



*A SunCam online continuing education course*

# Engineering Methods in *Microsoft Excel*

## Part 3: Data Analysis in Excel

by

Kwabena Ofosu, Ph.D., P.E., PTOE



Engineering Methods in Excel  
*A SunCam online continuing education course*

**Abstract**

This course is the part of a series on engineering methods in *Microsoft Excel*; tailored to practicing engineers. It features *Microsoft Excel* tools that can be used for a wide range of engineering analyses and data management. This course covers introductory topics on engineering data analyses. Each topic includes a review of the fundamental principles, followed by a real-life example encountered by a practicing engineer. The real-life examples are then formulated and implemented in *Microsoft Excel* and worked using the various *Excel* tools, spreadsheet techniques, and built-in functions. Examples from various engineering fields are used to demonstrate the concepts and methods learned throughout this course. Upon completion of this course, practitioners will be able to apply the methods learned to a variety of engineering problems, and identify situations in their fields of specialty where the innovative application of these tools and methods will be advantageous to their output and to their work product.



Engineering Methods in Excel  
*A SunCam online continuing education course*

**TABLE OF CONTENTS**

Abstract .....	2
1. INTRODUCTION .....	4
1.1 Overview .....	4
2. DATA ANALYSIS .....	5
2.1 Introduction .....	5
2.2 Data .....	5
2.3 Organizing and Summarizing Data .....	6
2.4 Frequency Diagrams .....	9
2.5 <i>Excel Data Analysis Tools</i> .....	15
2.6 Data Analysis in <i>Excel</i> .....	21
2.7 Descriptive Statistics .....	109
2.8 Quartiles and Percentiles .....	112
2.9 Descriptive Statistics in <i>Excel</i> .....	118
3. CONCLUSION .....	154
REFERENCES .....	155



*A SunCam online continuing education course*

## **1. INTRODUCTION**

### **1.1 Overview**

In their fields of specialty, Engineers are constantly challenged with solving a wide range of complex analytical and computational problems. These problems involve analysis methodologies and the management of data. The application of computers enables repetitive, time-consuming and often tedious calculations to be conducted rapidly, efficiently, and less prone to errors. The application of computer tools also enables the results and outputs of such engineering analyses to be readily transferred and incorporated into reports and other engineering documents. An even greater advantage—in terms of productivity and efficiency—is realized when these calculations and outputs are replicated across numerous projects. Competence in computer skills predisposes engineers to pursue and develop more creative and innovative solutions to problems than otherwise.

*Microsoft Excel* is an electronic spreadsheet program developed by the *Microsoft* company, and part of the *Microsoft Office* suite of products. A spreadsheet is a grid that organizes data and calculations into columns and rows. The intersection of a column and a row is called a cell. An electronic spreadsheet enables users to store, organize, manipulate, and analyze data in the cells of the spreadsheet. As of this publication, the software is into the 2016 release.

*Microsoft Excel* is widely and increasingly being used as a tool to assist engineers in conducting and replicating intricate calculations and analysis, designing complex systems, and managing large data sets. This course presents a selection of engineering analysis topics and how to formulate, implement and solve them within *Microsoft Excel*.

This course presents fundamental principles and engineering applications of engineering data analysis—and demonstrates the *Excel* tools, methods, and strategies that can be used to formulate and solve them.

Upon completion of this course, participants will have gained insight into applying *Excel* tools, methods, and strategies in analyzing, formulating and implementing a variety of engineering related problems, as well as managing large sets of complex data. Participants will also be able to identify professional situations where the application of innovative *Excel* techniques will be of great benefit and advantage, as it will enable practitioners to significantly improve their productivity, efficiency, and the quality of their work product.



Engineering Methods in Excel  
A SunCam online continuing education course

## 2. DATA ANALYSIS

### 2.1 Introduction

In this chapter, we shall cover some of the tools for statistical analysis that are part of the *Excel Data Analysis Tools* Add-in. Statistical analysis involves organizing, summarizing, analyzing, interpreting, manipulating etc., data. In this presentation, data from various fields of engineering shall be used for illustrative purposes.

### 2.2 Data

Data (singular **datum**) are observations or results that have been gathered. For example, measurements, experimental results, survey responses, test scores, etc. In a study, if data is collected on every single eligible member of the subject population, such a study is called a **census**. A **sample** is a subset of a population. The field of statistics involves using data from a sample—collected in an appropriate manner—to make valid inferences and conclusions about a population.

A numerical measure that describes a population is called a **parameter**. A numerical measure that describes a sample is called a **statistic**. For example, the average speed of all vehicles that use a roadway over a given period would be a parameter; whereas the average speed of a randomly selected sample of vehicles on this roadway would be a statistic.

**Quantitative data** consists of numerical values representing counts, measurements etc. For example, traffic volumes, or number of units produced by an industrial process. **Qualitative data**, also called categorical data or attribute data are nonnumeric in nature. For example, gender, hair color, hobby.

Quantitative data may be discrete or continuous. **Discrete data** consists of countable numeric values or integer values. For example, the number of packages processed by a mail service (e.g. 10, 15, 250, 681, etc.). **Continuous data** consists of integer and decimal values. For example, the volume in liters of a chemical product manufactured per day by a chemical plant (e.g. 4.8, 4.0, 9.85, 15.655, etc.). With continuous data, for any specified range of values, there are infinitely many possible values that may be obtained.



Engineering Methods in Excel  
A SunCam online continuing education course

Qualitative data may be described as **ordinal** or **nominal**. Ordinal data is such that the nonnumeric values express order, ranking, or scale. For example, letter grades (A, B, C, D, E, F) earned by students on a test; or the seniority of positions in a police department (corporal, lieutenant, captain, major, chief etc.). Nominal data on the other hand do not express a scheme of ordering. For example, eye color (brown, blue, green etc.) does not express a scale or order or ranking.

### 2.3 Organizing and Summarizing Data

The number of times a value occurs in a data set is called the **frequency** of the value. For grouped data (data arranged in groups, or intervals also known as **classes**), the frequency of an interval is the number of data values that fall within the interval. A **frequency distribution**, also called a frequency table, summarizes a data set by listing data values—individually or as intervals—along with their corresponding frequencies.

Consider the following speeds (in miles per hour) of vehicles on a highway: 53, 56, 72, 59, 63, 59, 56, 67, 59, 63. The values may be tallied and the frequencies appended as follows:

Speed (mph)	Tally	Frequency (f)
53	/	1
56	//	2
59	///	3
63	//	2
67	/	1
72	/	1
		$\Sigma f = 10$

Note that the total frequency must always equal the number of observations (the sample size).

For a larger dataset, however, tallying individual values is tedious and cumbersome—therefore the use of intervals (also called classes) is preferred. An example of a frequency table using intervals is as follows:



Engineering Methods in Excel  
A SunCam online continuing education course

Speed (mph)	Frequency (f)
[20, 30)	2
[30, 40)	8
[40, 50)	30
[50, 60)	40
[60, 70)	15
[70, 80)	5
	$\Sigma f = 100$

The interval convention used in the above table,  $[a, b)$  means that the lower limit  $a$  is included in the interval, and the upper limit  $b$  is not included in the interval, and will be included in the subsequent interval.

### Frequency Distributions

The procedure for constructing the frequency distribution (or frequency table) using intervals is as follows:

1. Compute the range of the data—which is the difference of the high value and the low value.
2. Select the number of intervals to be used—typically between 5 and 20 depending on the nature of the data.
3. Compute the interval width—which is the range divided by the number of intervals, and rounded to a convenient number to work with; typically, a multiple of 5, 10, 20, 50, 100 etc., depending on the nature of the data.
4. Select the lower limit of the first interval such that it encloses the low value of the data. The upper limit of the interval is the lower limit plus the interval width. The upper limit is not included in the interval per our convention, but becomes the lower limit (and included) of the next interval. This step is repeated until the last interval encloses the high value of the data.
5. Tally the frequency for each interval.
6. Record the frequency for each interval



Engineering Methods in Excel  
A SunCam online continuing education course

Cumulative Frequency

Another way of expressing frequency is the **cumulative frequency**. The cumulative frequency of an interval (or value) is the summation of the frequencies from the beginning of the distribution up to that interval (or value). The cumulative frequency table (or cumulative frequency distribution) of our sample speed interval data will therefore be as follows:

Speed (mph)	Frequency (f)	Cumulative Frequency (cf)
[20, 30)	2	2
[30, 40)	8	10
[40, 50)	30	40
[50, 60)	40	80
[60, 70)	15	95
[70, 80)	5	100
	$\Sigma f = 100$	

Note that the cumulative frequency of the last interval will always be equal to the total frequency.

Relative Frequency

The **relative frequency** of an interval (or value), is the frequency of the interval (or value), divided by the total frequency. Just like cumulative frequency, the **cumulative relative frequency** of an interval (or value) is the summation of the relative frequencies from the beginning of the distribution up to that interval (or value).

Class Mark

The class mark, or **class midpoint**, is the midpoint of an interval. It is the average of the lower limit and the upper limit of an interval (even though by our convention the upper limit value does not belong to the interval).

We may now update our distribution to show: the relative frequency, the cumulative relative frequency, and the class mark for each interval.





Engineering Methods in Excel  
A SunCam online continuing education course

Speed (mph)	Class Mark	Frequency (f)	Cumulative Frequency (cf)	Relative Frequency	Cumulative Relative Frequency
[20, 30)	25	2	2	0.02	0.02
[30, 40)	35	8	10	0.08	0.10
[40, 50)	45	30	40	0.30	0.40
[50, 60)	55	40	80	0.40	0.80
[60, 70)	65	15	95	0.15	0.95
[70, 80)	75	5	100	0.05	1.00
		$\Sigma f = 100$			

Note that the relative frequencies will always add up to unity; and the cumulative relative frequency of the last interval will be equal to unity.

## 2.4 Frequency Diagrams

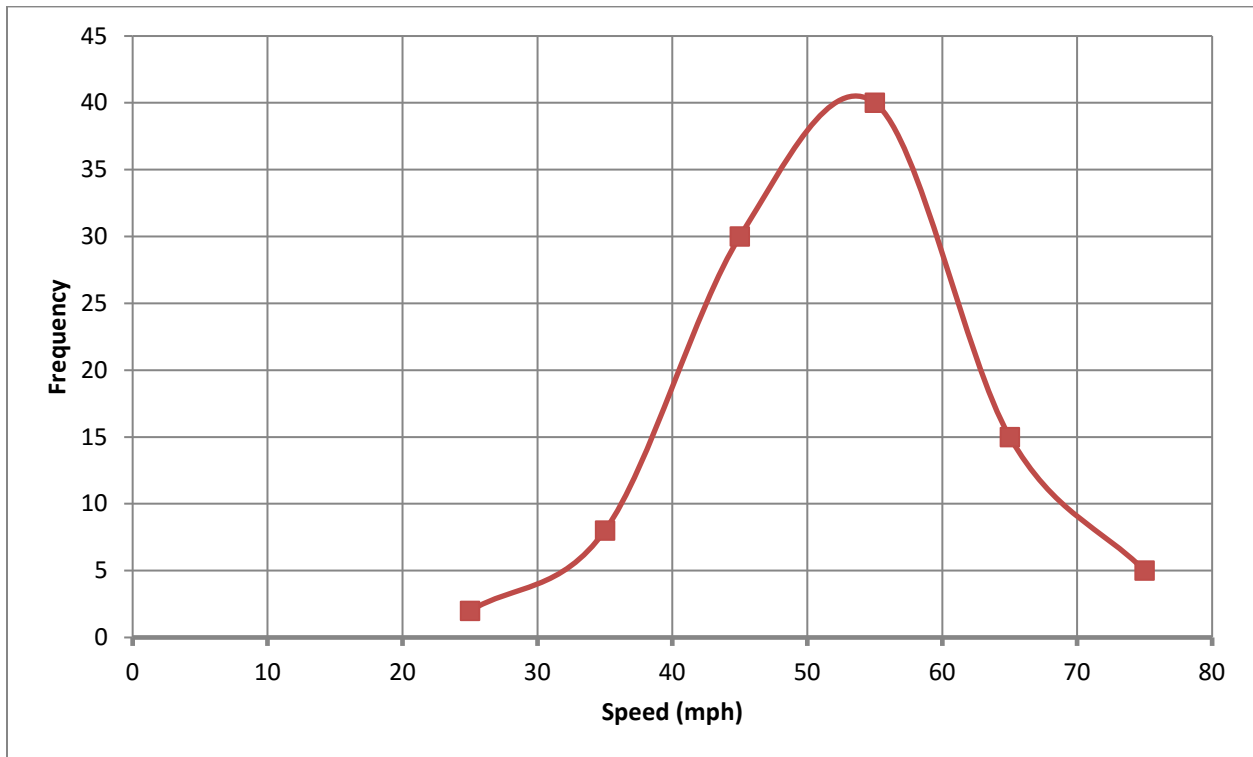
A frequency diagram is a graphical representation of a frequency distribution. There are many types of frequency diagrams in use. In this section, we shall review a selection of frequency diagrams commonly used by engineers.



Engineering Methods in Excel  
*A SunCam online continuing education course*

Frequency Polygon

A frequency polygon is a plot of frequency versus the data values. For grouped data, it is the frequency versus the class midpoint (class mark).

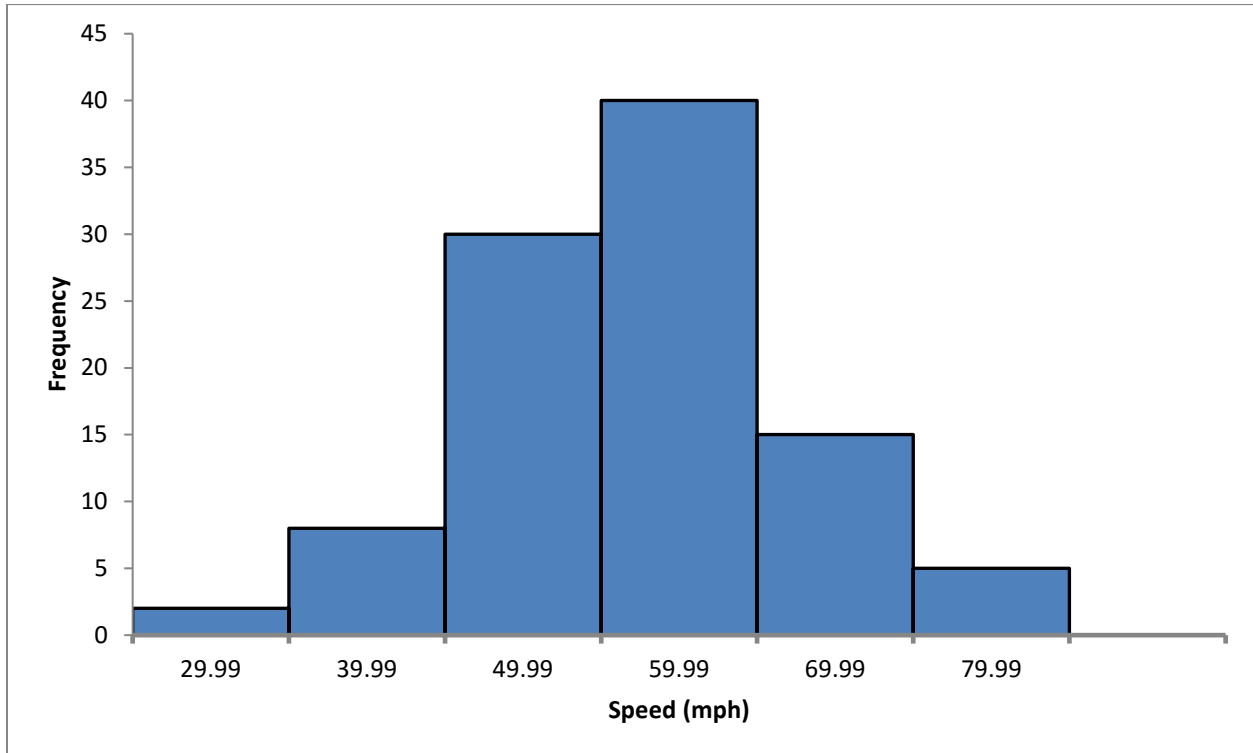




Engineering Methods in Excel  
*A SunCam online continuing education course*

Histogram

This is constructed by plotting a bar graph of frequency versus each interval.

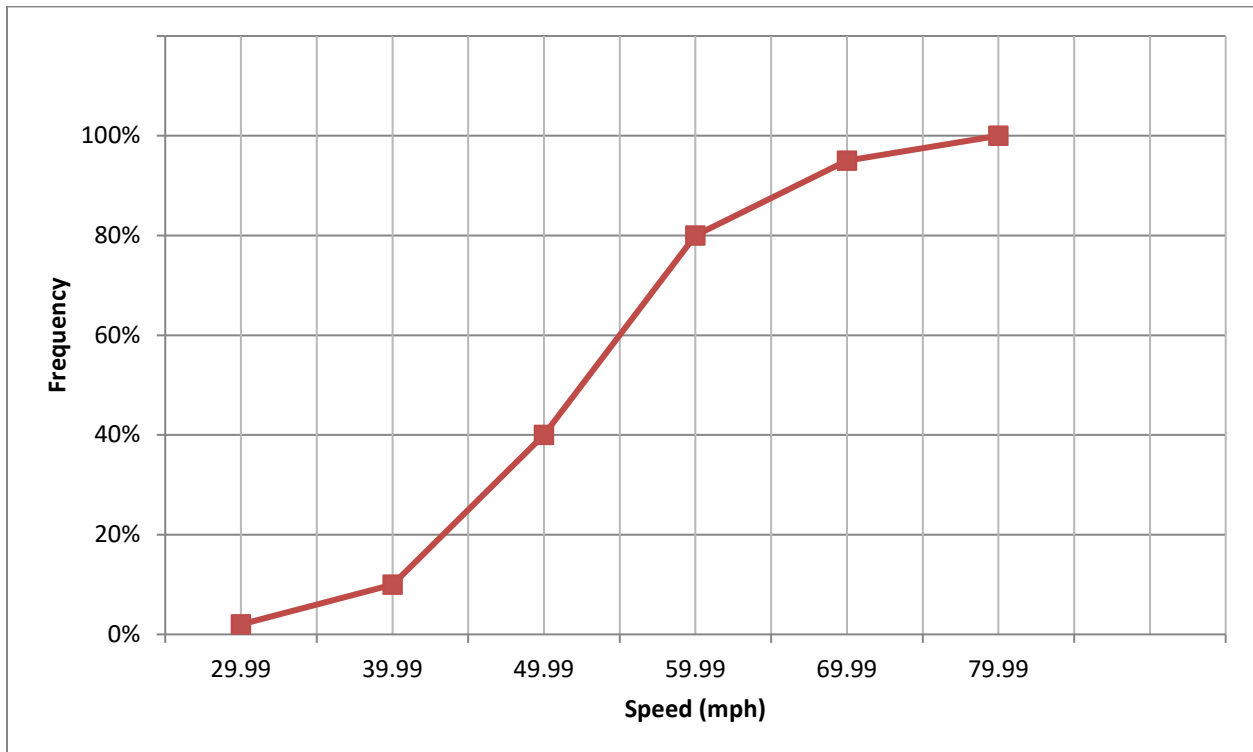




Engineering Methods in Excel  
A SunCam online continuing education course

Ogive

The ogive (“oh-jive”), or **cumulative frequency curve**, is a plot of cumulative relative frequency (in %) versus the class upper limits (or alternately, the class midpoints).



Note that an ogive based on class upper limits compared to an ogive based on class midpoints will have the exact same shape and values on the vertical axis; however, the latter will be shifted horizontally to the left—relative to the former.





Engineering Methods in Excel  
A SunCam online continuing education course

Stem and Leaf Diagram

In a stem-and-leaf plot, the data values are broken up such that all digits—except the last digit of the value—form a “stem”; and the last digit represents the data value in the “leaf”. Once a leaf is complete, the digits in the leaf are then rearranged in ascending numerical order. For grouped data, an alternative strategy is to have the stem formed by the intervals (classes); and each individual value is represented on the diagram by its last digit.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

Thus, in the example below, the arrows point to values 39, 66, and 75.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 (9)
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 (6) 9
7	0 2 2 (5) 9

We see that the stem-and-leaf diagram graphically depicts the entire data arranged in numerical order.



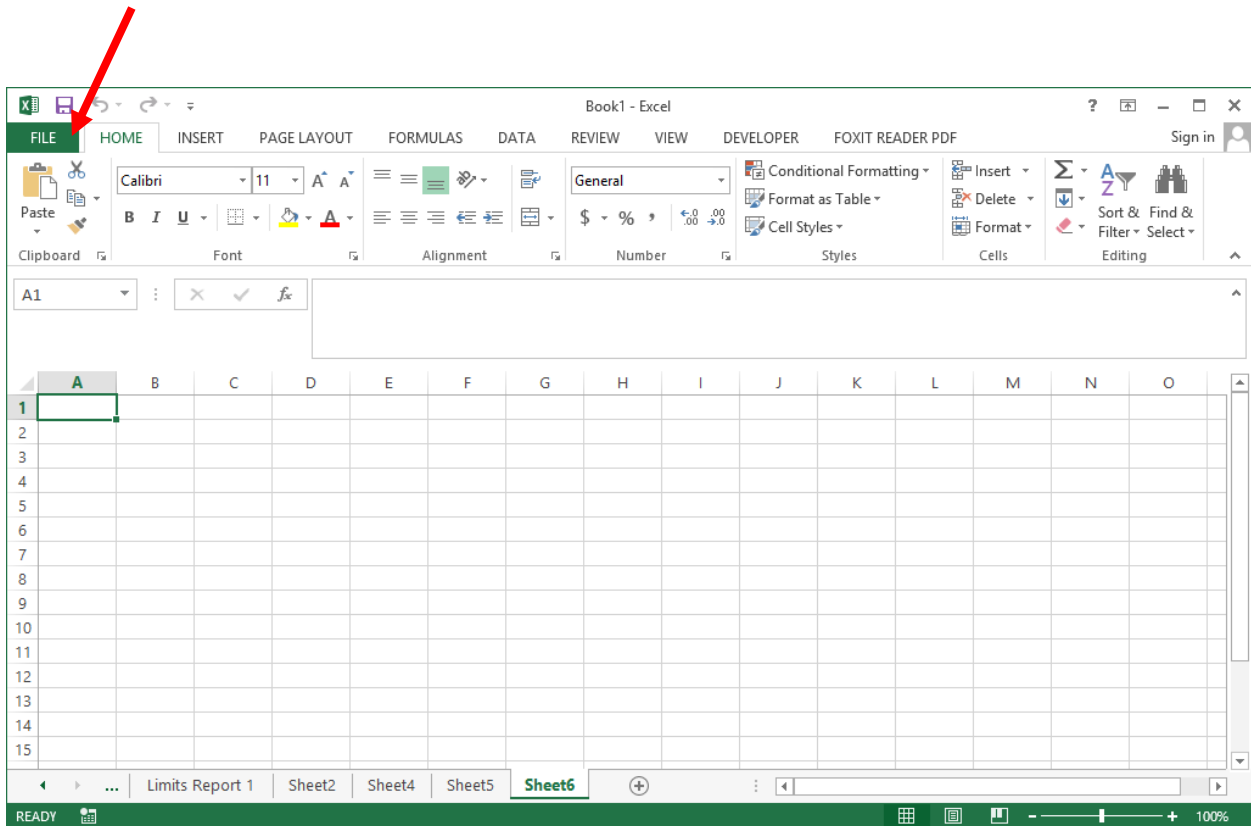
Engineering Methods in Excel  
A SunCam online continuing education course

### 2.5 Excel Data Analysis Tools

The *Data Analysis Tools* Add-in consists of a suite of tools for probability and statistics computations and analyses. As an Add-in, *Data Analysis Tools* is not readily available with a regular *Excel* install and must be specifically activated from the Add-in window. Once activated, however, *Data Analysis Tools* remains available for use; unless or until the user deactivates the add-in or resets their computer to a setting that predates the *Data Analysis Tools* activation.

Let us perform the following steps to activate *Data Analysis Tools*.

