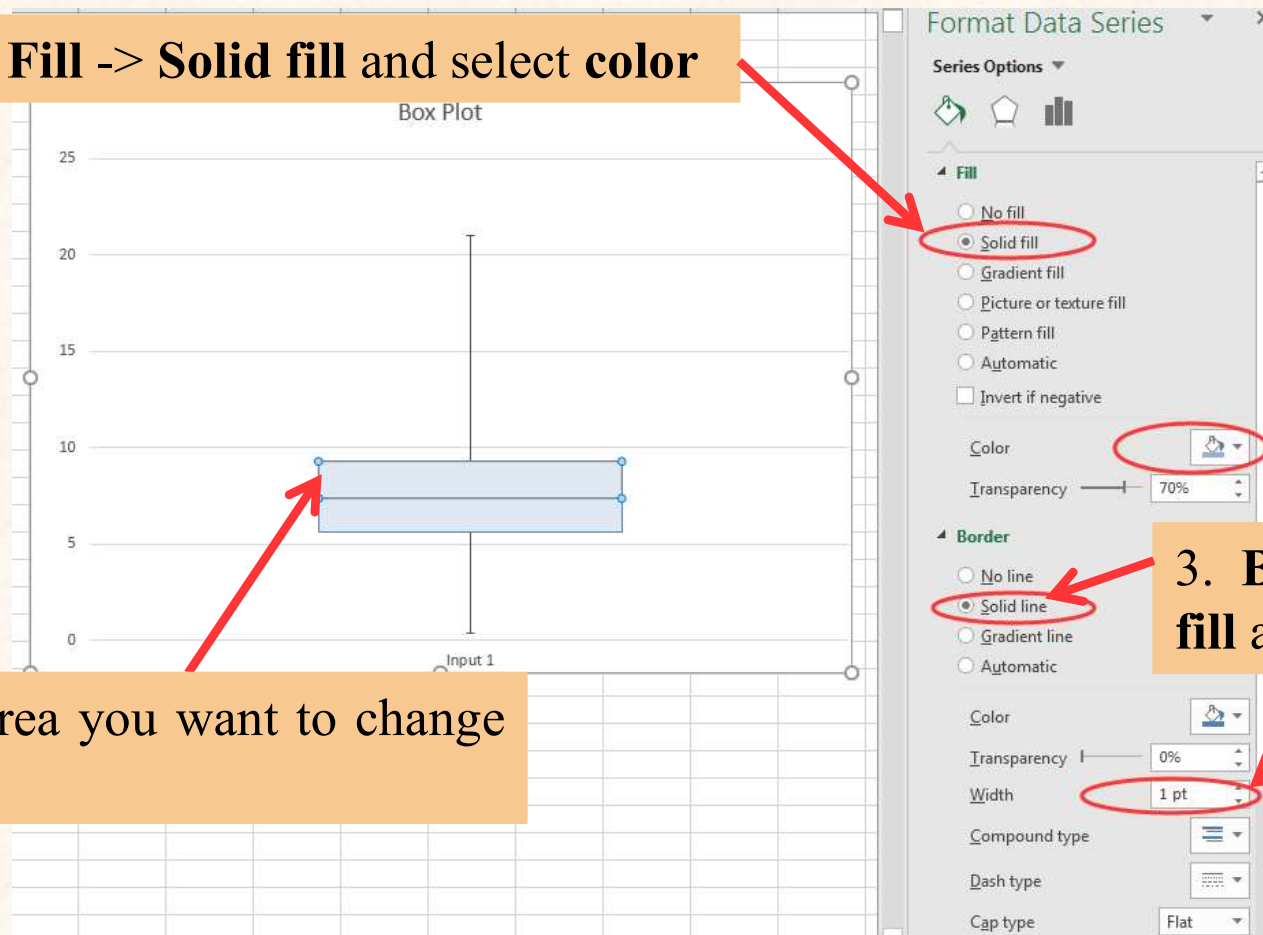


Part B.2 Box Plot (Box-whisker Plot)

- Step 4: Convert the stacked column chart to the box plot
 - 4.3 color the middle area. Box plots are usually drawn in one fill color, with a slight outline border.

2. Fill -> Solid fill and select color

1. Select the area you want to change its color.



3. Border -> Solid fill and select width

Exercise



- ❑ 1. Calculate descriptive statistics of Input2, Input 3 and Response
- ❑ 2. Plot histograms and box plots for Input 2, Input 3 and Response. You can try to plot box- whiskers of Input 1, Input 2, Input 3 and response in the same figure.