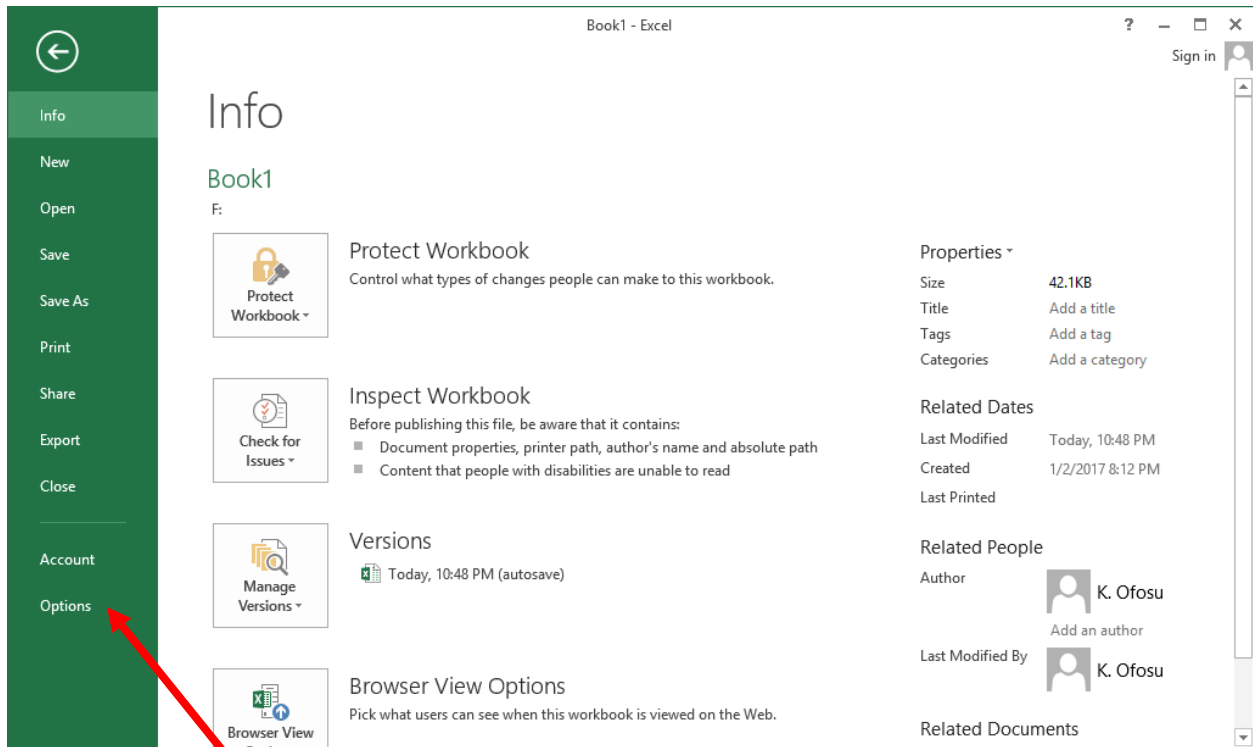
Click on **FILE**





Engineering Methods in Excel  
A SunCam online continuing education course

Click on **Options**

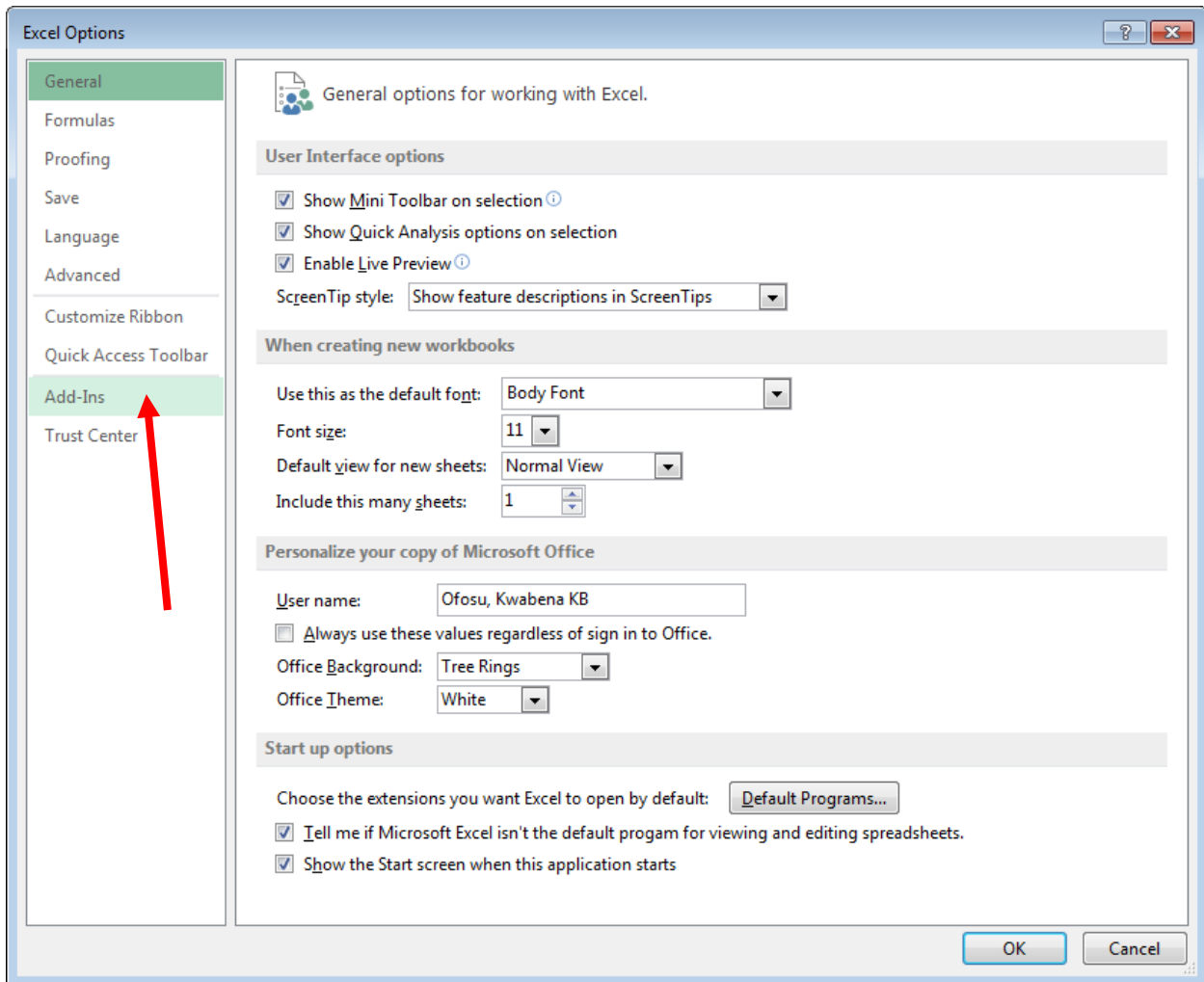






Engineering Methods in Excel  
A SunCam online continuing education course

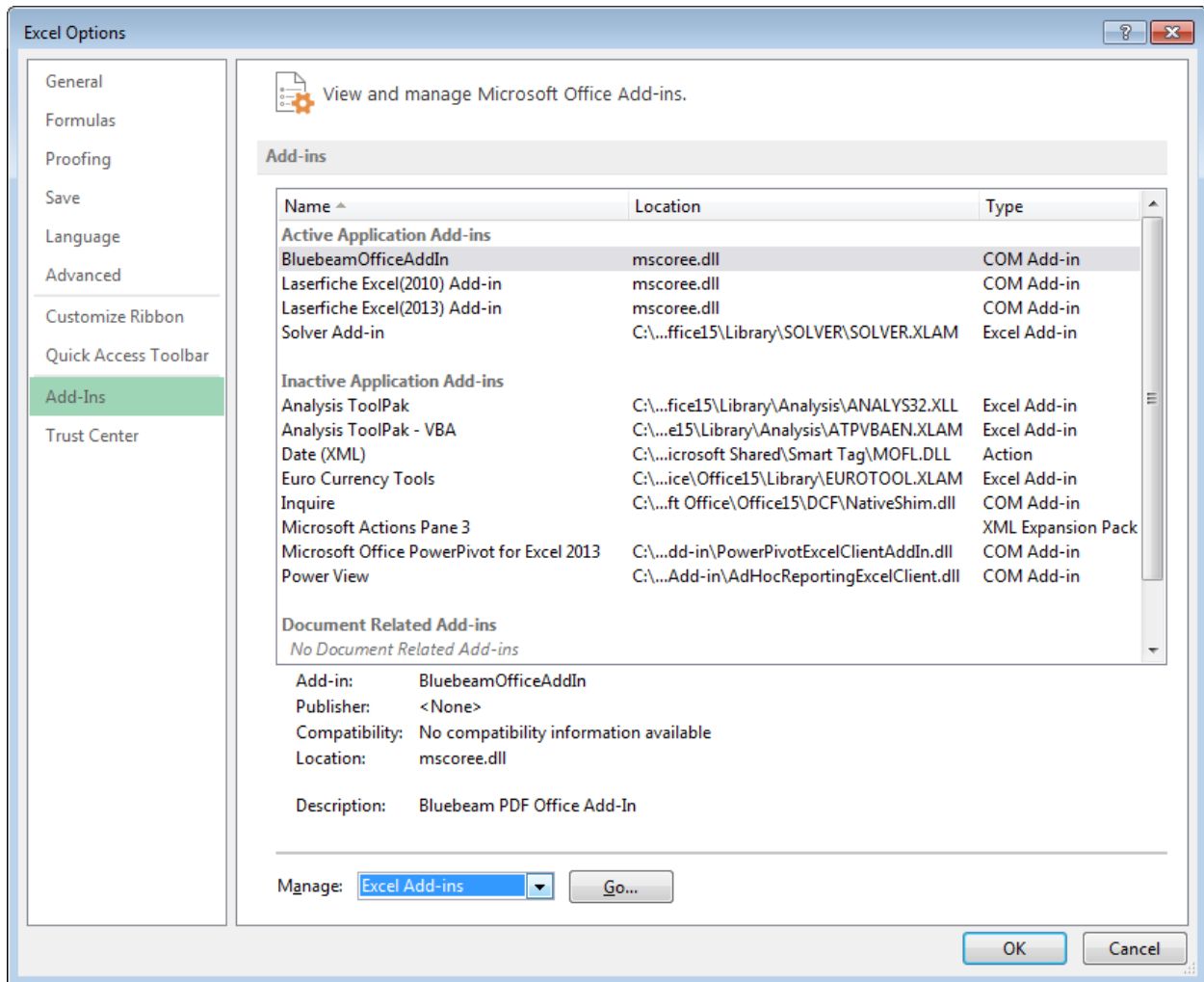
Click on **Add-Ins**.





Engineering Methods in Excel  
*A SunCam online continuing education course*

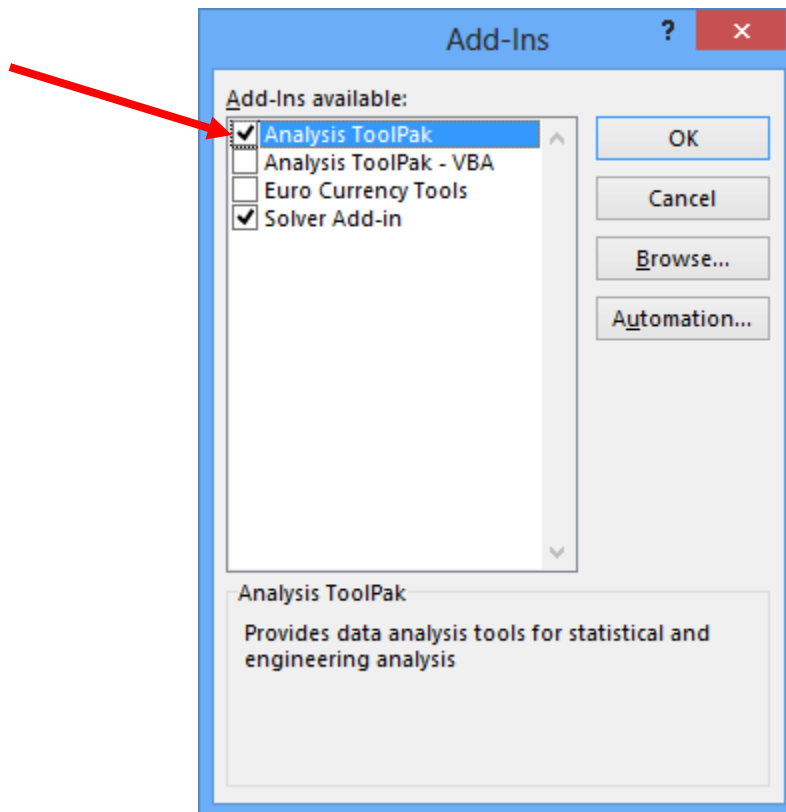
Under **Manage**, select **Excel-Add-Ins**, and hit **Go**





Engineering Methods in Excel  
*A SunCam online continuing education course*

Under **Add-Ins available:**, check the **Analysis ToolPak** add-In to activate it.



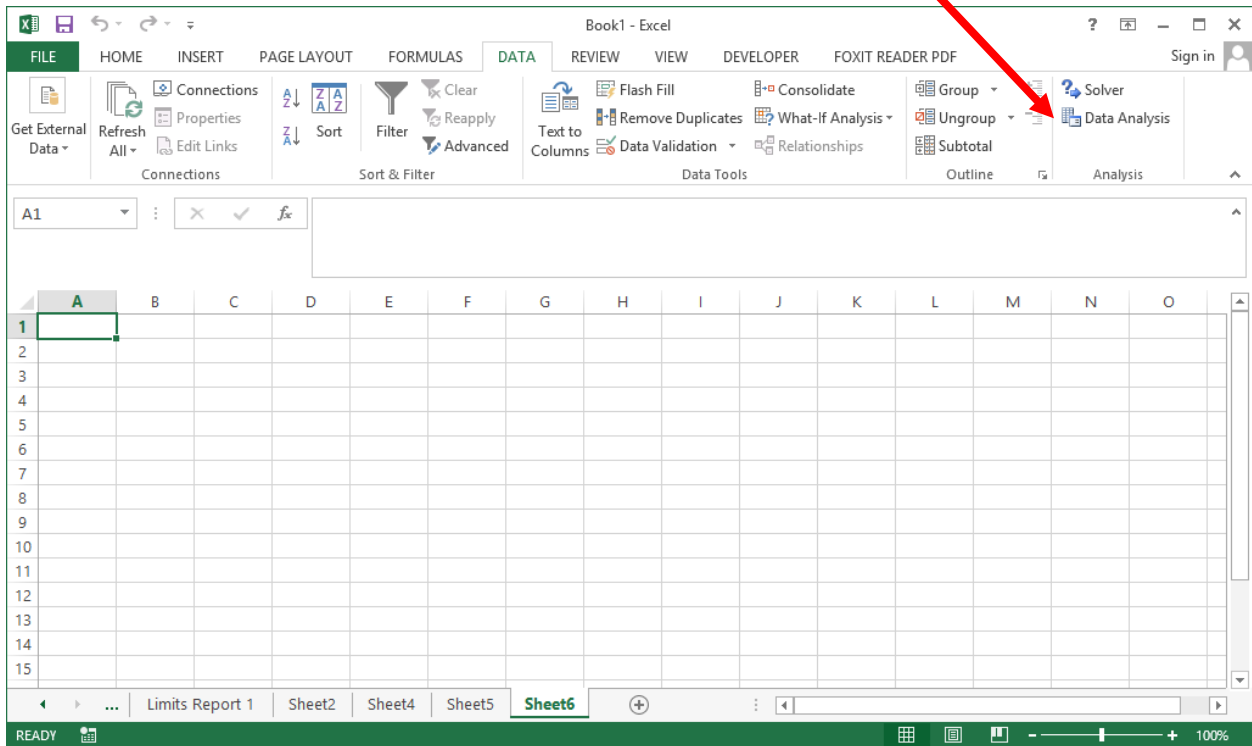
Hit **OK**.

The Add-Ins window is dismissed.



Engineering Methods in Excel  
*A SunCam online continuing education course*

Return to your main *Excel* window and go to the **DATA** tab.  
Notice that the **Analysis** palette now has the **Data Analysis Tools** icon.

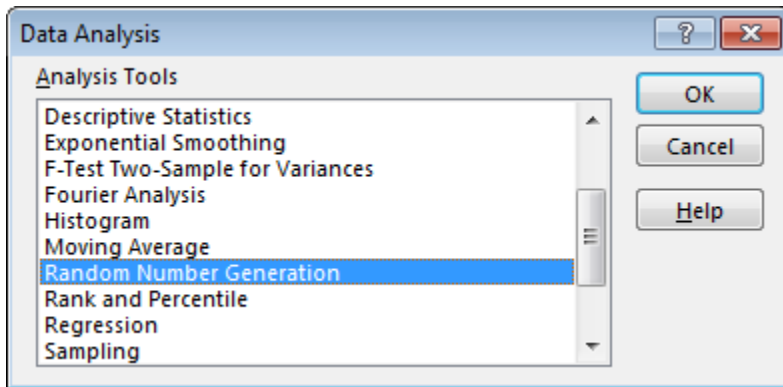




Engineering Methods in Excel  
*A SunCam online continuing education course*

Click on **Data Analysis Tools**.

In the **Data Analysis** window review the various statistical tools available. Use the scrollbar to review tools towards the bottom of the list.



Click on **Cancel** to close the **Data Analysis** window.

## 2.6 Data Analysis in *Excel*

In this section, we shall work through a real-life example with the *Excel Data Analysis Tools*—using the speed data provided with your course packet.

Open your speed data file and review the data which consists of 100 vehicles speeds measured on a rural highway by a municipal traffic engineer.



Engineering Methods in Excel  
A SunCam online continuing education course

The screenshot shows the Microsoft Excel interface. The ribbon includes FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, DEVELOPER, BLUEBEAM, and LASERFICHE. The HOME ribbon is active, showing options for Clipboard, Font, Alignment, Number, Styles, Cells, and Editing. The spreadsheet has a column labeled 'Speed (mph)' in row 1. The data in this column is as follows:

Row	Speed (mph)
2	52
3	51
4	53
5	54
6	50
7	55
8	57
9	53
10	55
11	56
12	58
13	55
14	59
15	58
16	55
17	55
18	52
19	54
20	55
21	55
22	42
23	63
24	35
25	44

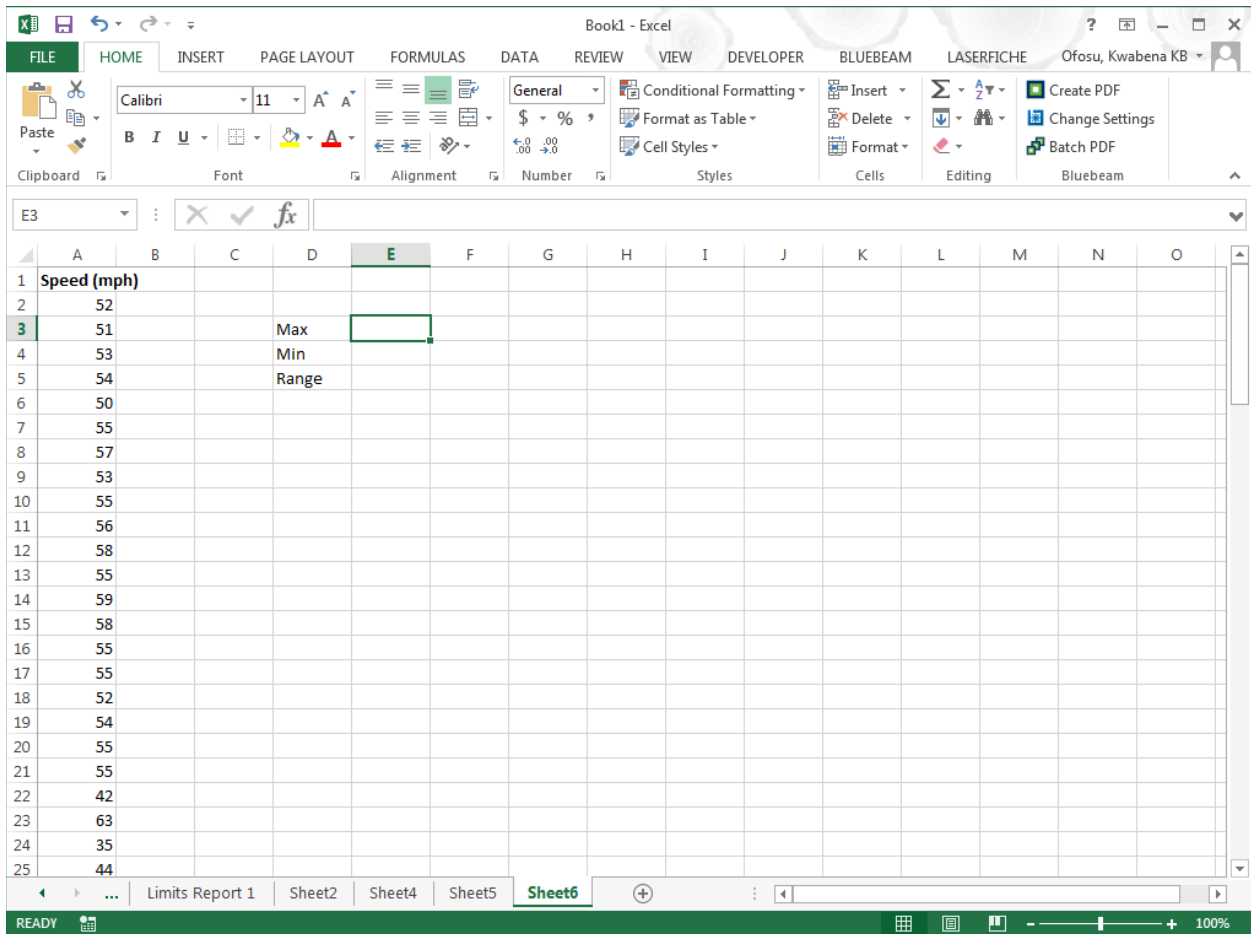
The status bar at the bottom shows 'READY' and a zoom level of 100%.



Engineering Methods in Excel  
A SunCam online continuing education course

Let us go through the process to set up the frequency distribution (or frequency table):

1. Compute the range of the data—which is the difference of the high value and the low value. Select cells to compute the maximum value and the minimum value in the data.





Engineering Methods in Excel  
A SunCam online continuing education course

For the maximum value type “=Max(“

The screenshot shows the Microsoft Excel interface. The ribbon is set to the 'FORMULAS' tab. The formula bar at the top displays the formula `=Max(`. The spreadsheet contains the following data:

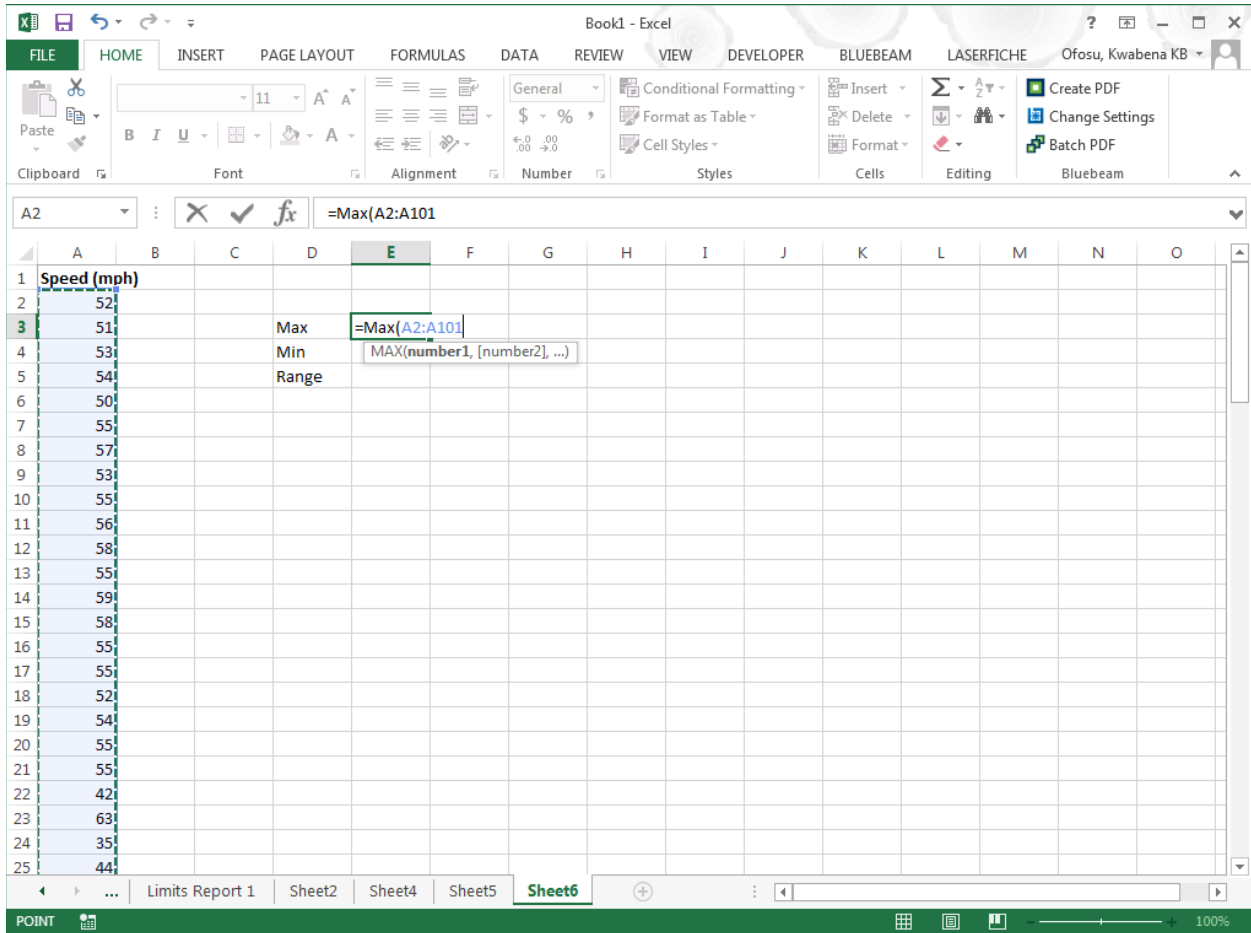
Row	Column A	Column B	Column C	Column D	Column E	Column F	Column G	Column H	Column I	Column J	Column K	Column L	Column M	Column N	Column O
1	Speed (mph)														
2		52													
3		51		Max	=Max(										
4		53		Min	MAX(number1, [number2], ...)										
5		54		Range											
6		50													
7		55													
8		57													
9		53													
10		55													
11		56													
12		58													
13		55													
14		59													
15		58													
16		55													
17		55													
18		52													
19		54													
20		55													
21		55													
22		42													
23		63													
24		35													
25		44													





Engineering Methods in Excel  
A SunCam online continuing education course

Select the data, or type the range of cells of the data in the formula bar.



Hit **Enter** on your keyboard.



Engineering Methods in Excel  
A SunCam online continuing education course

The maximum value in the data is displayed.

The screenshot shows an Excel spreadsheet with the following data in column A:

Speed (mph)
52
51
53
54
50
55
57
53
55
56
58
55
59
58
55
52
54
55
42
63
35
44

In column D, the following text is present:

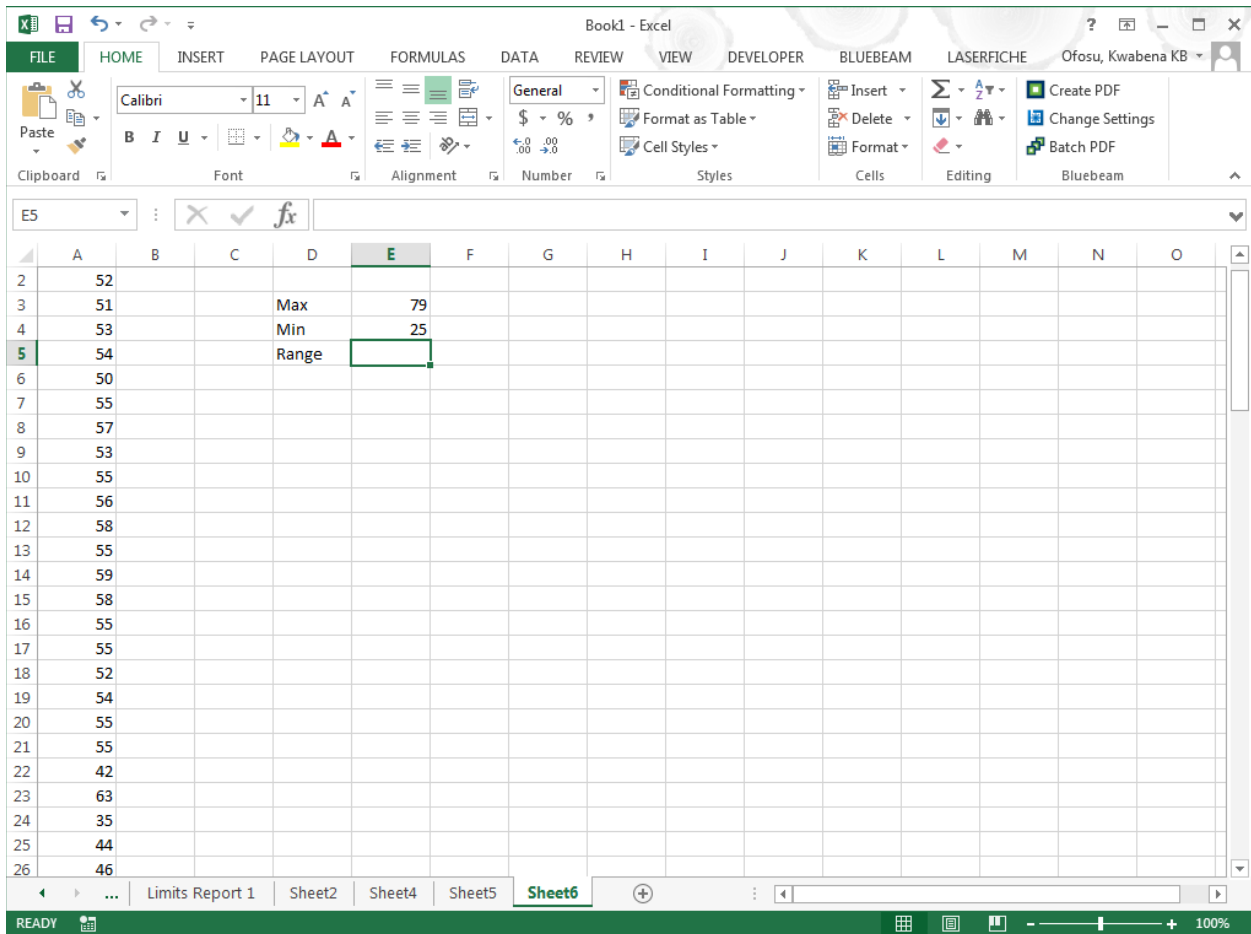
Max
Min
Range

In cell E3, the value 79 is displayed, representing the maximum value of the data set. Cell E4 is currently empty and selected.



Engineering Methods in Excel  
A SunCam online continuing education course

At the cell location for the minimum value, type “=Min(“  
Select the data, or type the range of cells of the data in the formula bar.  
Hit **Enter** on your keyboard.  
The minimum value in the data is displayed.





Engineering Methods in Excel  
A SunCam online continuing education course

The Range of the data is the maximum value minus the minimum value.

By point-and click,

The screenshot shows an Excel spreadsheet with the following data in column A:

Row	Value
2	52
3	51
4	53
5	54
6	50
7	55
8	57
9	53
10	55
11	56
12	58
13	55
14	59
15	58
16	55
17	55
18	52
19	54
20	55
21	55
22	42
23	63
24	35
25	44
26	46

In column D, the following statistics are calculated:

Label	Value
Max	79
Min	25
Range	=E3-E4

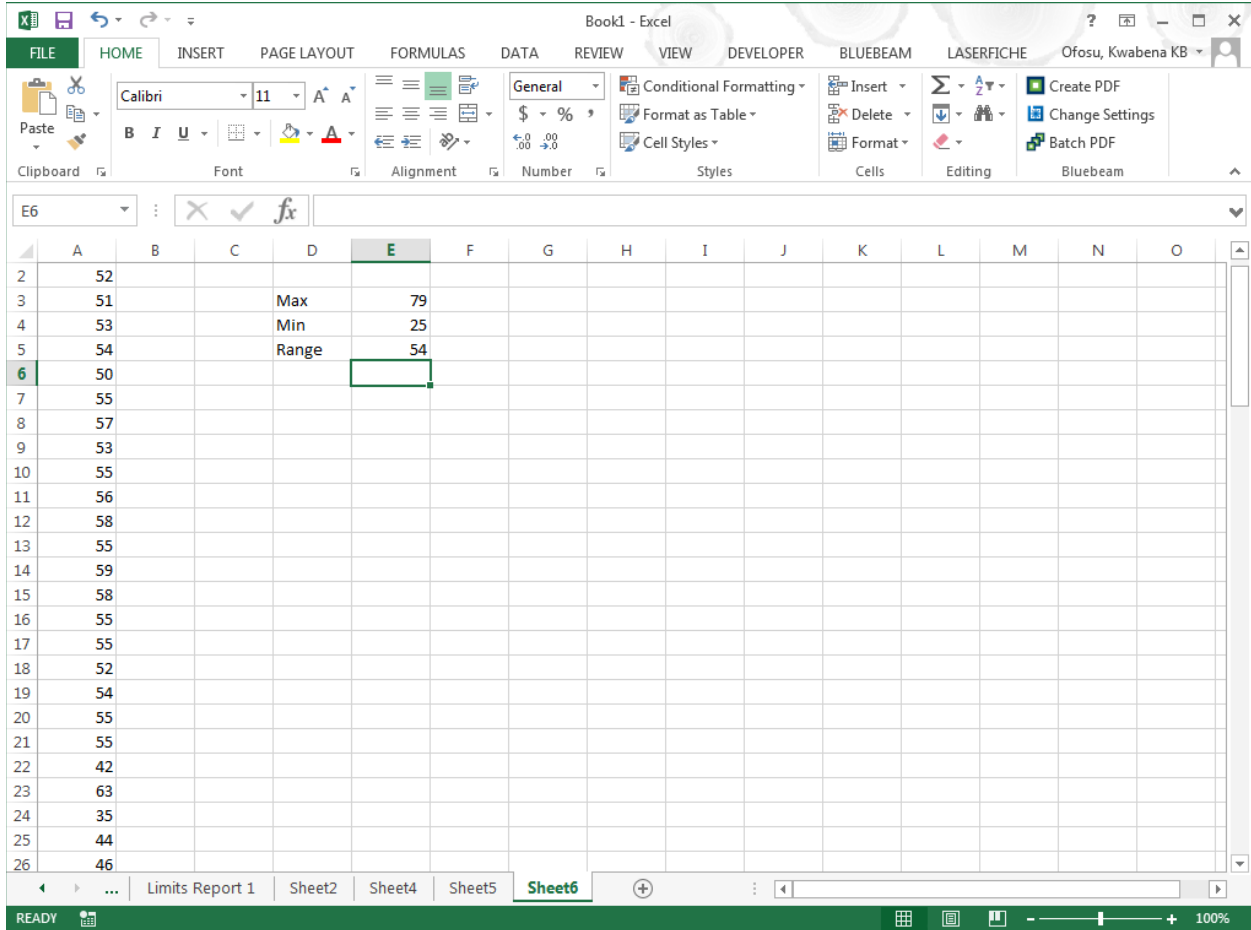
The formula bar shows the active cell E4 contains the formula  $=E3-E4$ .



Engineering Methods in Excel  
A SunCam online continuing education course

Hit **Enter** on your keyboard.

The range of the data is computed.

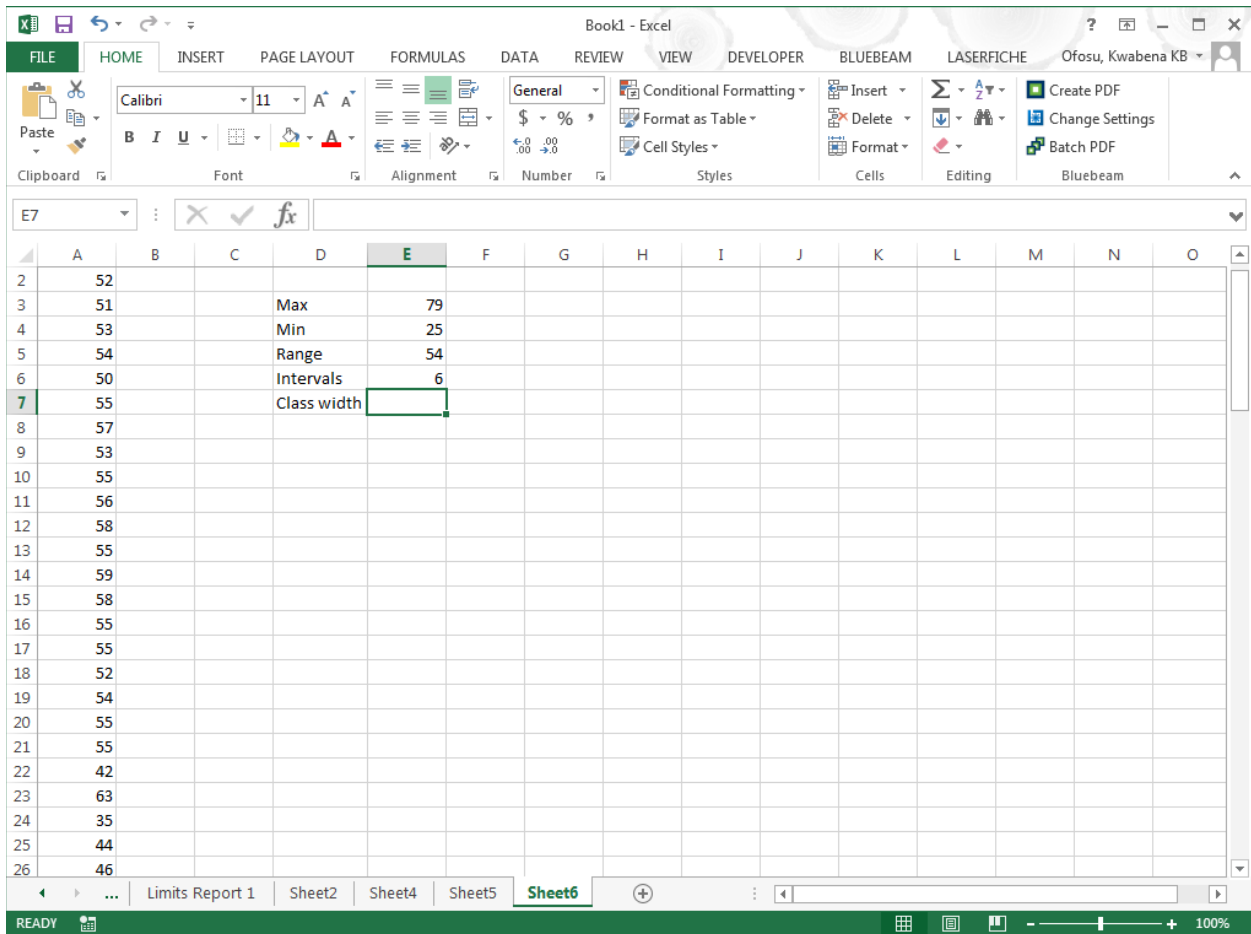




Engineering Methods in Excel  
A SunCam online continuing education course

2. Select the number of intervals to be used—typically between 5 and 20, depending on the nature of data.

Let us pick 6 intervals to use for our data.





Engineering Methods in Excel  
A SunCam online continuing education course

3. Compute the interval width—which is the range divided by the number of intervals—and round to a convenient number to work with; typically, a multiple of 5, 10, 20, 50, 100 etc., depending on the nature of the data.

Enter the division formula by point-and-click.

The screenshot shows an Excel spreadsheet with the following data in columns A through E:

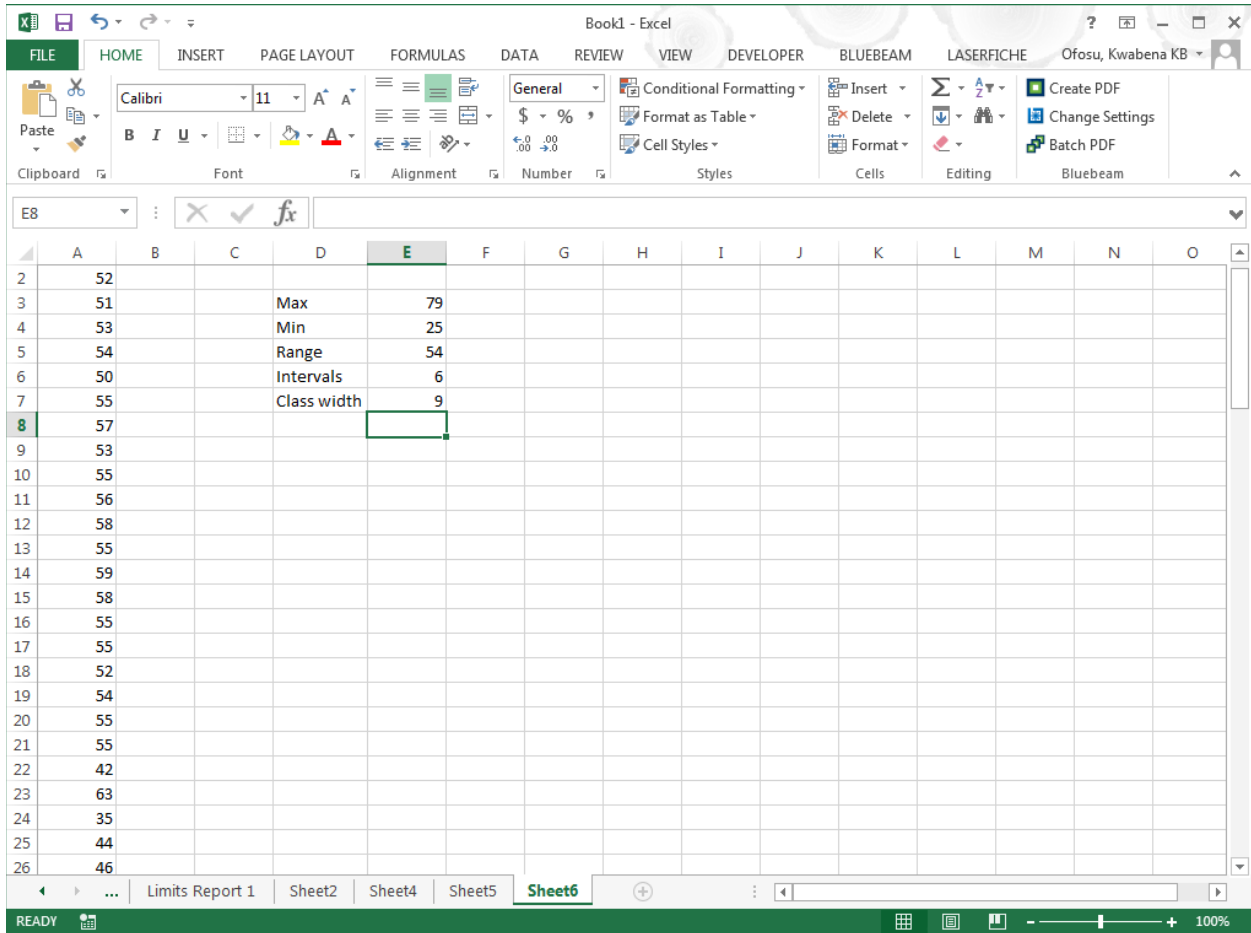
	A	B	C	D	E
2	52				
3	51			Max	79
4	53			Min	25
5	54			Range	54
6	50			Intervals	6
7	55			Class width	=E5/E6
8	57				
9	53				
10	55				
11	56				
12	58				
13	55				
14	59				
15	58				
16	55				
17	55				
18	52				
19	54				
20	55				
21	55				
22	42				
23	63				
24	35				
25	44				
26	46				

The formula bar shows the active cell E6 containing the formula  $=E5/E6$ . The spreadsheet interface includes the ribbon (FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, DEVELOPER, BLUEBEAM, LASERFICHE) and the status bar at the bottom (POINT, 100%).



Engineering Methods in Excel  
*A SunCam online continuing education course*

**Hit Enter**



So, for convenience, we shall round up and use a class width of 10. Note that this rounding may result in a change in the number of intervals than originally selected, but that is fine; it will have no adverse impacts on our analysis.



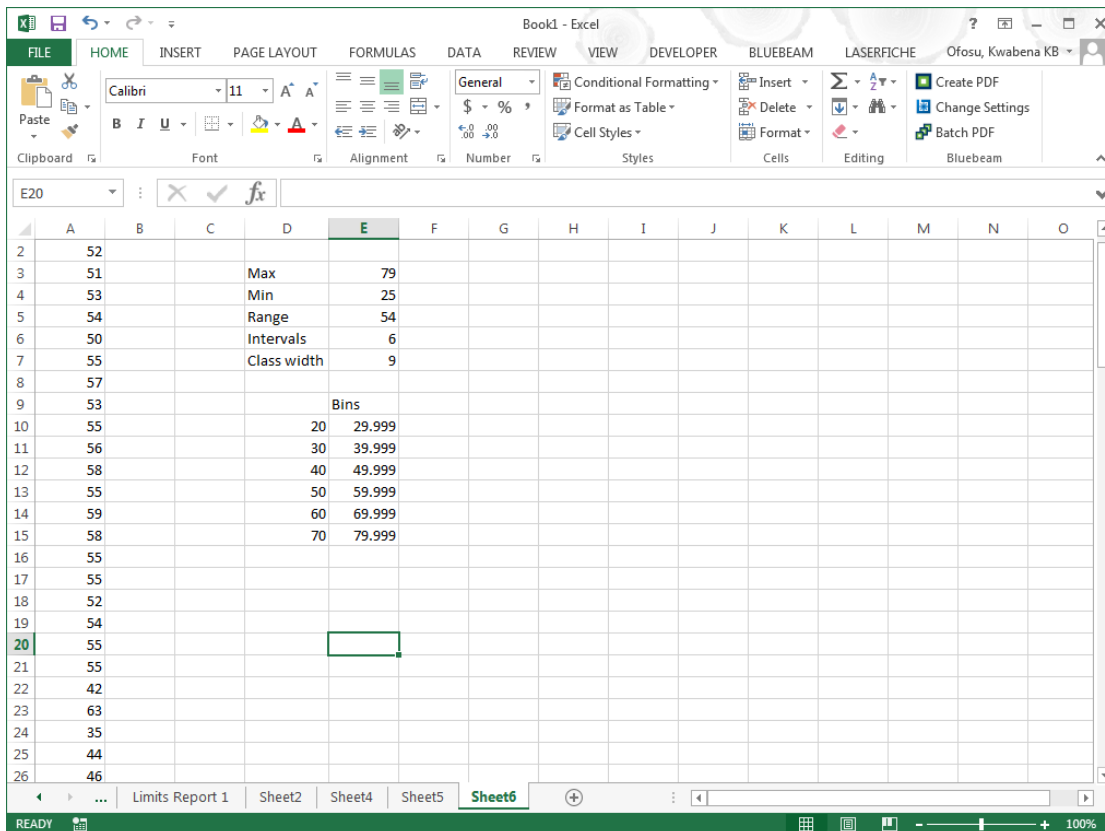


Engineering Methods in Excel  
A SunCam online continuing education course

4. Select the lower limit of the first interval so that it encloses the low value of the data. The upper limit of the interval is the lower limit plus the interval width. The upper limit is not included in the interval per our convention; but becomes the lower limit (and included) of the next interval. This step is repeated such that the upper limit of the last interval encloses the high value of the data.

For our data, let us pick a lower limit of 20 for the first interval. The upper limit will therefore be 30. However, based on our  $[a, b)$  convention, the first interval will be  $[20, 30)$ —where the value 30 is not included in the interval. In other words, our interval is effectively from 20 through 29.999999... We shall use this format of the upper limit of the intervals in the *Excel* manipulations.

The last interval is therefore  $[70, 80)$ —which is effectively 70 through 79.999999... In *Excel*, the intervals are referred to as **bins**. Our intervals (or bins) for *Excel* purposes are therefore:



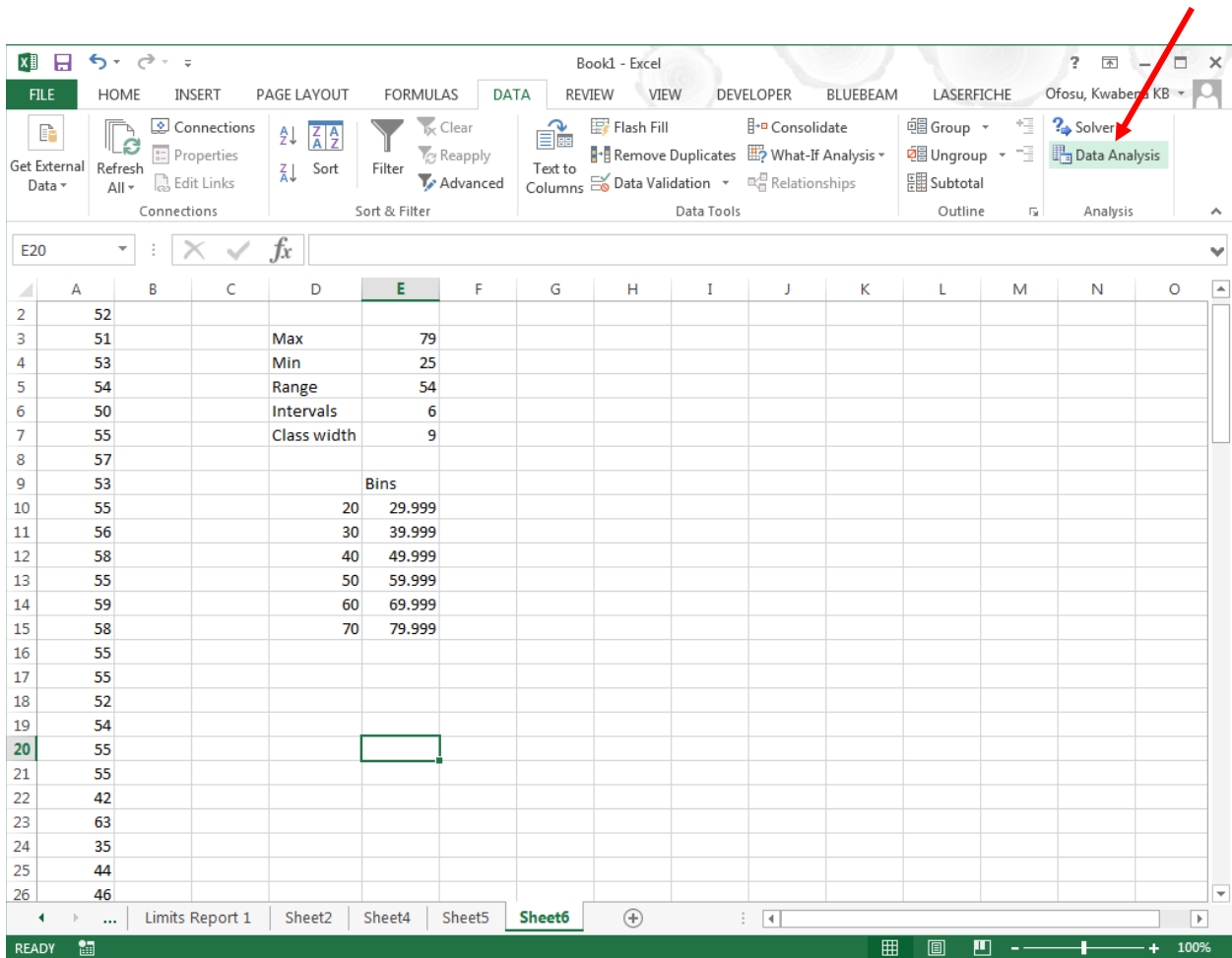


Engineering Methods in Excel  
A SunCam online continuing education course

5. Tally the frequency for each interval.
6. Record the frequency for each interval

We shall use the *Excel Data Analysis Tools* to tally and record the frequency of each interval (bin).

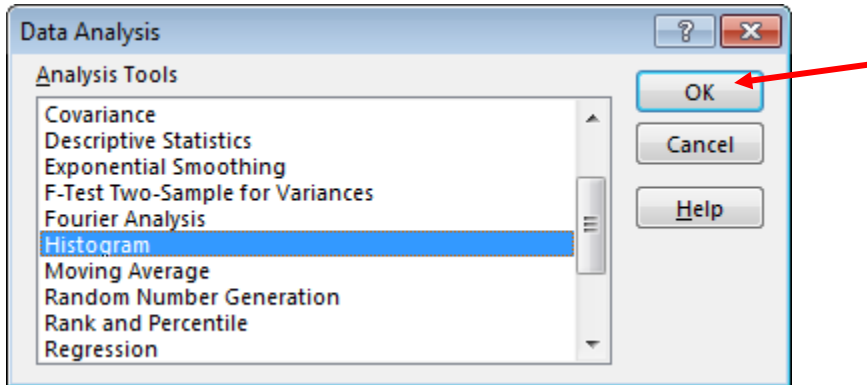
Click on the **DATA** tab.  
Click on **Data Analysis**.



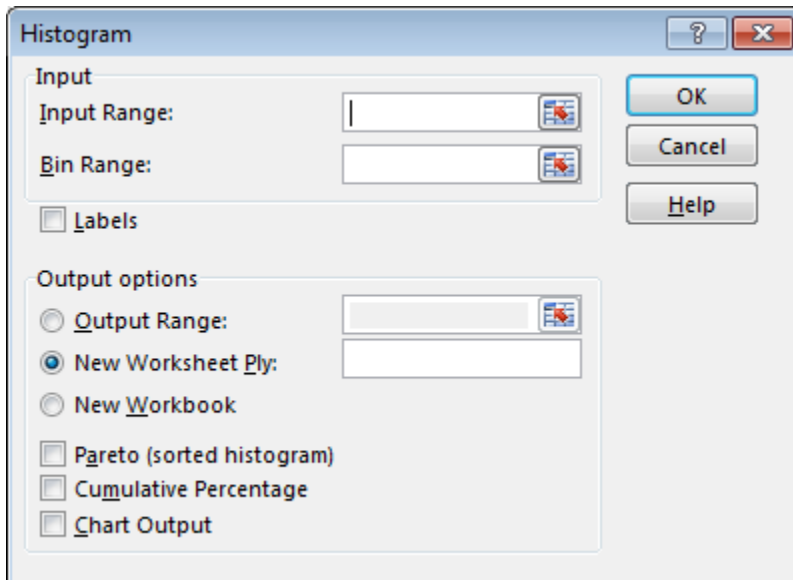


Engineering Methods in Excel  
A SunCam online continuing education course

In the **Data Analysis** window, select **Histogram**.  
Press **OK**.



The **Histogram** window opens.





Engineering Methods in Excel  
A SunCam online continuing education course

With the cursor flashing in the **Input Range:** textbox, go to the spreadsheet and select your data. (Alternately, type in the range of cells holding the data directly into the **Input Range:** textbox).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2	52														
3	51			Max	79										
4	53			Min	25										
5	54			Range	54										
6	50			Intervals	6										
7	55			Class width	9										
8	57														
9	53			Bins											
10	55			20	29.999										
11	56			30	39.999										
12	58			40	49.999										
13	55			50	59.999										
14	59			60	69.999										
15	58			70	79.999										
16	55														
17	55														
18	52														
19	54														
20	55														
21	55														
22	42														



Engineering Methods in Excel  
A SunCam online continuing education course

Click in the **Bin Range:** textbox.

Go to the spreadsheet and select the bins. The upper limits only are needed.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2	52														
3	51			Max	79										
4	53			Min	25										
5	54			Range	54										
6	50			Intervals	6										
7	55			Class width	9										
8	57														
9	53				Bins										
10	55			20	29.999										
11	56			30	39.999										
12	58			40	49.999										
13	55			50	59.999										
14	59			60	69.999										
15	58			70	79.999										
16	55														
17	55														
18	52														
19	54														
20	55														
21	55														
22	55														



Engineering Methods in Excel  
A SunCam online continuing education course

For the **Output options**, we would like to place all outputs on a range of cells on the current worksheet.

Select the option button for the **Output Range**:.  
Click in the Output Range textbox.

Select a range of cells on the current worksheet where you want your results to be posted.

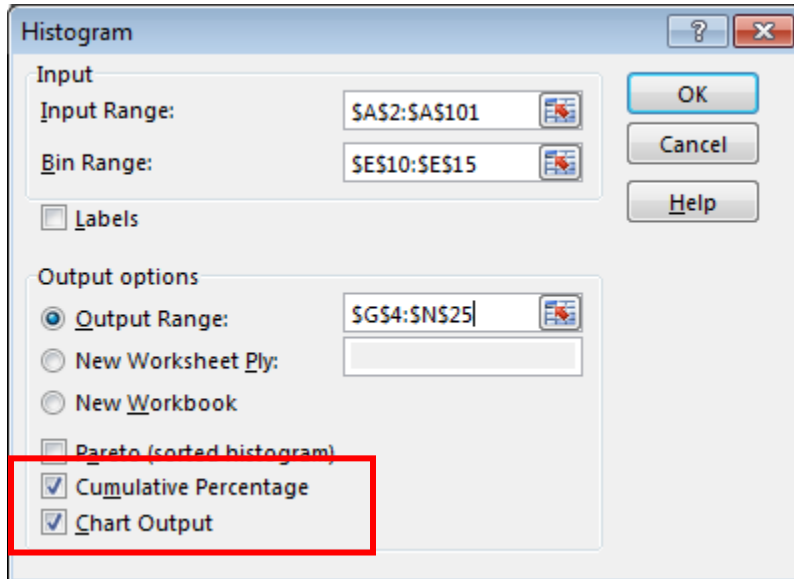
The screenshot shows the Histogram dialog box in Excel. The 'Input' section has 'Input Range' set to '\$A\$2:\$A\$101' and 'Bin Range' set to '\$E\$10:\$E\$15'. The 'Output options' section has 'Output Range' set to '\$G\$4:\$N\$25', which is highlighted with a dashed green box on the worksheet. Other options include 'Labels' (unchecked), 'New Worksheet Ply' (radio button), 'New Workbook' (radio button), 'Pareto (sorted histogram)' (checkbox), 'Cumulative Percentage' (checked), and 'Chart Output' (checked). The worksheet data is as follows:

	53			Bins	
10	55			20	29.999
11	56			30	39.999
12	58			40	49.999
13	55			50	59.999
14	59			60	69.999
15	58			70	79.999
16	55				
17	55				
18	52				
19	54				
20	55				
21	55				
22	42				
23	63				
24	35				
25	44				



Engineering Methods in Excel  
*A SunCam online continuing education course*

We want a **Chart Output** and a **Cumulative Frequency (in percentage)** to be recorded and displayed on the chart.

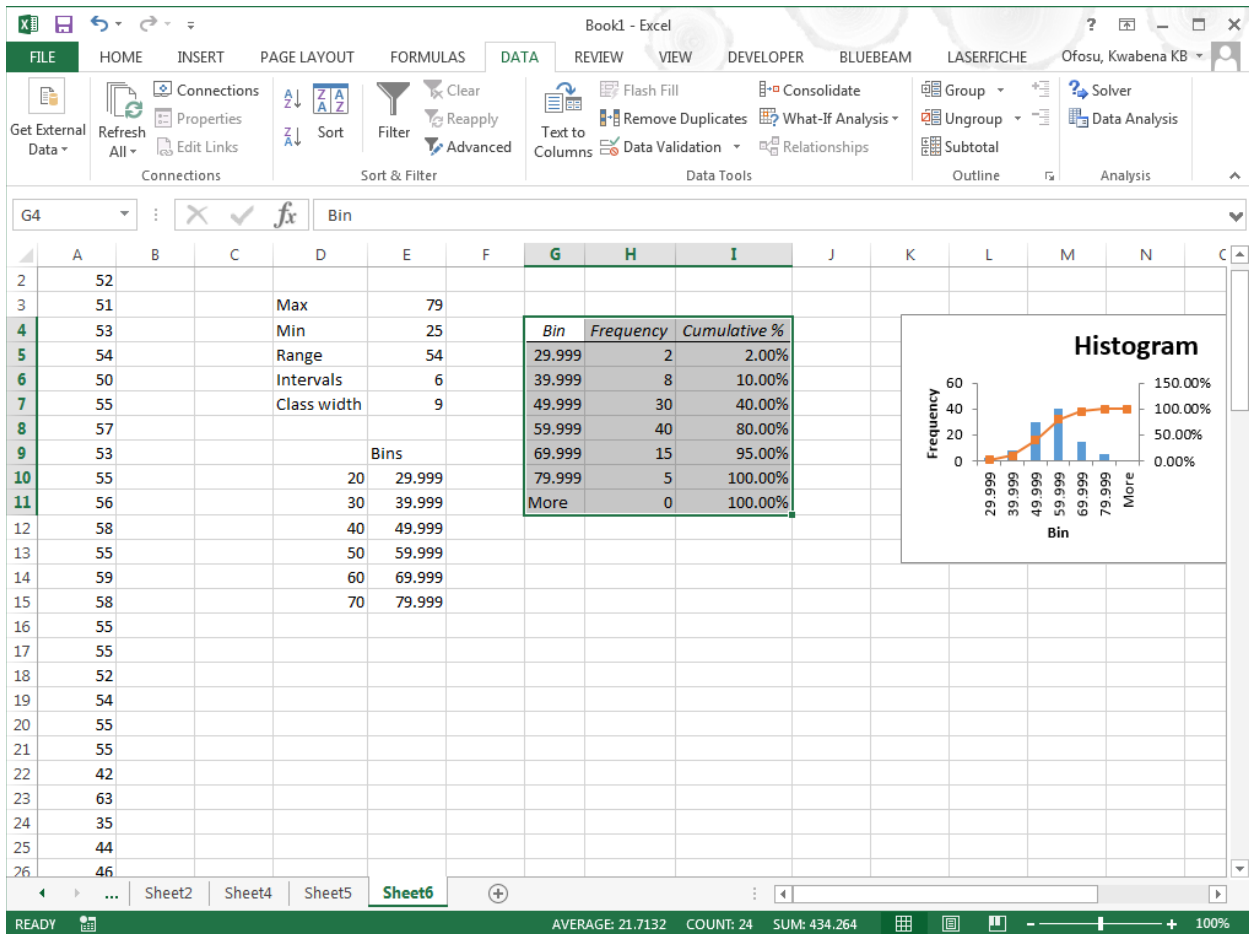




Engineering Methods in Excel  
A SunCam online continuing education course

Hit OK.

The frequency table and the histogram, with an ogive superimposed on it, are displayed.







Engineering Methods in Excel  
A SunCam online continuing education course

Use **Copy** and **Paste** on the **HOME** tab to make a copy of the chart.

The screenshot shows an Excel spreadsheet with a histogram and a data table. The histogram is titled "Histogram" and displays Frequency (blue bars) and Cumulative % (orange line) for various bins. The data table is as follows:

Bin	Frequency	Cumulative %
29.999	2	2.00%
39.999	8	10.00%
49.999	30	40.00%
59.999	40	80.00%
69.999	15	95.00%
79.999	5	100.00%
More	0	100.00%

The spreadsheet also contains summary statistics in cells D3-D8:

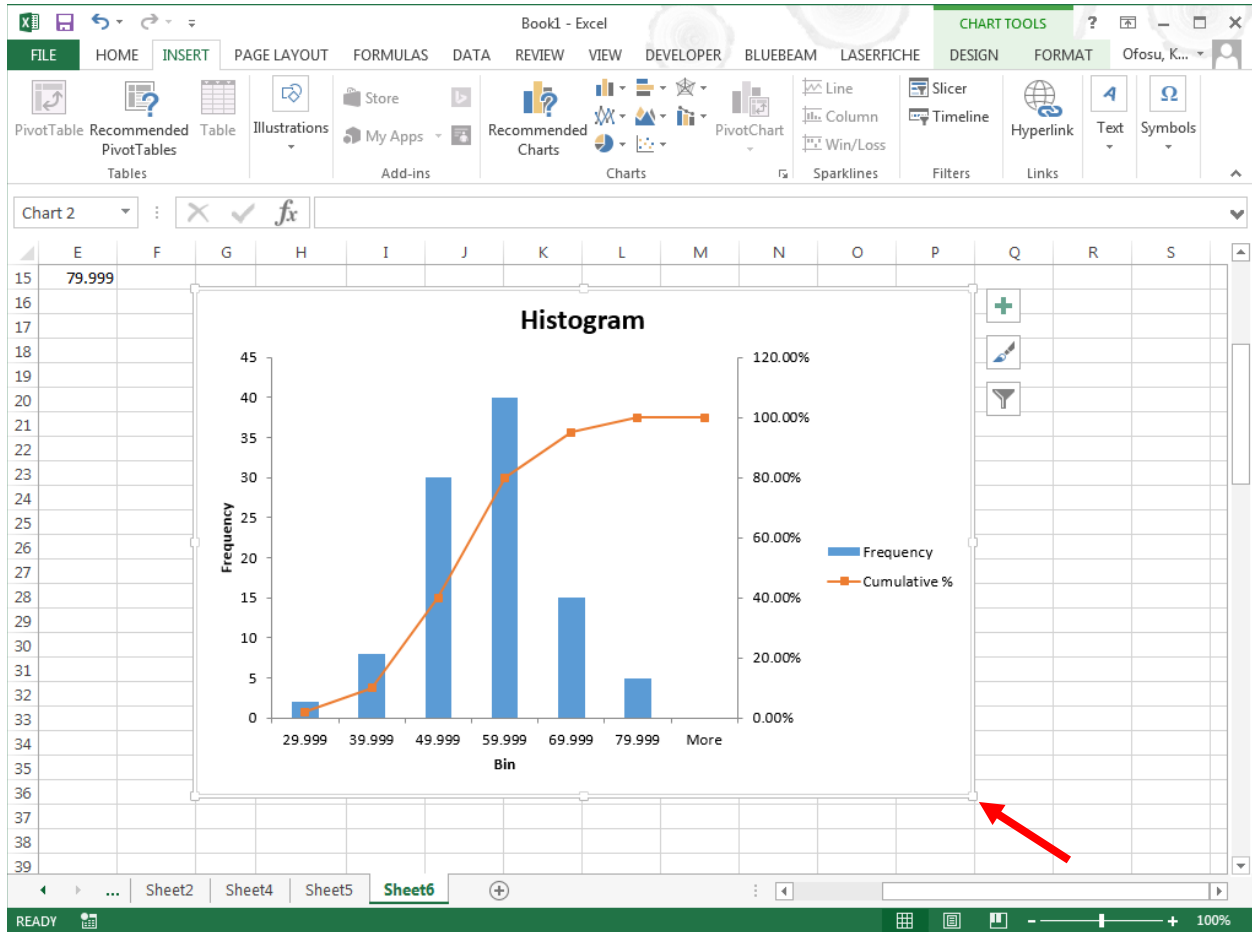
Statistic	Value
Max	79
Min	25
Range	54
Intervals	6
Class width	9

The histogram shows a distribution of data points across bins, with the highest frequency occurring in the 49.999-59.999 bin. The cumulative percentage line shows the total percentage of data points up to each bin.



Engineering Methods in Excel  
A SunCam online continuing education course

On the copied chart click and hold down the cursor on the bottom right corner.  
Drag the cursor to resize the chart.

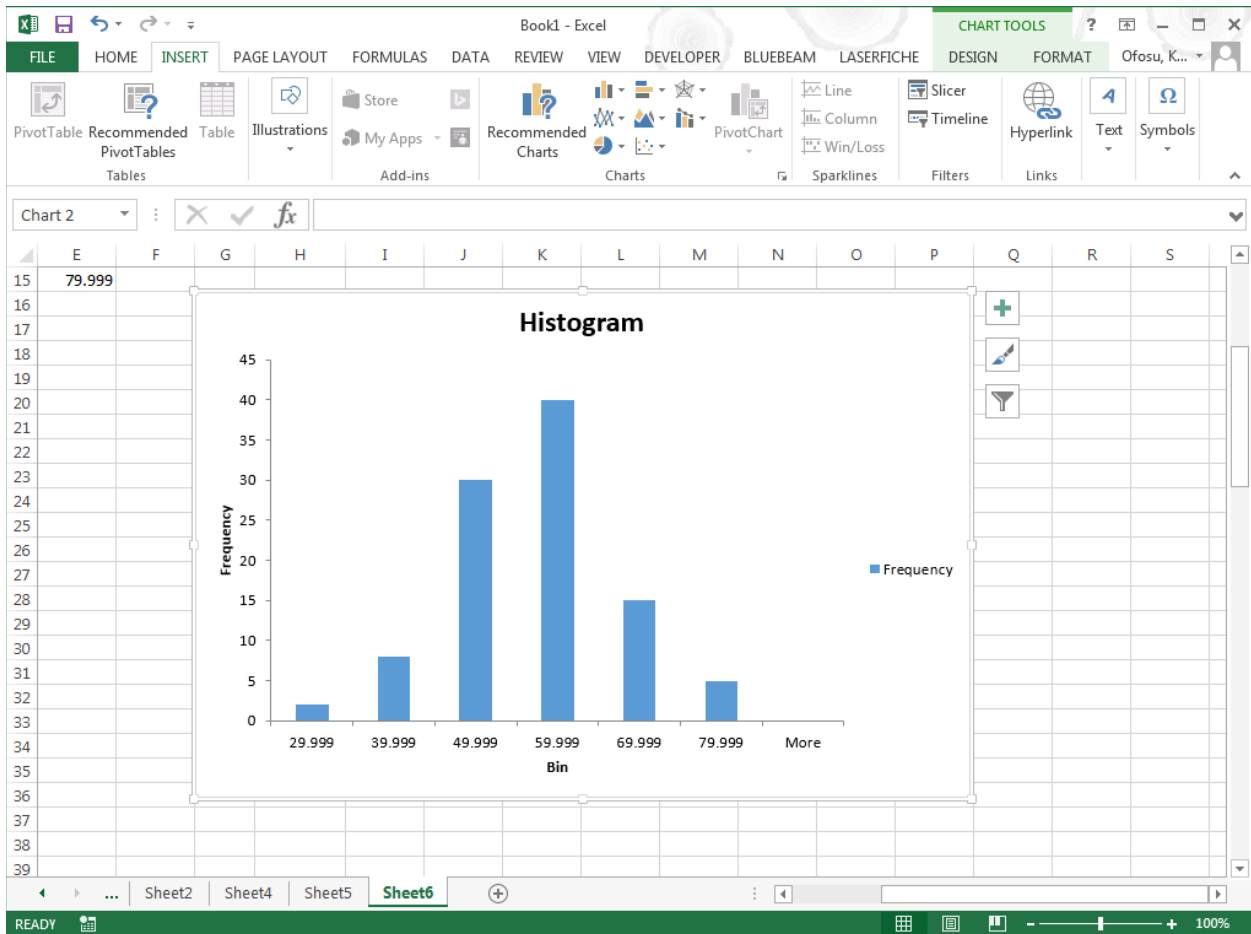


We shall perform some manipulations on the chart to “clean it up”. These manipulations apply to all types of charts in *Excel* and not specifically to charts created via the *Data Analysis Tools*.



Engineering Methods in Excel  
A SunCam online continuing education course

On the copied chart, we would like to focus on the histogram only.  
Click anywhere on the ogive to select it.  
Hit **Delete** on your keyboard.

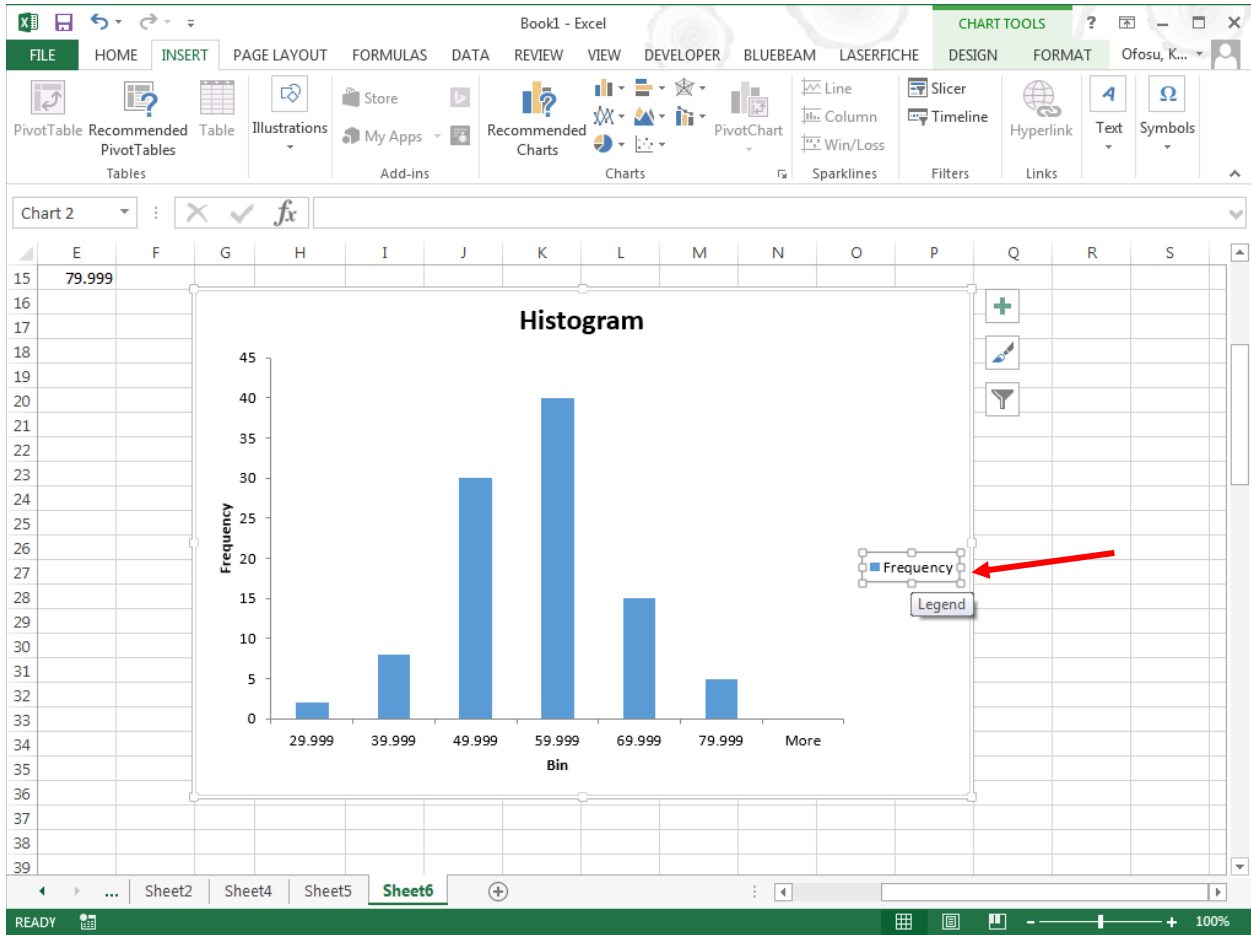


Notice the ogive and its associated vertical axis are deleted.



Engineering Methods in Excel  
A SunCam online continuing education course

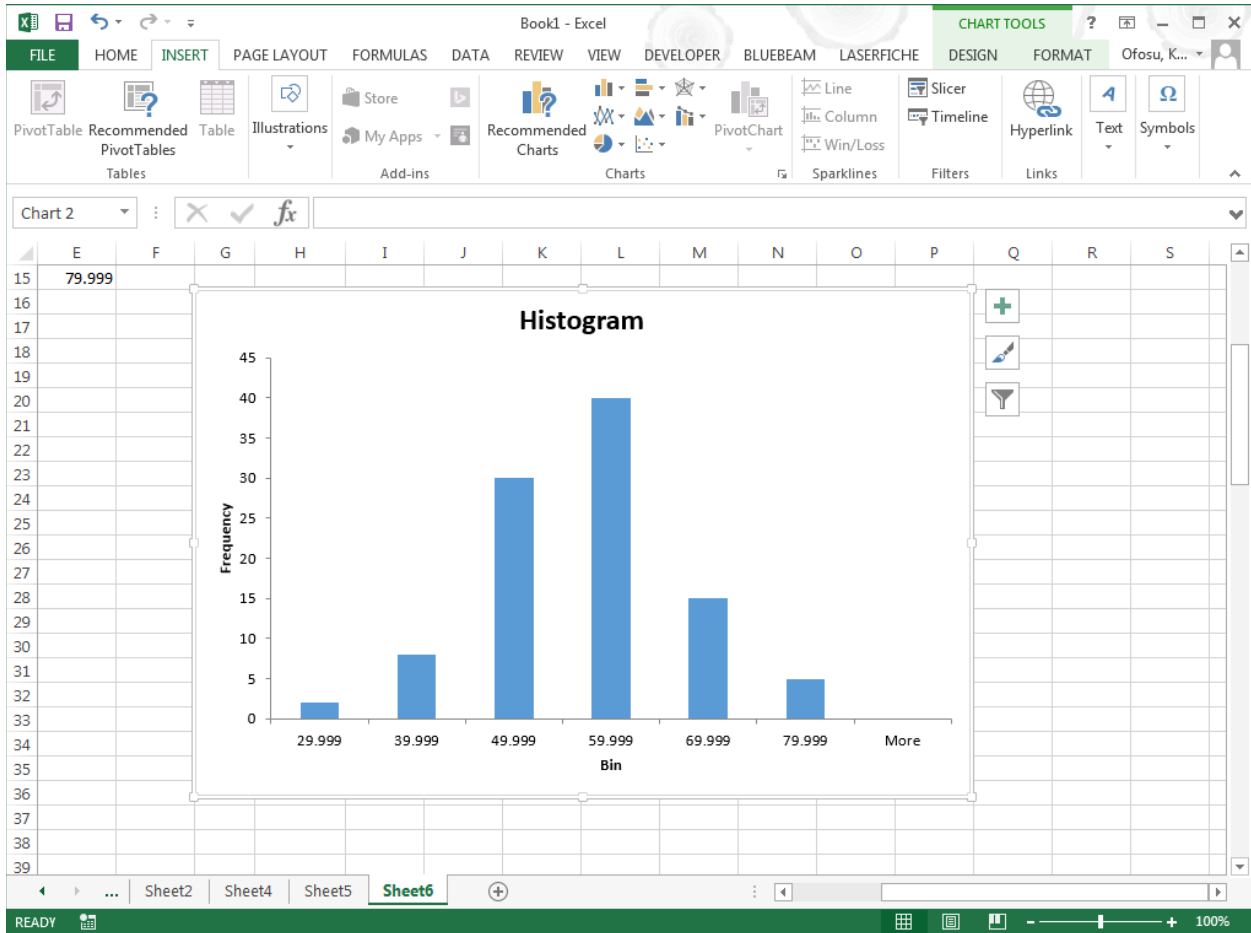
Select the legend to the right.





Engineering Methods in Excel  
A SunCam online continuing education course

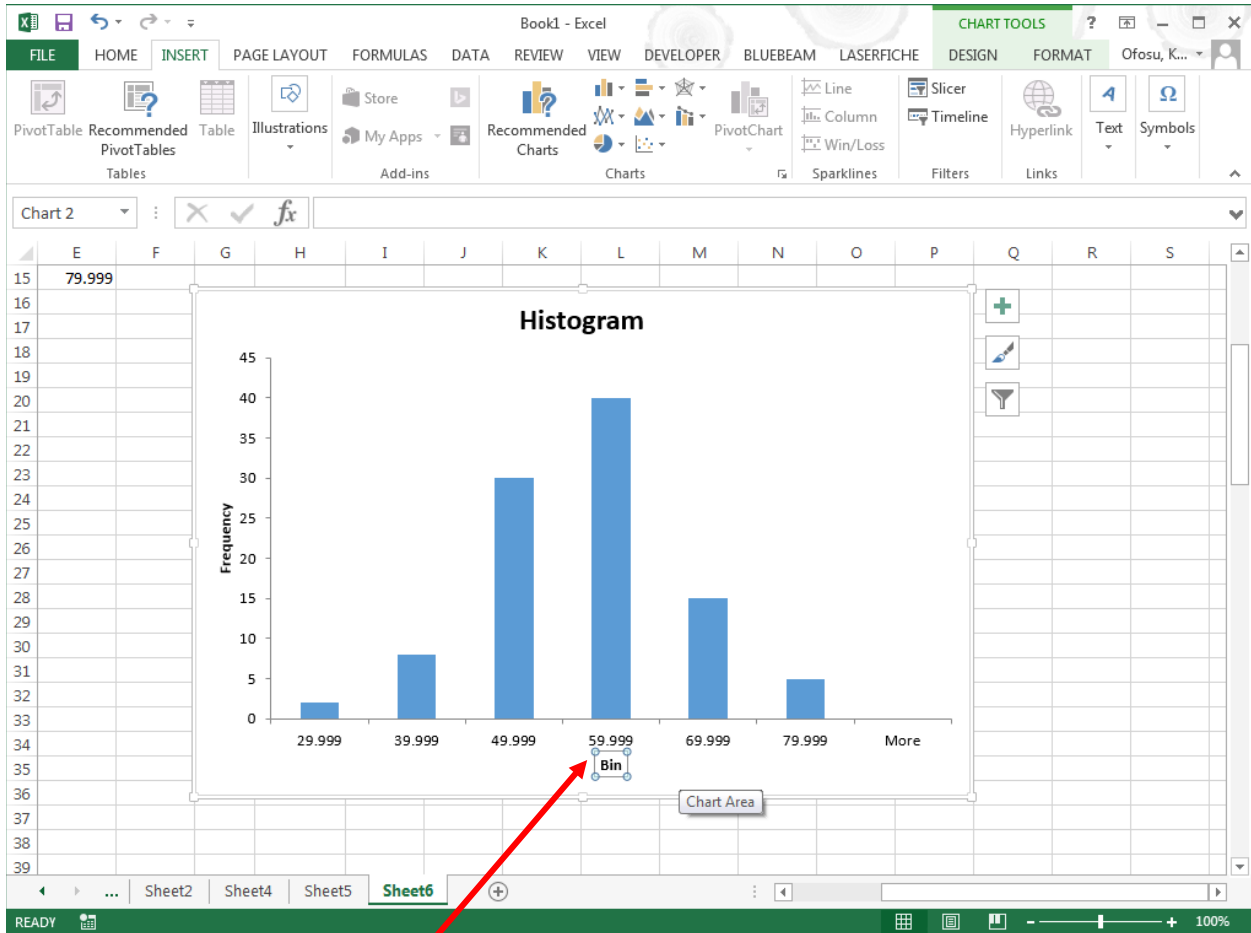
Hit **Delete** on your keyboard.





Engineering Methods in Excel  
A SunCam online continuing education course

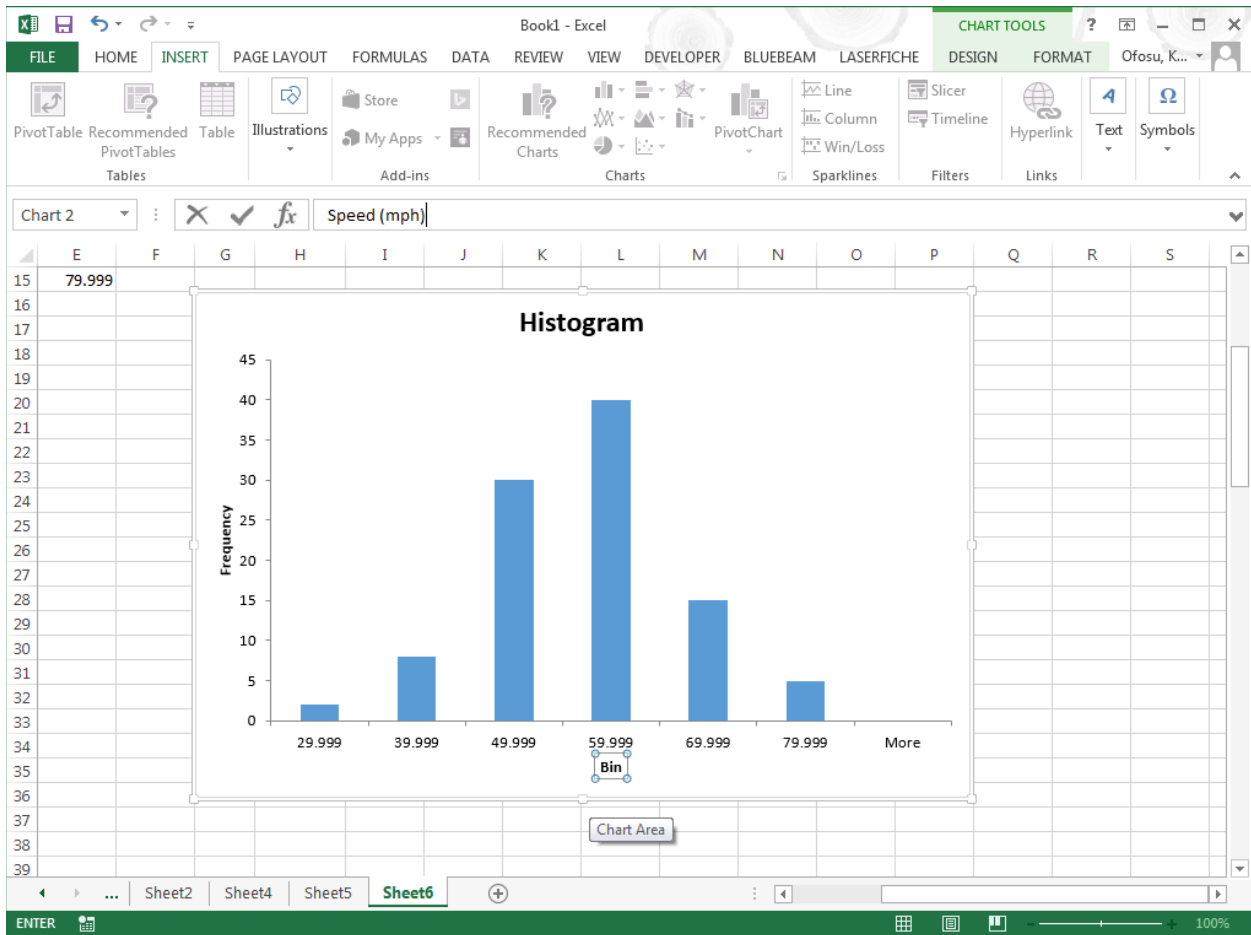
Click on the “Bin” label on the horizontal axis to select it.





Engineering Methods in Excel  
A SunCam online continuing education course

Type “Speed (mph)”

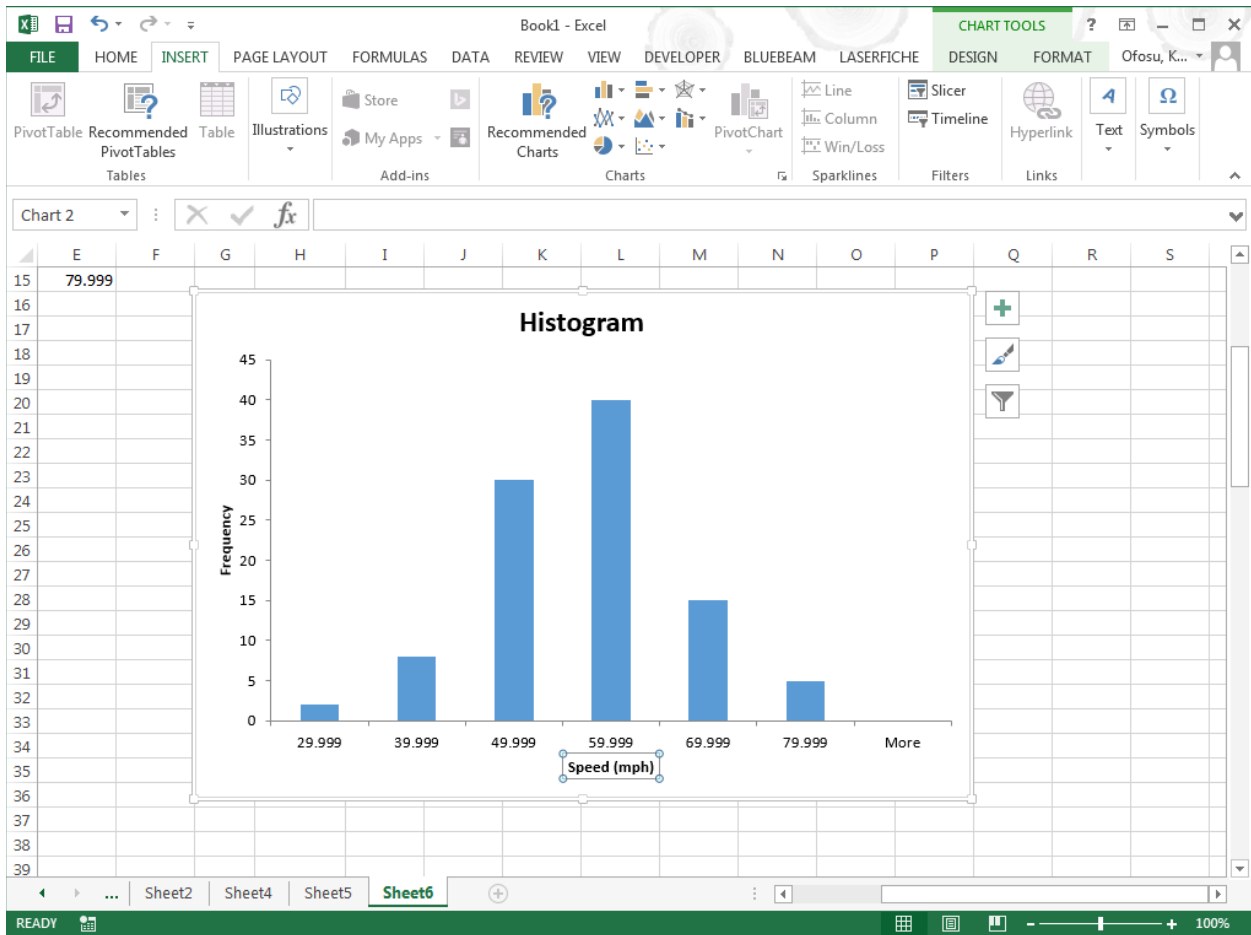


Notice that the text update appears in the formula bar.



Engineering Methods in Excel  
A SunCam online continuing education course

Hit **Enter** on the keyboard.

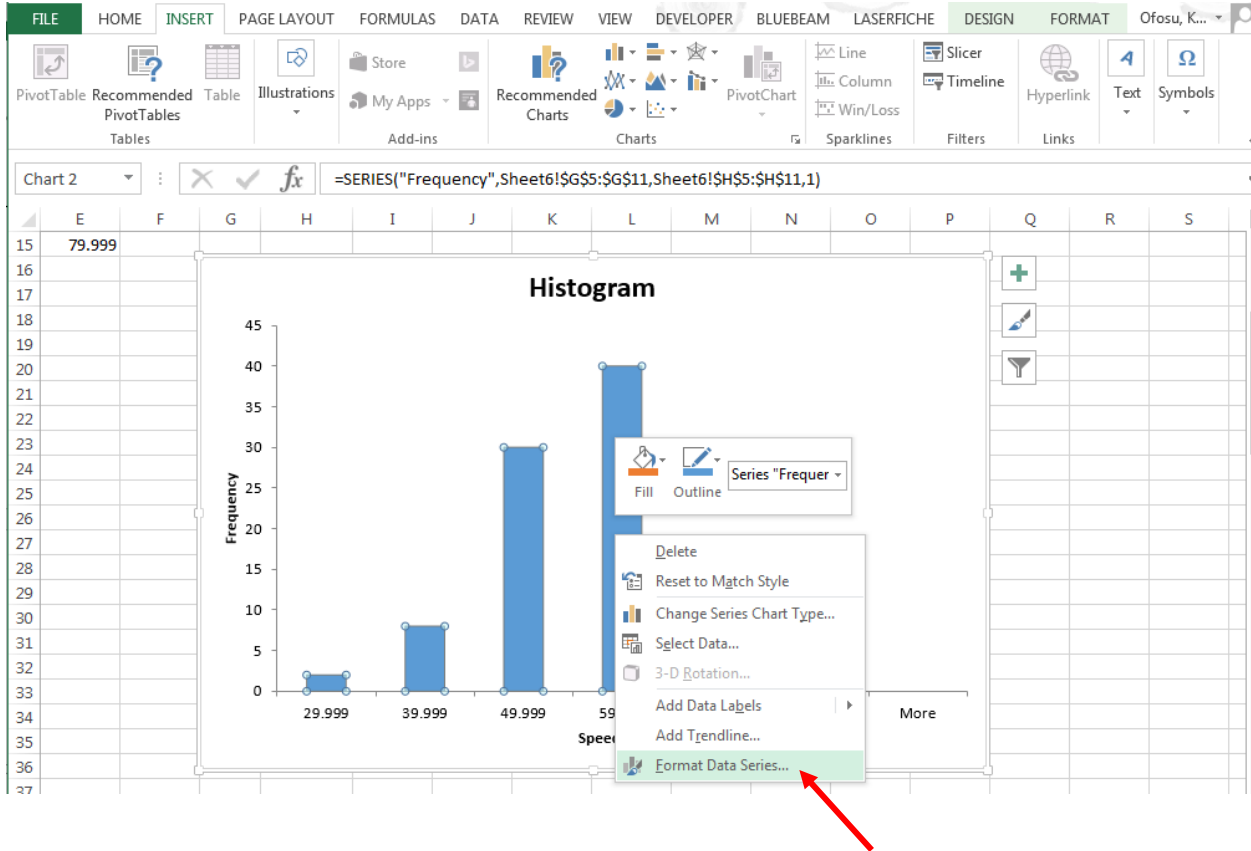






Engineering Methods in Excel  
A SunCam online continuing education course

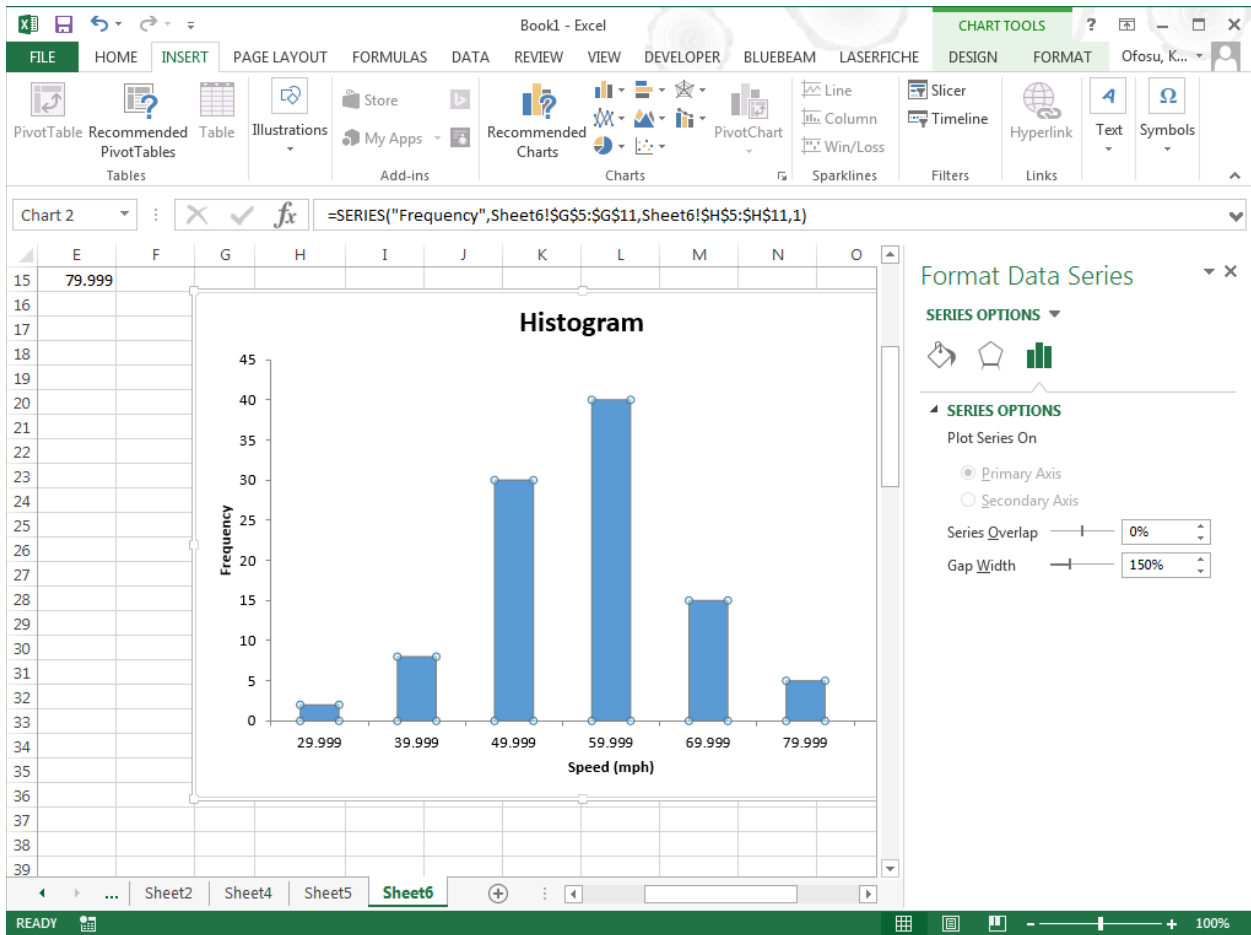
Right click over any of the bars of the histogram.  
Click on **Format Data Series**.





Engineering Methods in Excel  
A SunCam online continuing education course

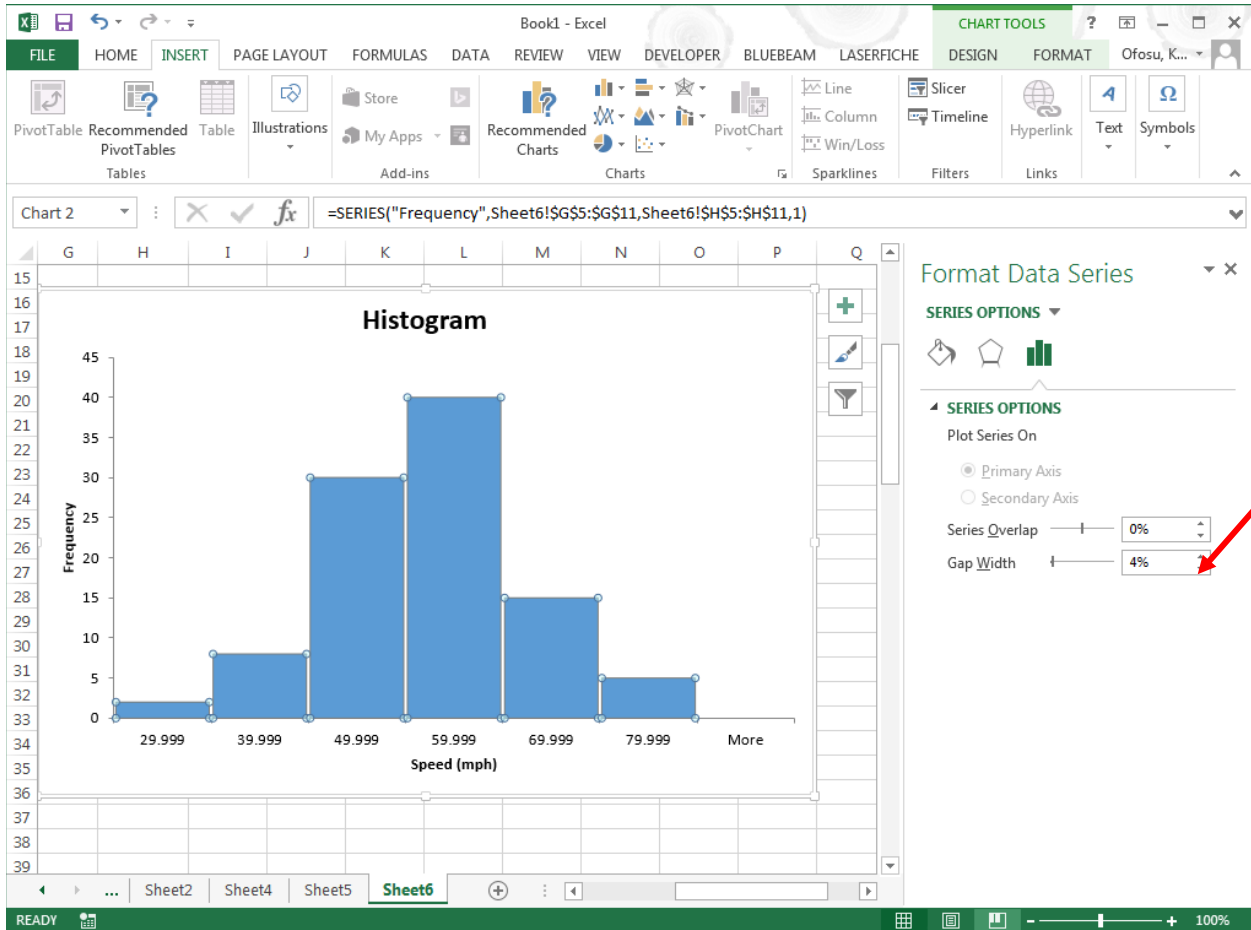
This opens the **Format Data Series** in a pane to the right.





Engineering Methods in Excel  
A SunCam online continuing education course

Manipulate the **Gap Width** value until the histogram bars are barely touching, to obtain a more professionally looking histogram.

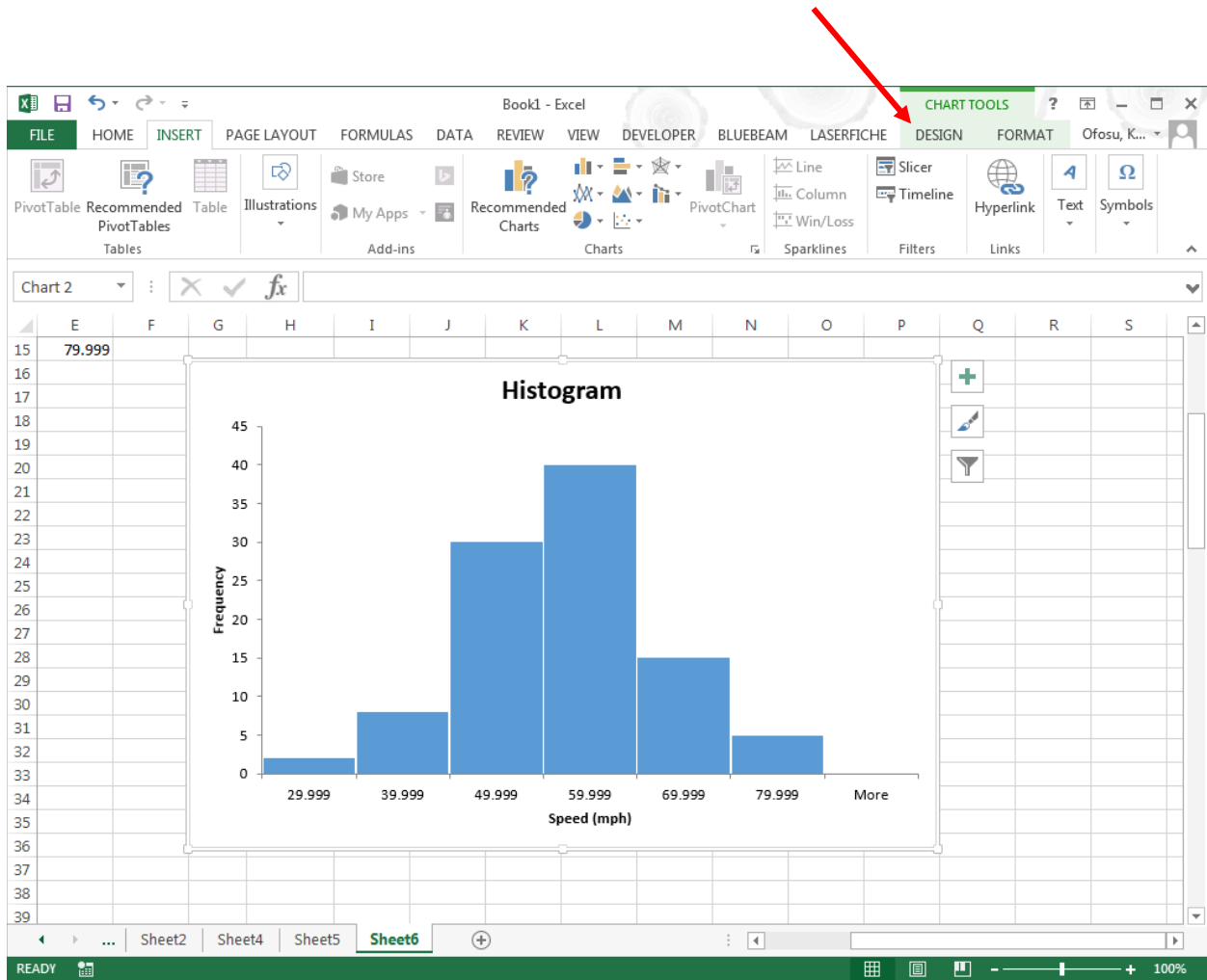


Close the **Format Data Series** pane.  
Click on any cell to take the cursor off the data series.



Engineering Methods in Excel  
A SunCam online continuing education course

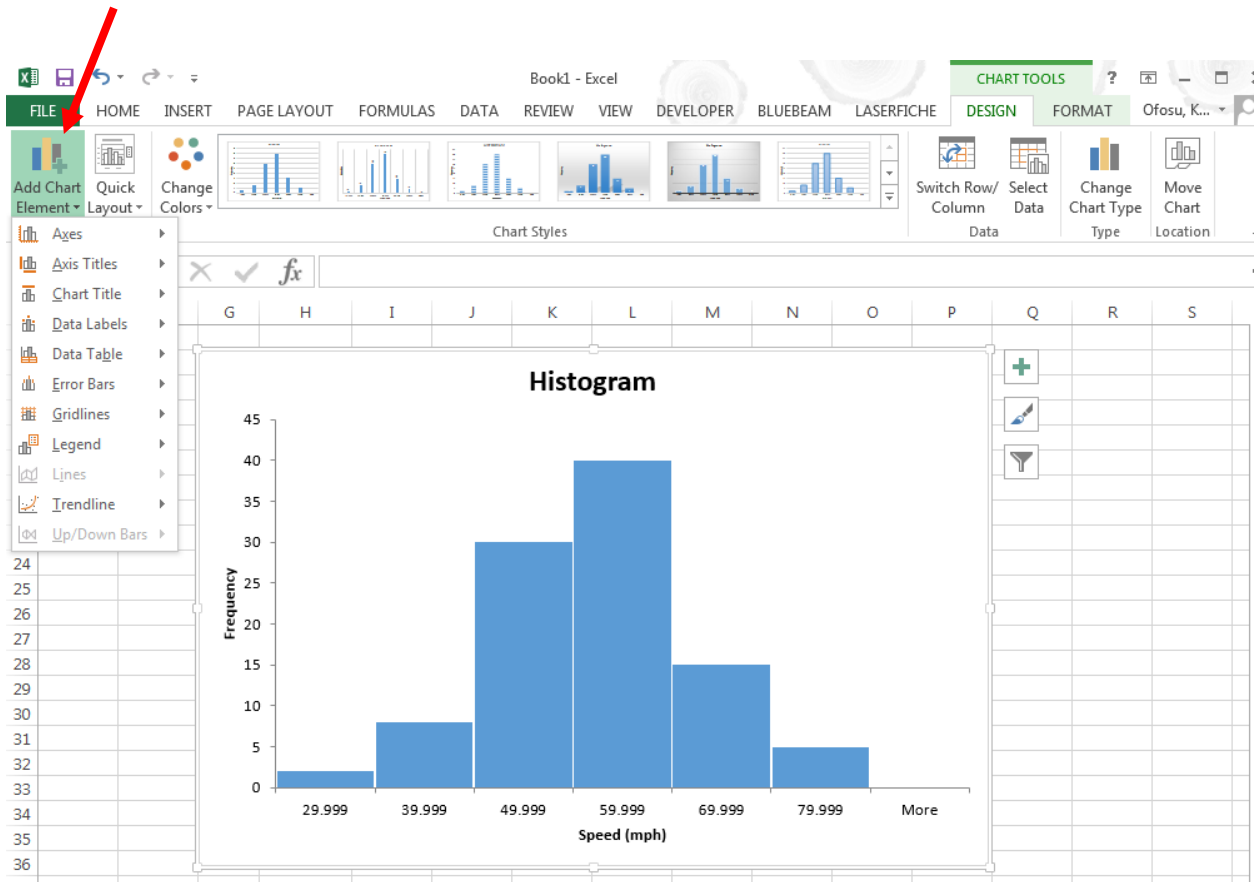
Now, click on the border of the chart to select it.  
Notice the **CHART TOOLS** tab becomes available.  
Under the **CHART TOOLS** tab, click on the **DESIGN** tab.





Engineering Methods in Excel  
A SunCam online continuing education course

Click on **Add Chart Elements**, and review the available options.



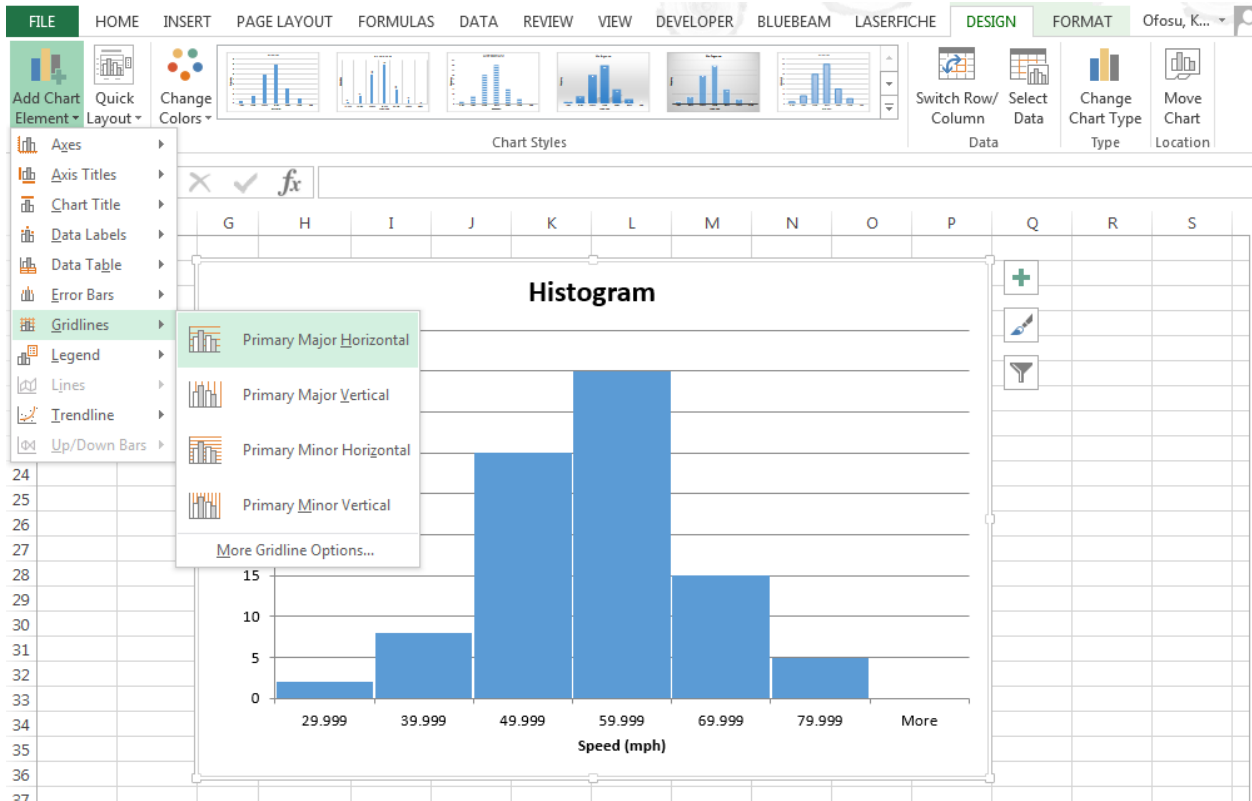
To keep it simple, we shall not add any new elements to the histogram; but review the available elements under the various options by hovering over them.



Engineering Methods in Excel  
A SunCam online continuing education course

For example, let us look at the options for adding gridlines to the chart.

Notice that any option hovered over provides a preview of what the chart will look like if that chart element is added.

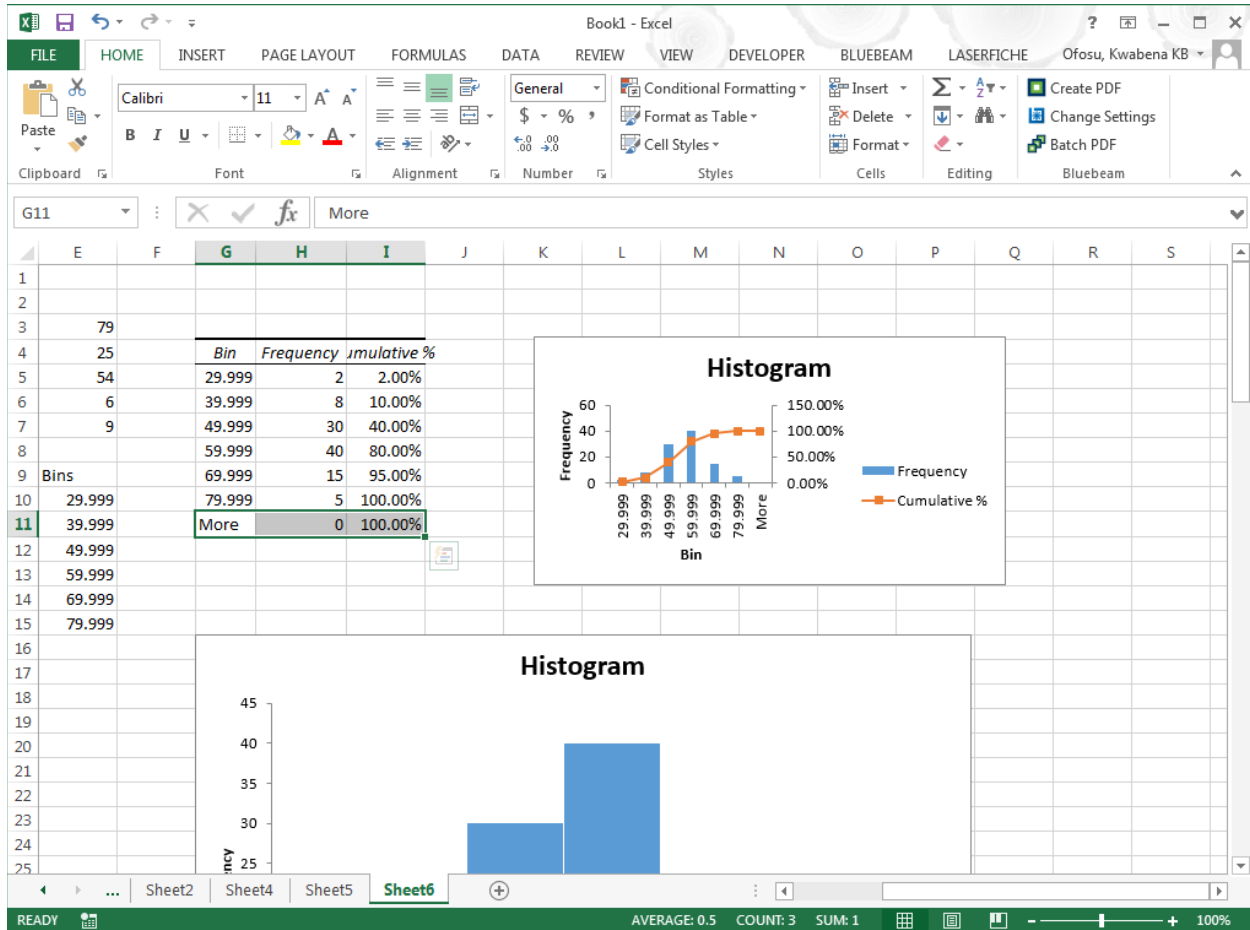


Click on the **Add Chart Elements** to close it.



## Engineering Methods in Excel A SunCam online continuing education course

Let's go to the frequency table. Note that the table has an interval named "More". Since there are no data values beyond the bins we defined the "More" interval has a frequency of zero.



We may therefore remove it from the table—which will also remove it from the histogram, and any charts on the worksheet that pull their data from this frequency table.



Engineering Methods in Excel  
A SunCam online continuing education course

Select all elements on the “More” row on the frequency table.  
Hit **Delete** on your keyboard.

The screenshot shows an Excel spreadsheet with a frequency table and two histograms. The frequency table is as follows:

Bin	Frequency	Cumulative %
29.999	2	2.00%
39.999	8	10.00%
49.999	30	40.00%
59.999	40	80.00%
69.999	15	95.00%
79.999	5	100.00%

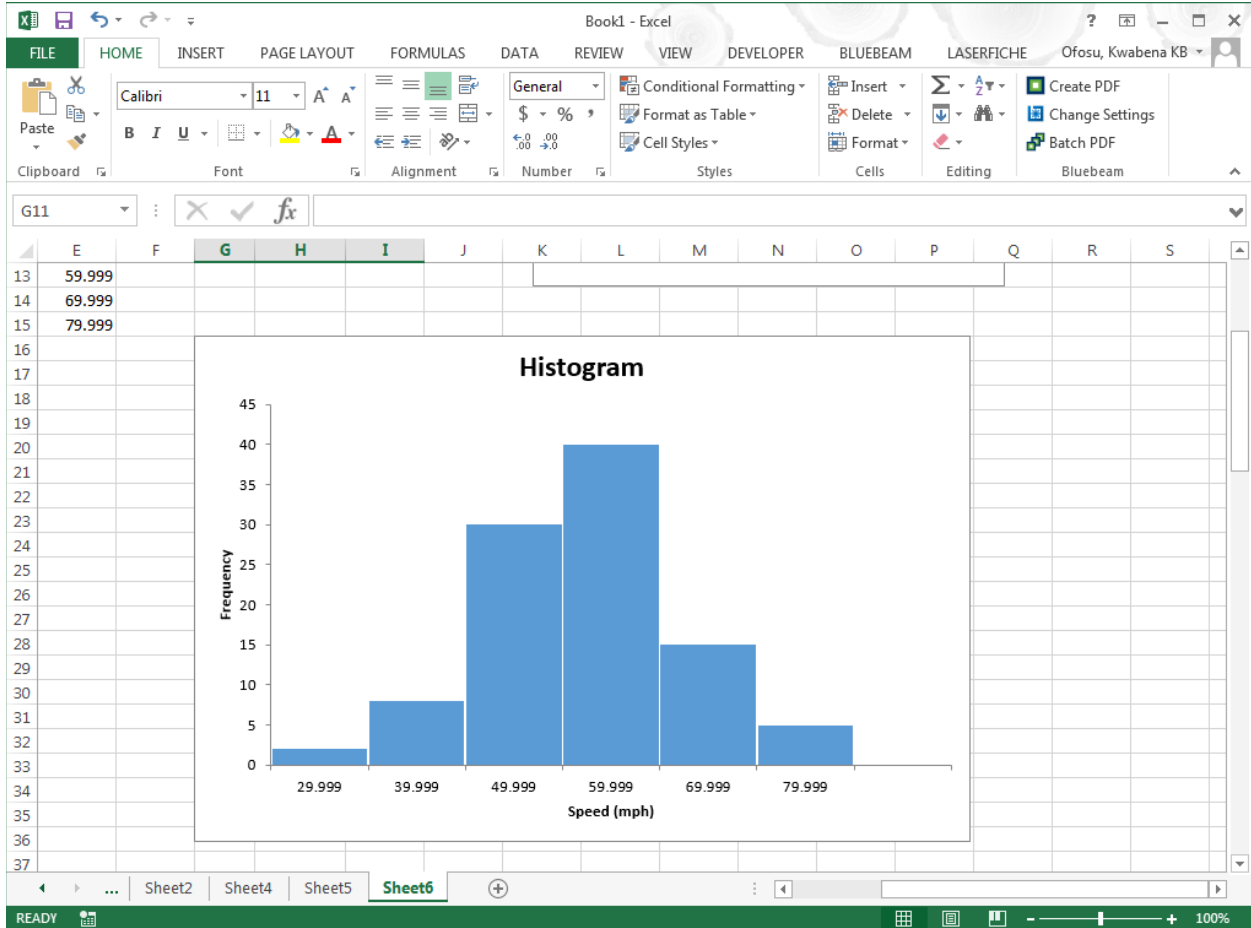
The top histogram is a combined bar and line chart showing Frequency (blue bars) and Cumulative % (orange line) against Bin values. The bottom histogram is a bar chart showing Frequency (blue bars) against Bin values.





Engineering Methods in Excel  
A SunCam online continuing education course

Notice that the “More” interval is removed from the histogram and any chart that is connected to the frequency table.



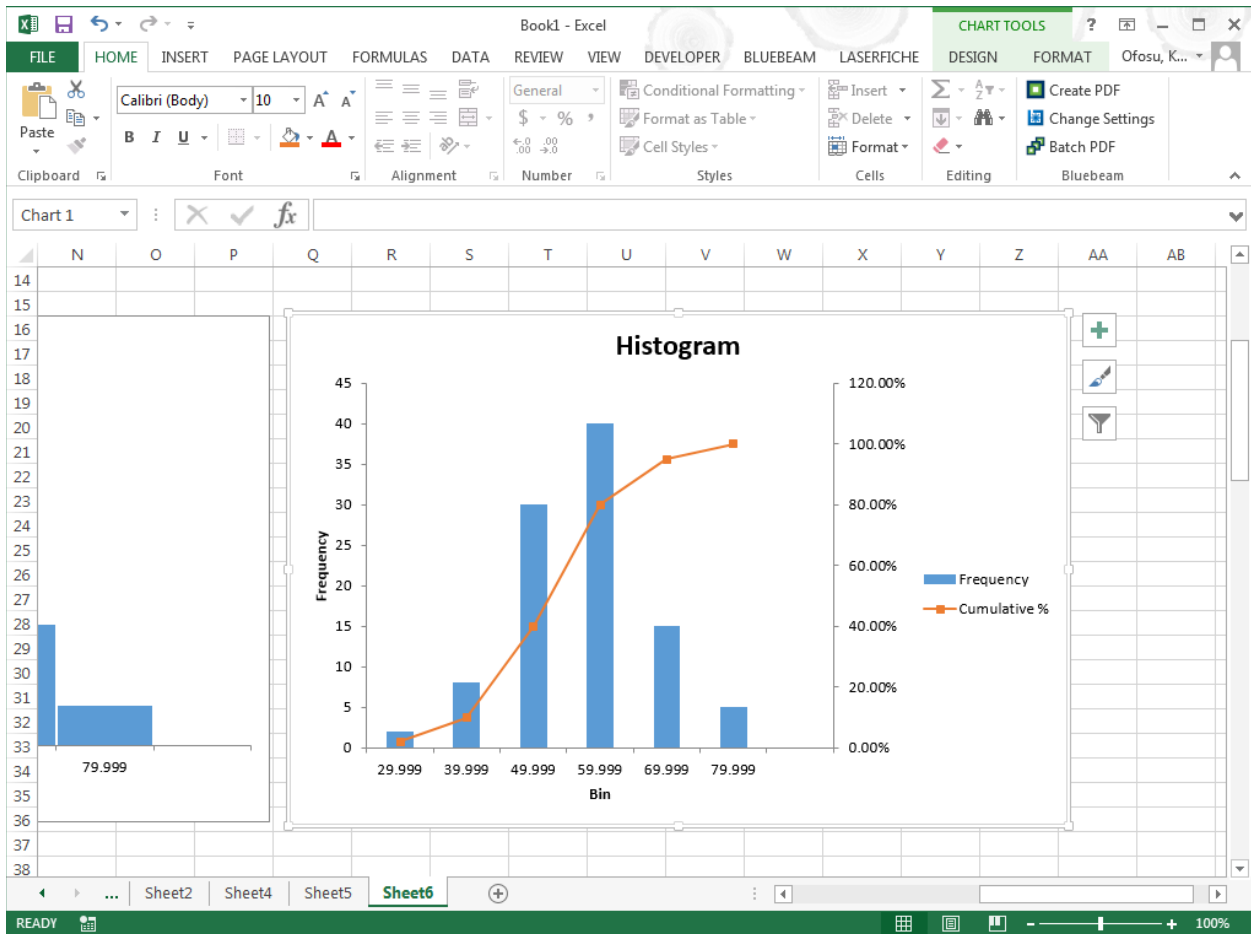
We have completed our work with the histogram.



Engineering Methods in Excel  
A SunCam online continuing education course

We shall now focus on the ogive.

Resize and reposition the original chart accordingly.





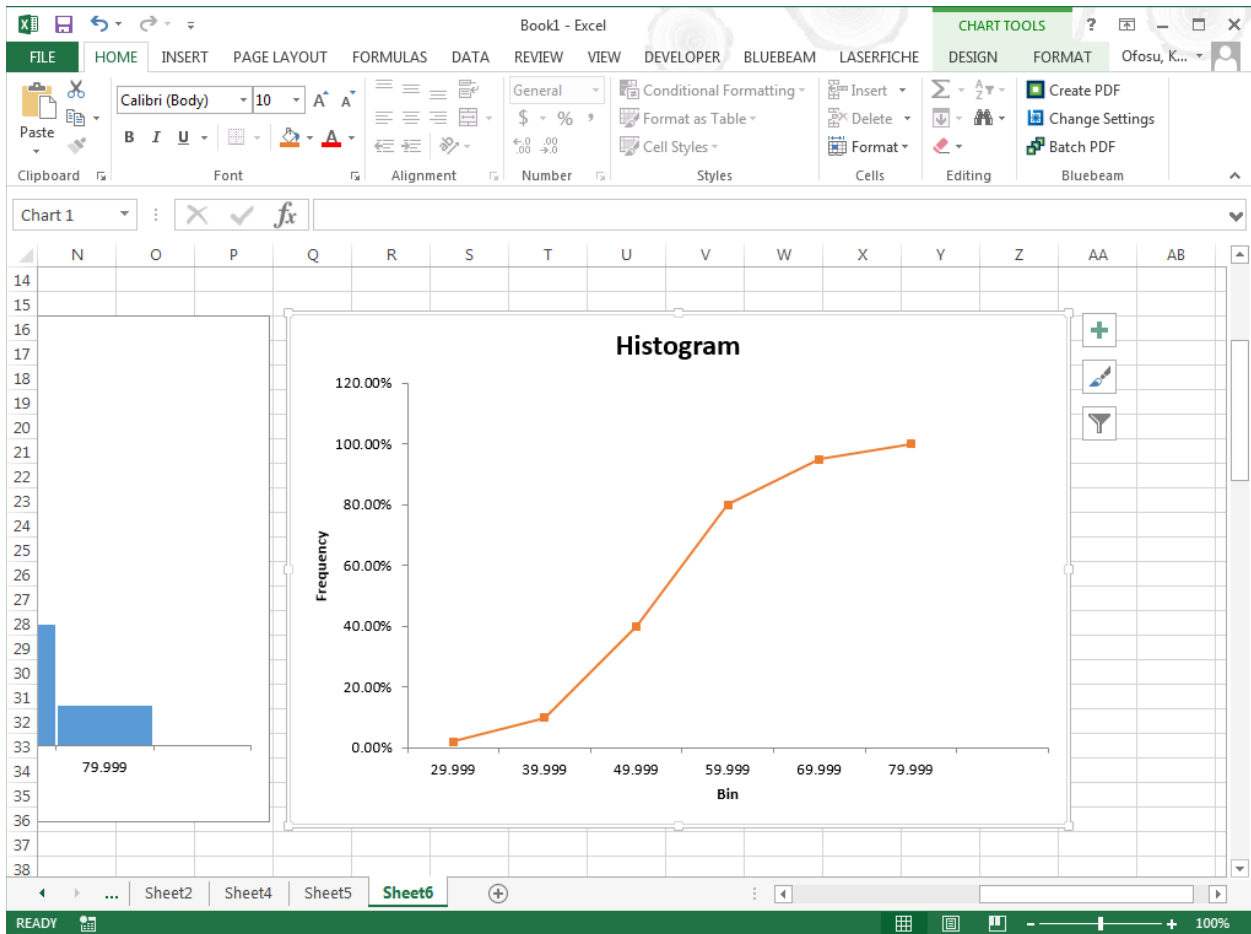
Engineering Methods in Excel  
A SunCam online continuing education course

Click on any of the bars of the histogram to select it.

Hit **Delete** on your keyboard.

Click on the legend to select it.

Hit **Delete** on your keyboard.

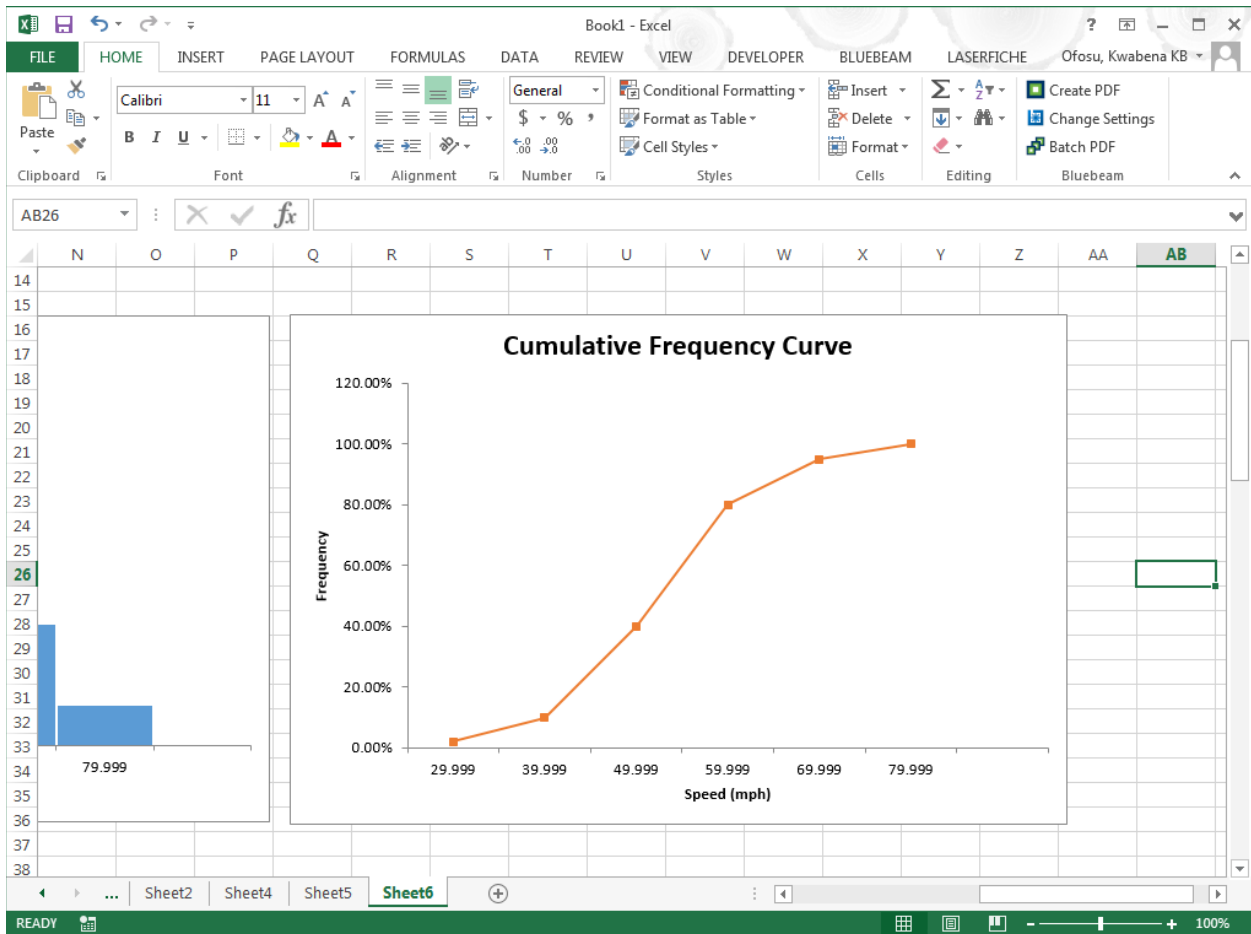




Engineering Methods in Excel  
A SunCam online continuing education course

Change the label on the horizontal axis to “Speed (mph)”.

Change the Chart title from “Histogram” to “Ogive” or “Cumulative Frequency Curve” in the same way as the horizontal axis label was changed.





Engineering Methods in Excel  
A SunCam online continuing education course

Make sure the chart is selected (by clicking in its border).

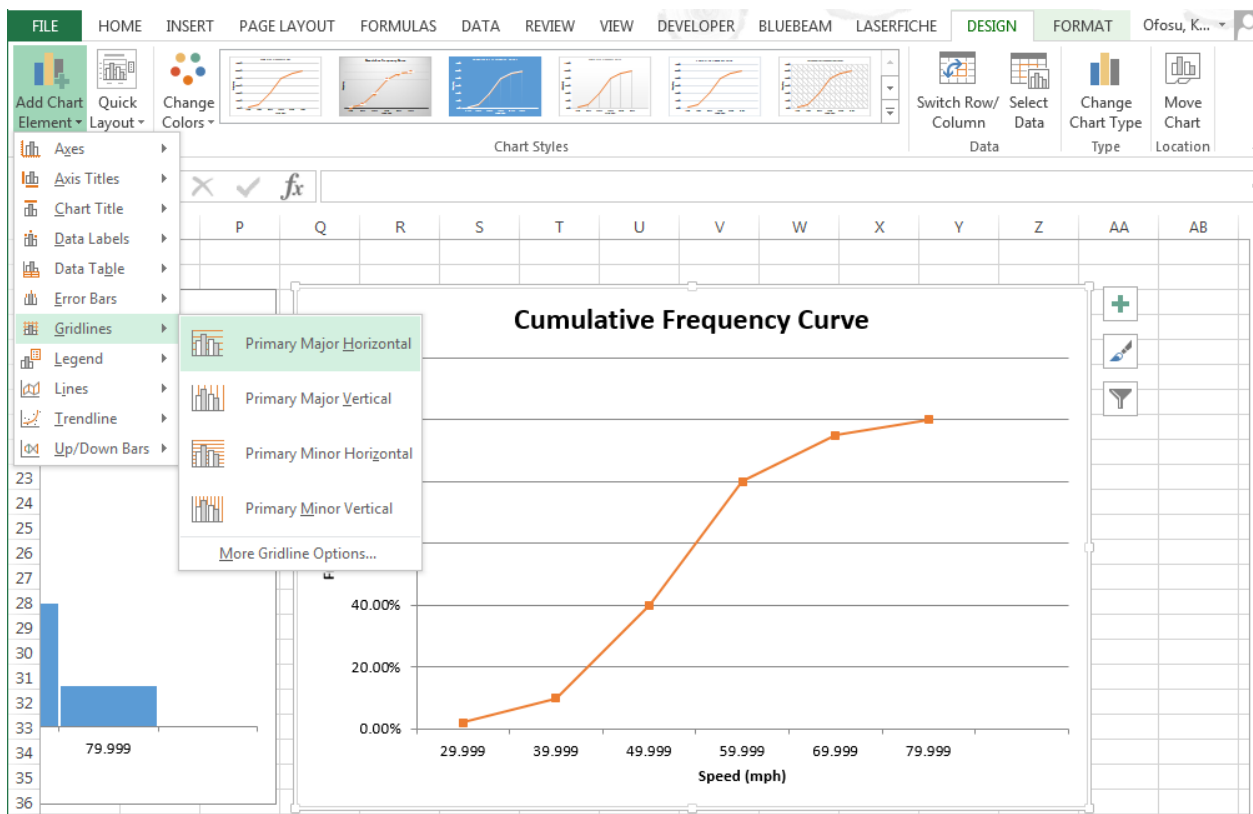
Go to **CHART TOOLS**.

Click on **DESIGN**.

Click on **Add Chart Elements**.

Hover over **Gridlines**.

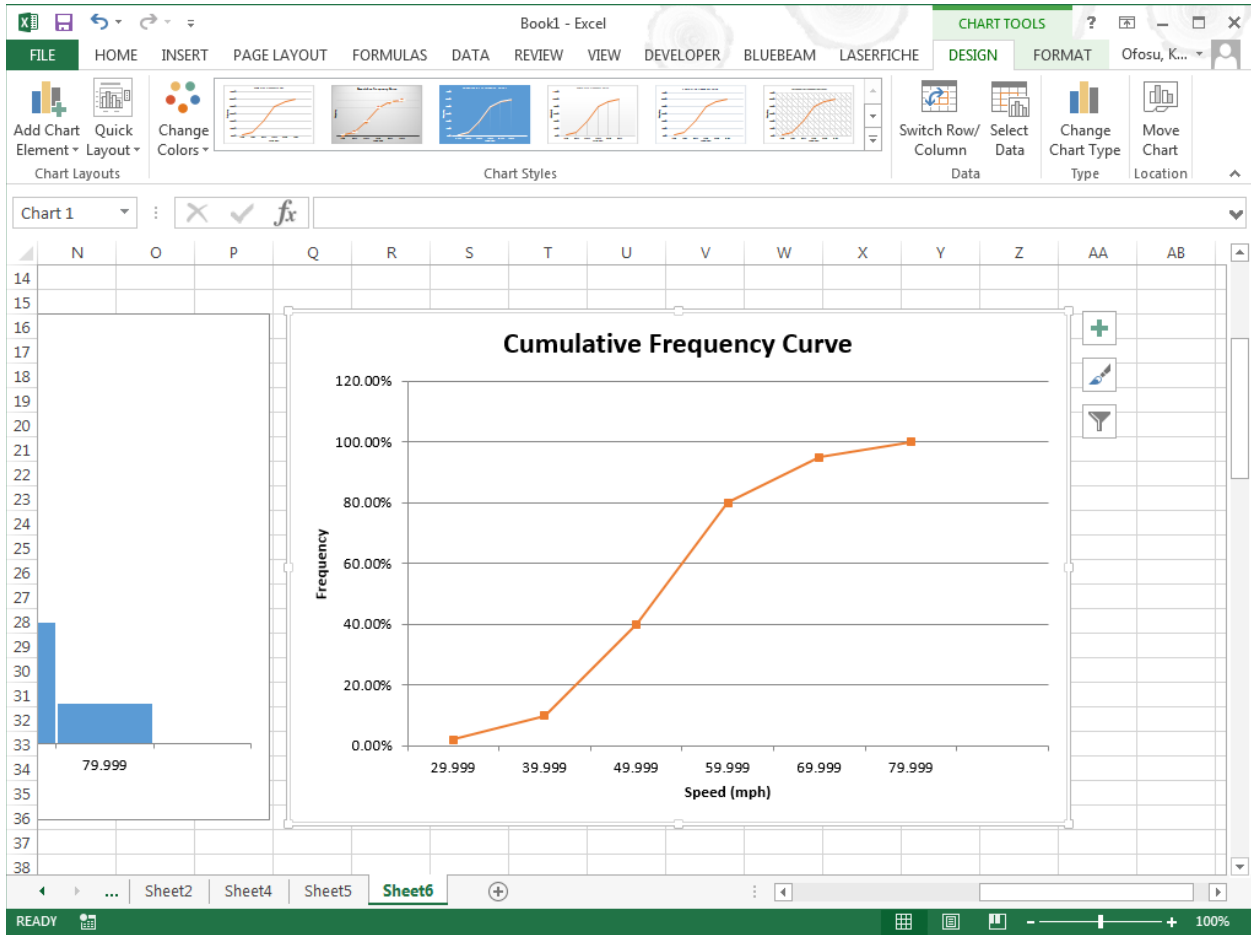
Click on **Primary Major Horizontal**.





Engineering Methods in Excel  
A SunCam online continuing education course

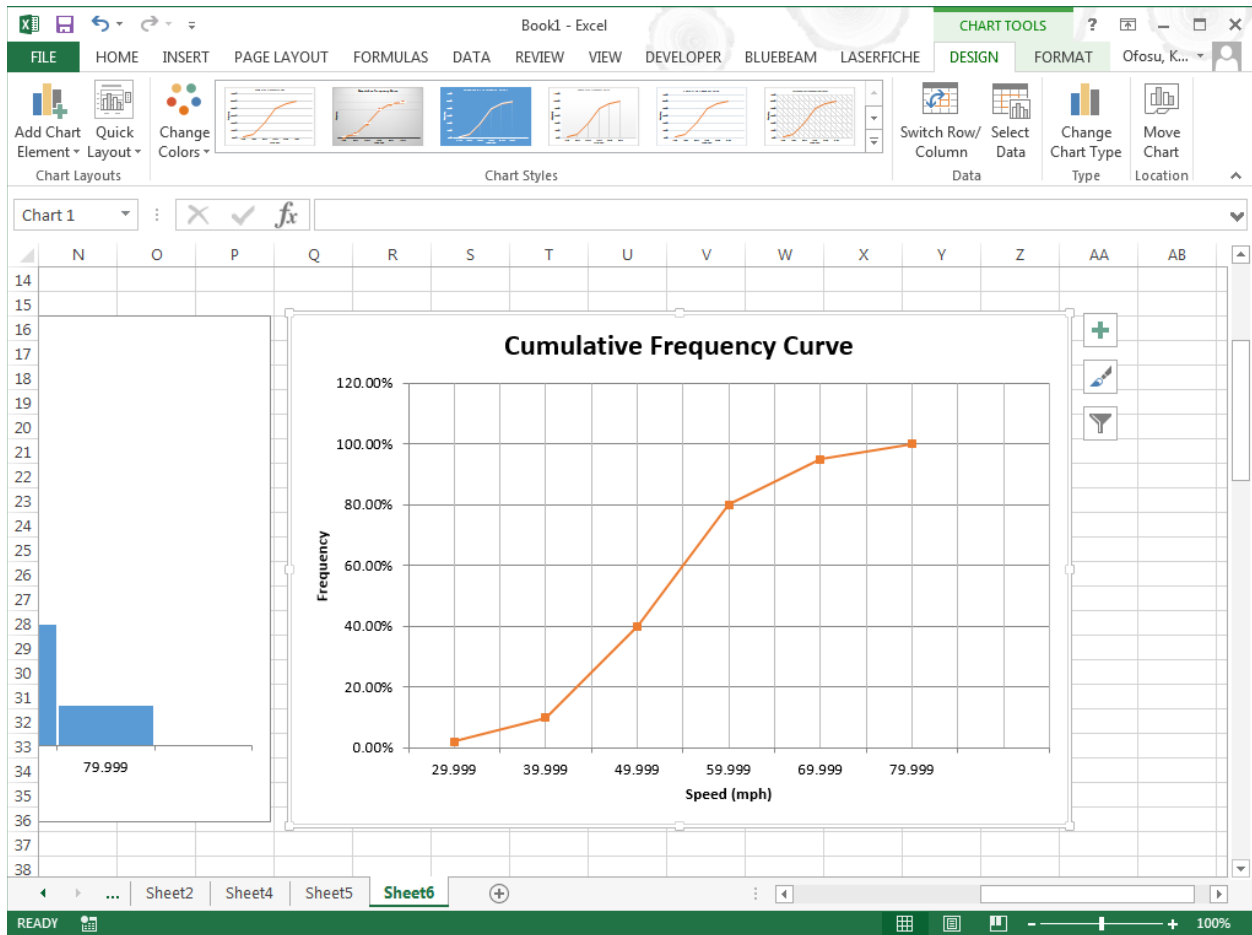
The gridlines are added to the chart.





Engineering Methods in Excel  
A SunCam online continuing education course

Repeat the process to add **Primary Minor Vertical** gridlines.

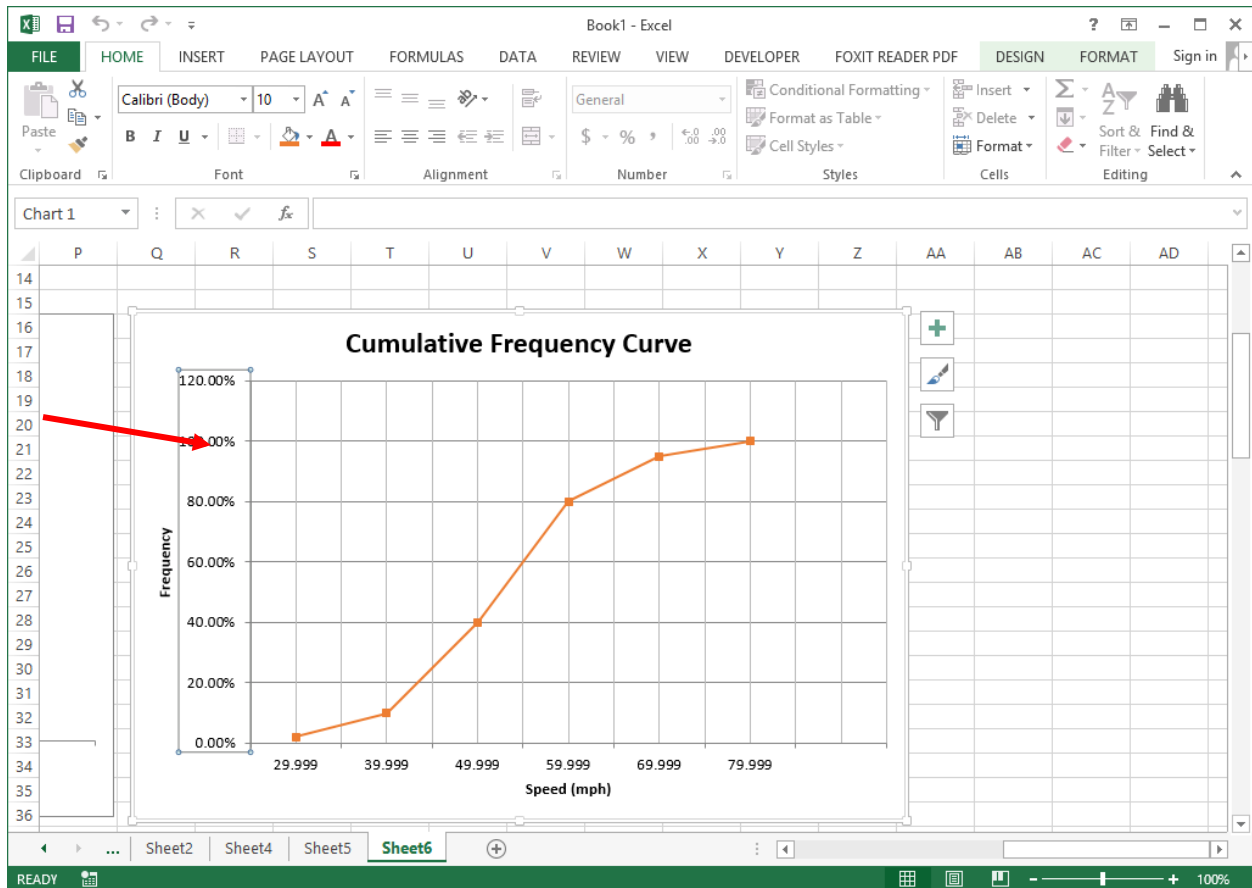


The scale on the vertical axis shows: the values with tick marks in steps of 20 and ending at 120%. We would like to change that to show: steps of 10 labeled, with intermediary marks for every 5 units—but not labelled—and ending at 100% (as any percentage on an ogive in excess of 100%, is meaningless). We shall also change the number format on the vertical axis to have no decimal places.



Engineering Methods in Excel  
A SunCam online continuing education course

Click on any of the values on the vertical axis to select the vertical axis.

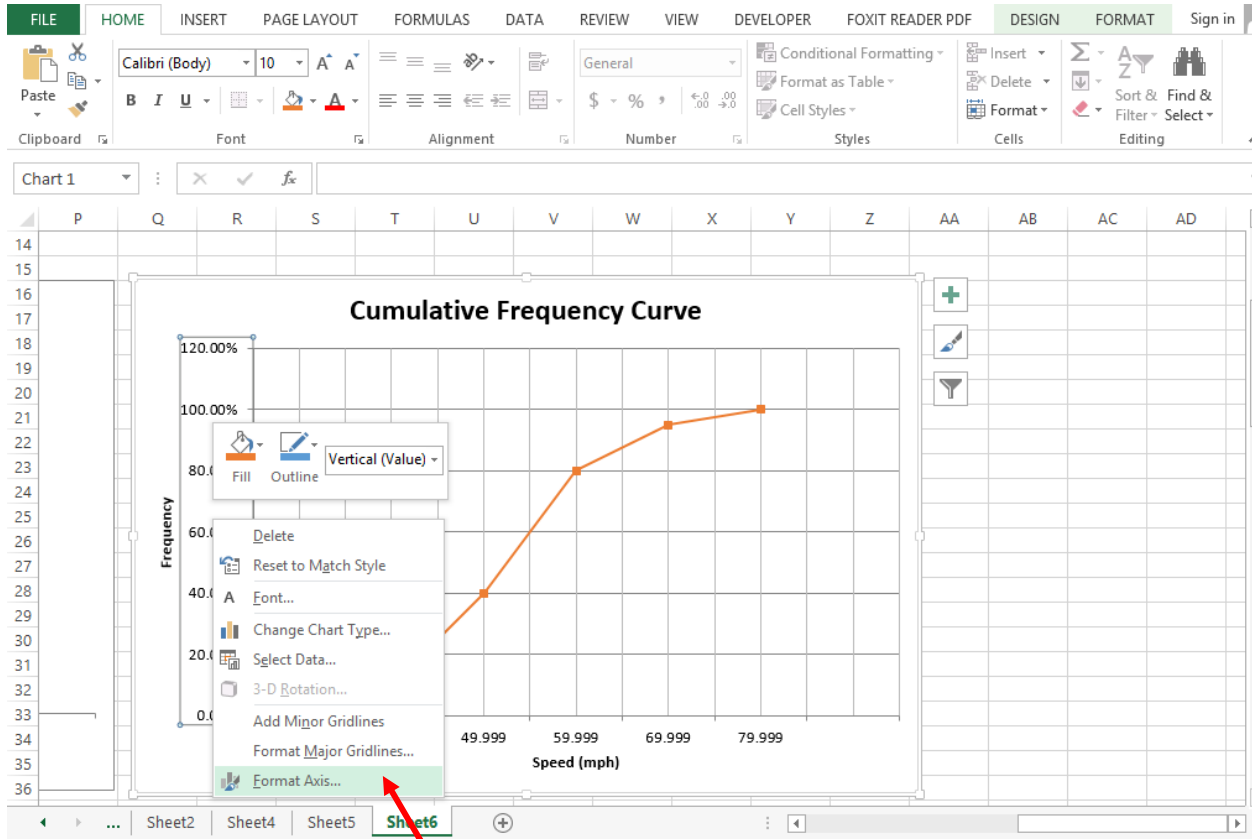






Engineering Methods in Excel  
*A SunCam online continuing education course*

Right click.  
Select **Format Axis**.





Engineering Methods in Excel  
A SunCam online continuing education course

The **Format Axis** pane opens to the right.

The screenshot displays the Microsoft Excel interface with a chart titled "Cumulative Frequency Curve". The chart's vertical axis is labeled "Frequency" and ranges from 0.00% to 120.00% in 20.00% increments. The horizontal axis is labeled "Speed (mph)" and has major tick marks at 29.999, 39.999, 49.999, 59.999, 69.999, and 79.999. The chart shows a smooth orange curve connecting several data points. To the right of the chart, the "Format Axis" task pane is open, showing the "AXIS OPTIONS" tab. The "Bounds" section has "Minimum" set to 0.0 and "Maximum" set to 1.2, both with "Auto" units. The "Units" section has "Major" set to 0.2 and "Minor" set to 0.04, both with "Auto" units. The "Horizontal axis crosses" section has "Automatic" selected. The "Display units" dropdown is set to "None".

Speed (mph)	Frequency (%)
29.999	0.00
39.999	10.00
49.999	40.00
59.999	80.00
69.999	95.00
79.999	100.00



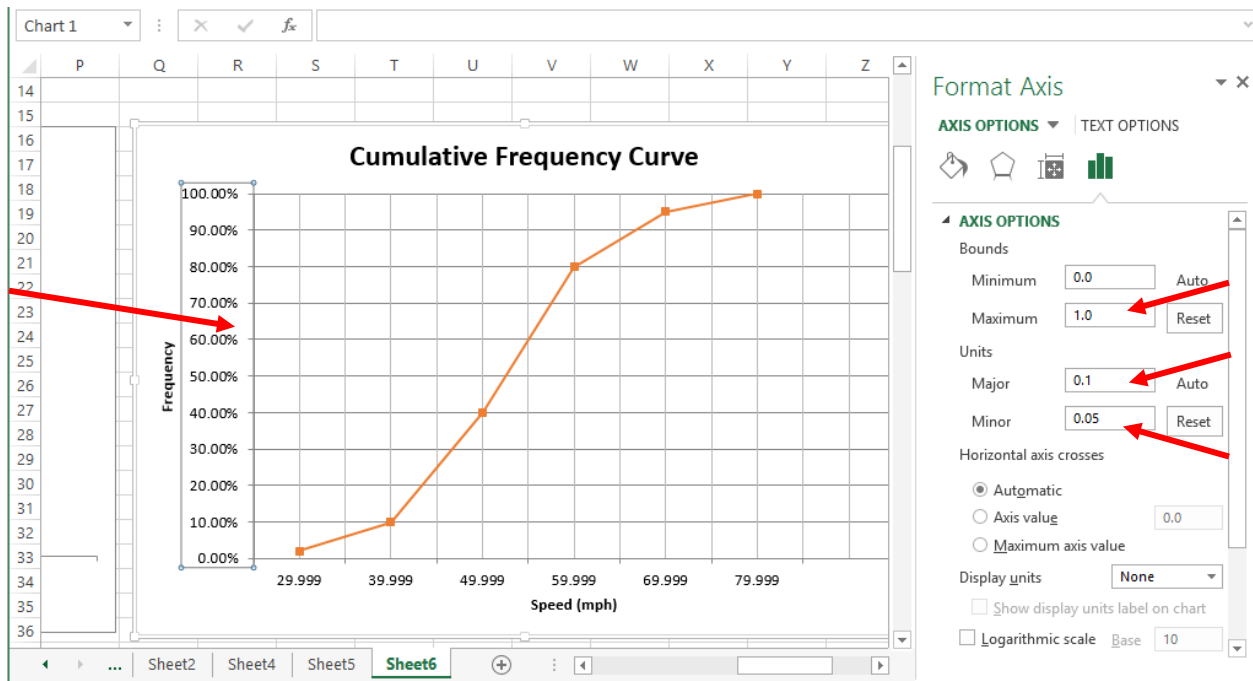
Engineering Methods in Excel  
A SunCam online continuing education course

Under **Bounds – Maximum**, change the value to 1.0.

Under **Units – Major**, change the value to 0.1

Under **Units – Minor**, change the value to 0.05.

Observe the axis values and the graph shifting accordingly.





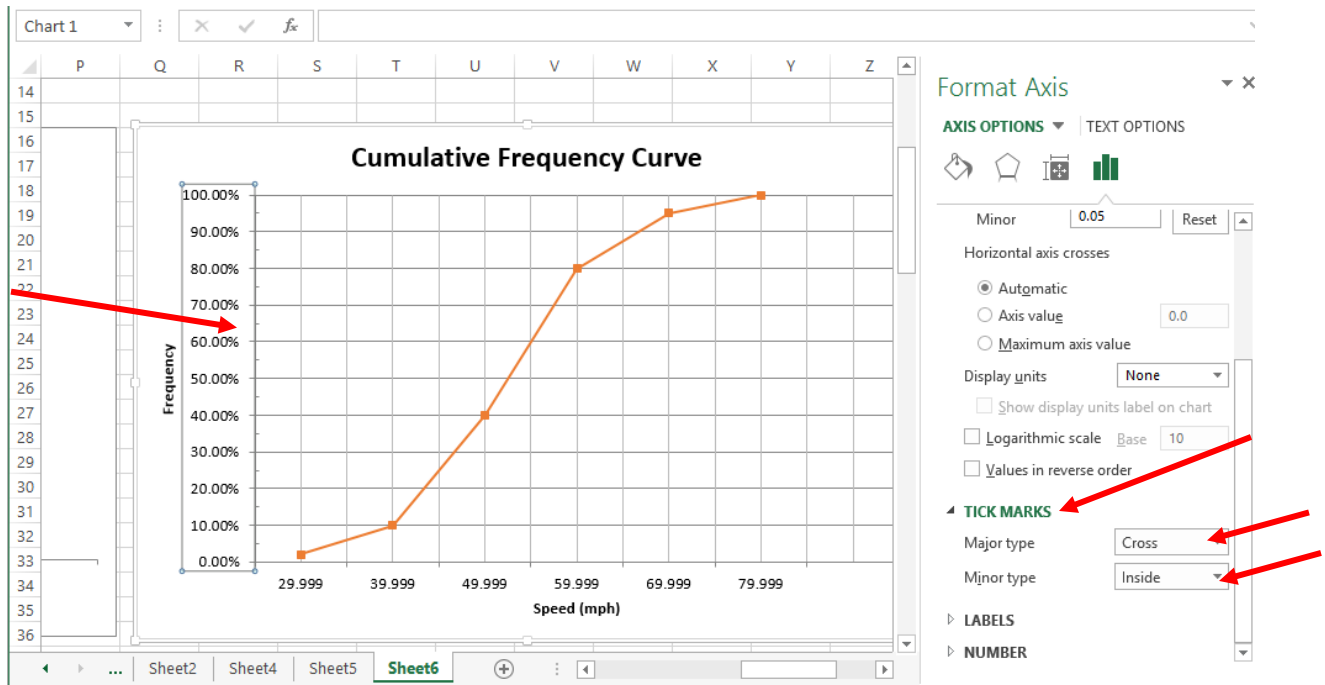
Engineering Methods in Excel  
A SunCam online continuing education course

Click on **TICK MARKS** to expand it to view the available options. (You may have to scroll down the **Format Axis** pane to see the **TICK MARKS** option).

Change the **Major type** to **Cross**.

Change the **Minor type** to **Inside**.

Observe how the tick marks on the vertical axis have changed accordingly.

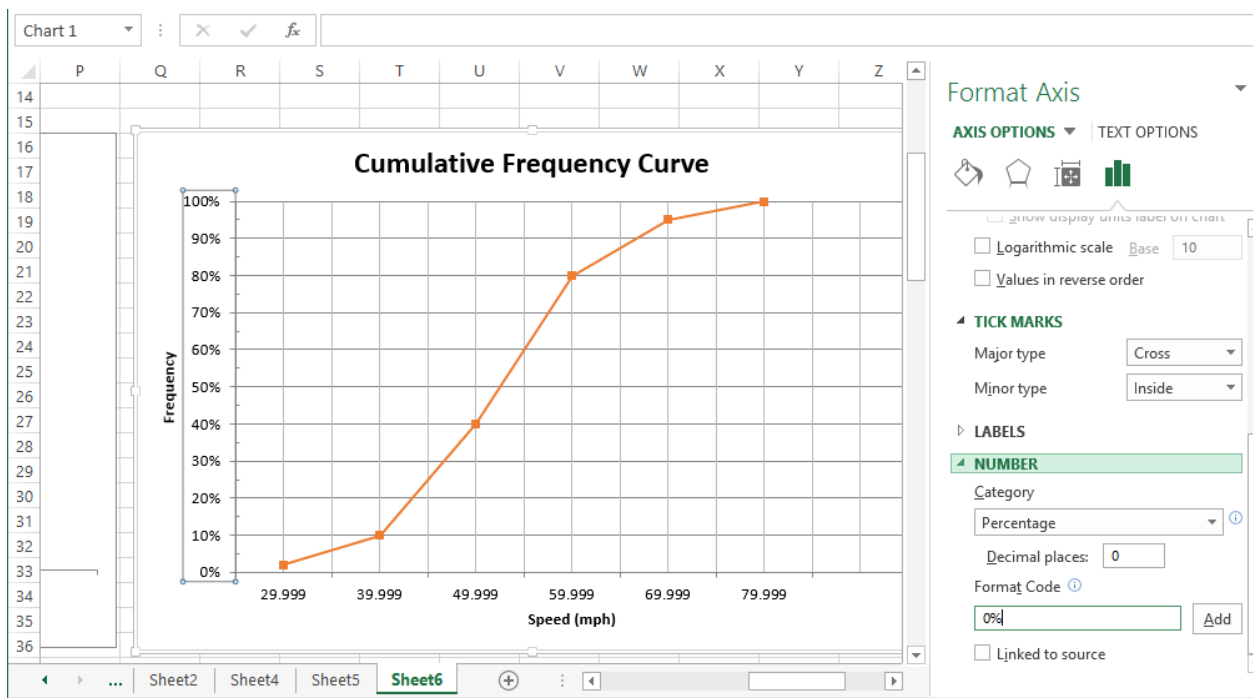




Engineering Methods in Excel  
A SunCam online continuing education course

Click on **NUMBER** to expand it to view the available options.  
Change the **Decimal places** value to zero.

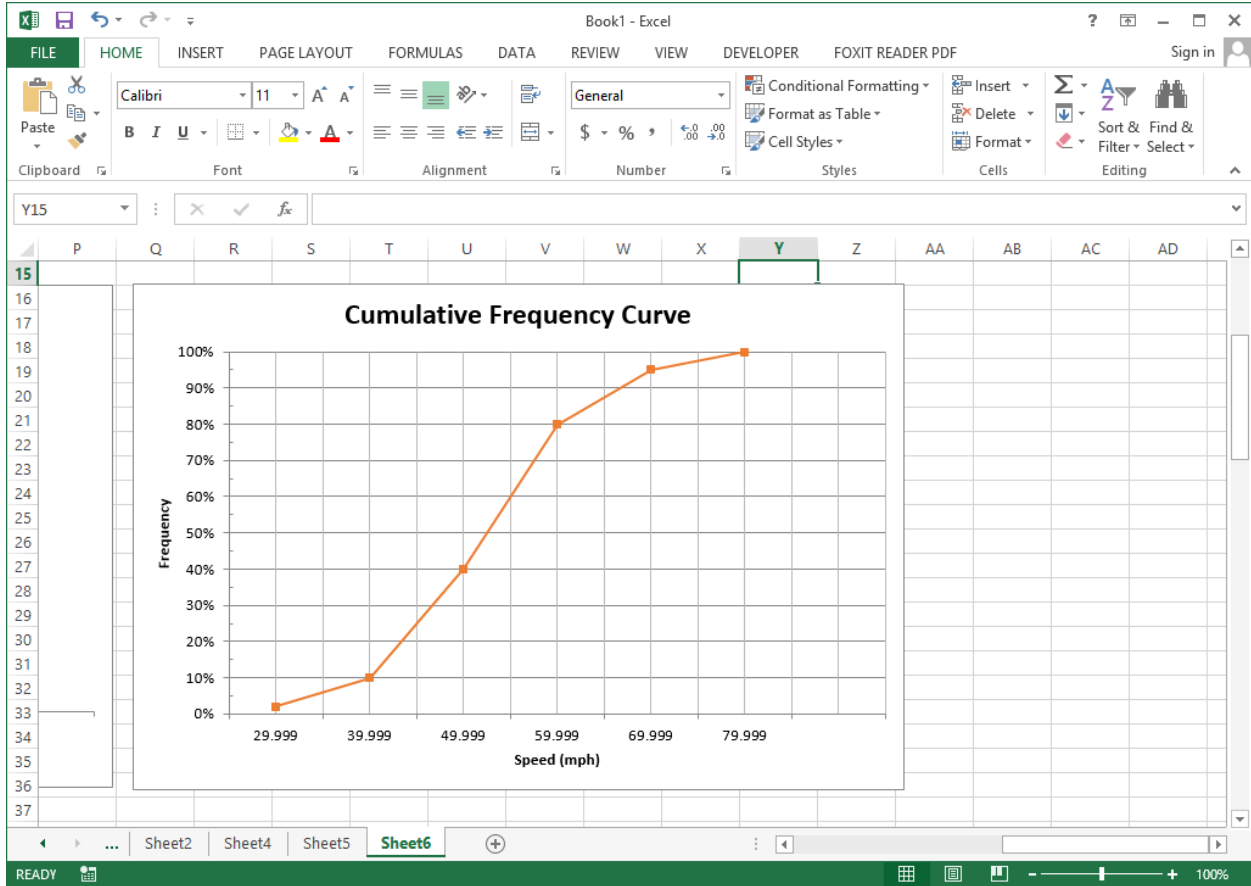
Observe how the number formatting on the vertical axis has changed accordingly.





Engineering Methods in Excel  
A SunCam online continuing education course

Close the **Format Axis** pane.  
Click away from the chart.



Cumulative frequency curve complete!



Engineering Methods in Excel  
A SunCam online continuing education course

Alternate Method for Histograms (Excel 2016 and beyond users only)

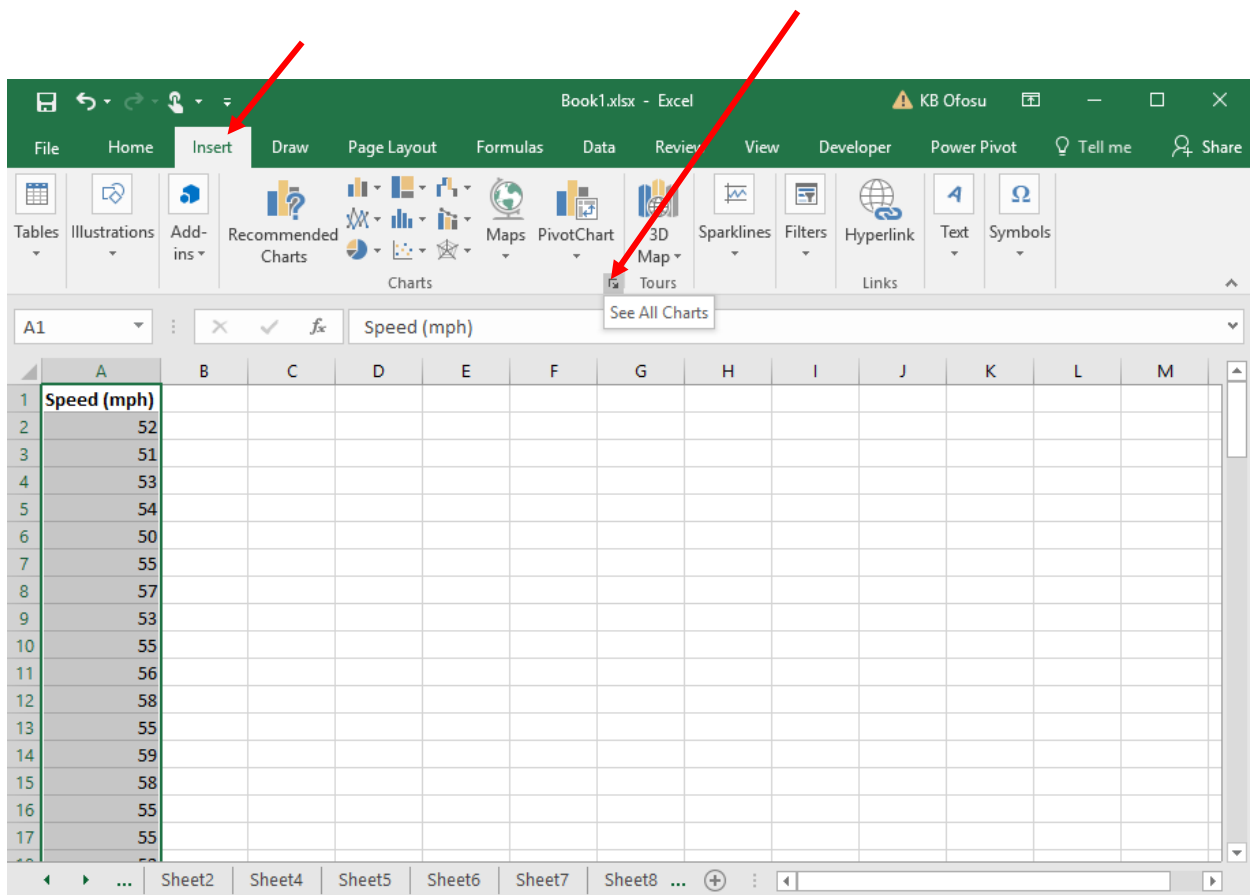
This sub-section can be performed only if you have the 2016 version and above, of *Excel*, otherwise move on the next section. \*There is no penalty for not completing this section if you do not have the software upgrades and/or updates for it.\*

Starting with the 2016 release, *Excel* provides a built-in functionality for plotting histograms directly from the general tools for creating charts.

Select the data.

Click on the **Insert** tab.

Click on **See All Charts**.



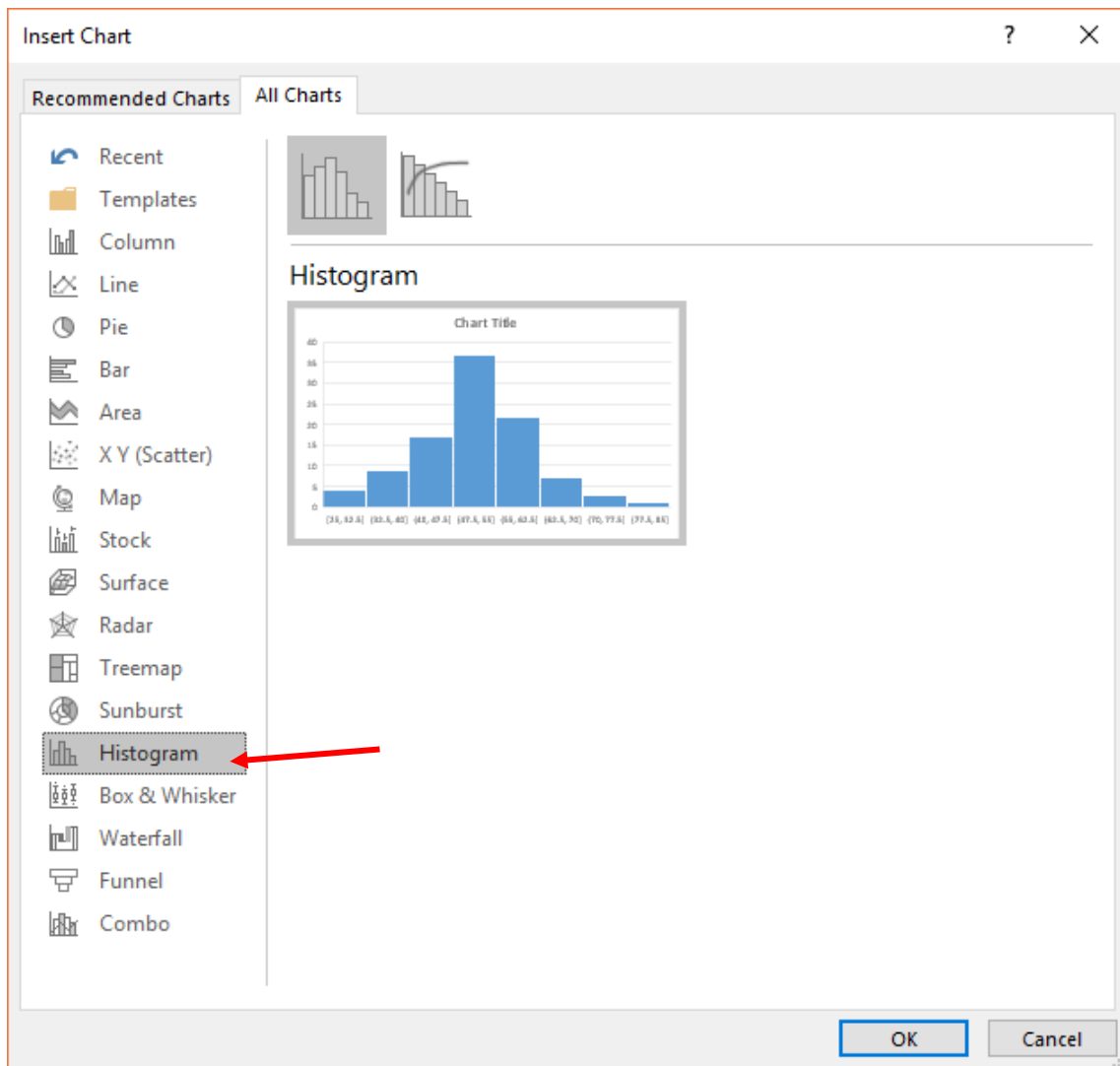


Engineering Methods in Excel  
*A SunCam online continuing education course*

The **Insert Chart** window opens.

Click on the tab **All Charts**.

Click on **Histogram**.



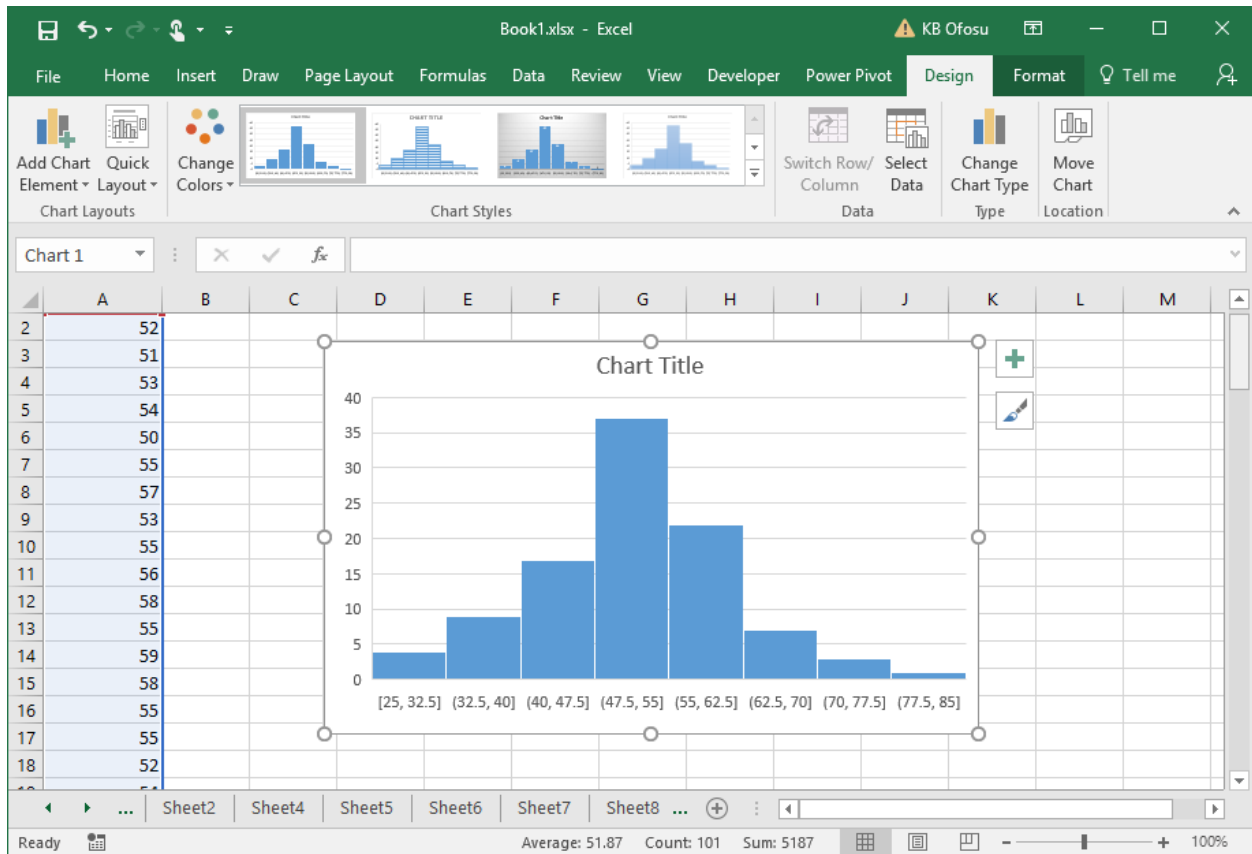
Click **OK**.





Engineering Methods in Excel  
A SunCam online continuing education course

A histogram is inserted on the spreadsheet.

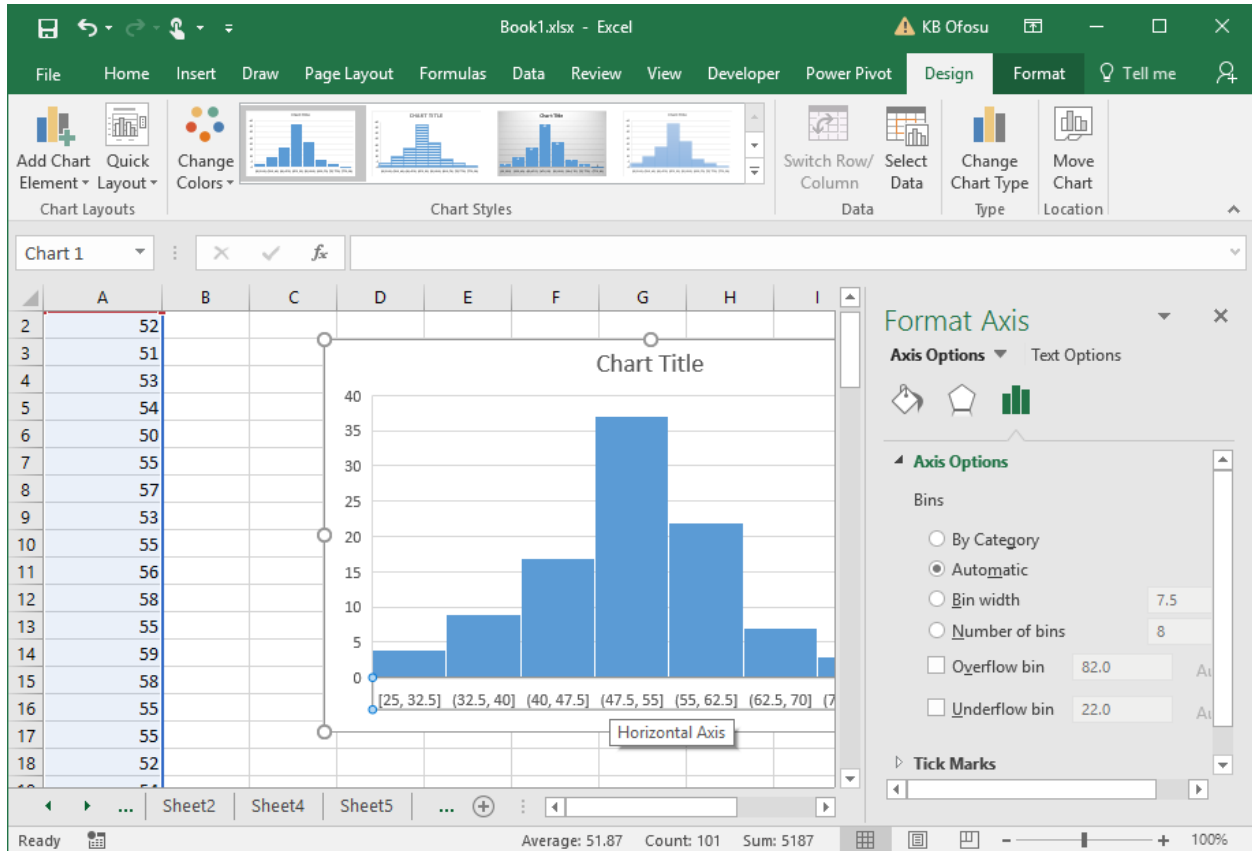


Notice that we did not have the opportunity to set the intervals (bins). The program inspected the data and chose appropriate bins for us. Let us change the bins.



Engineering Methods in Excel  
A SunCam online continuing education course

Double click anywhere on the horizontal axis.  
The **Format Axis** pane opens to the right.



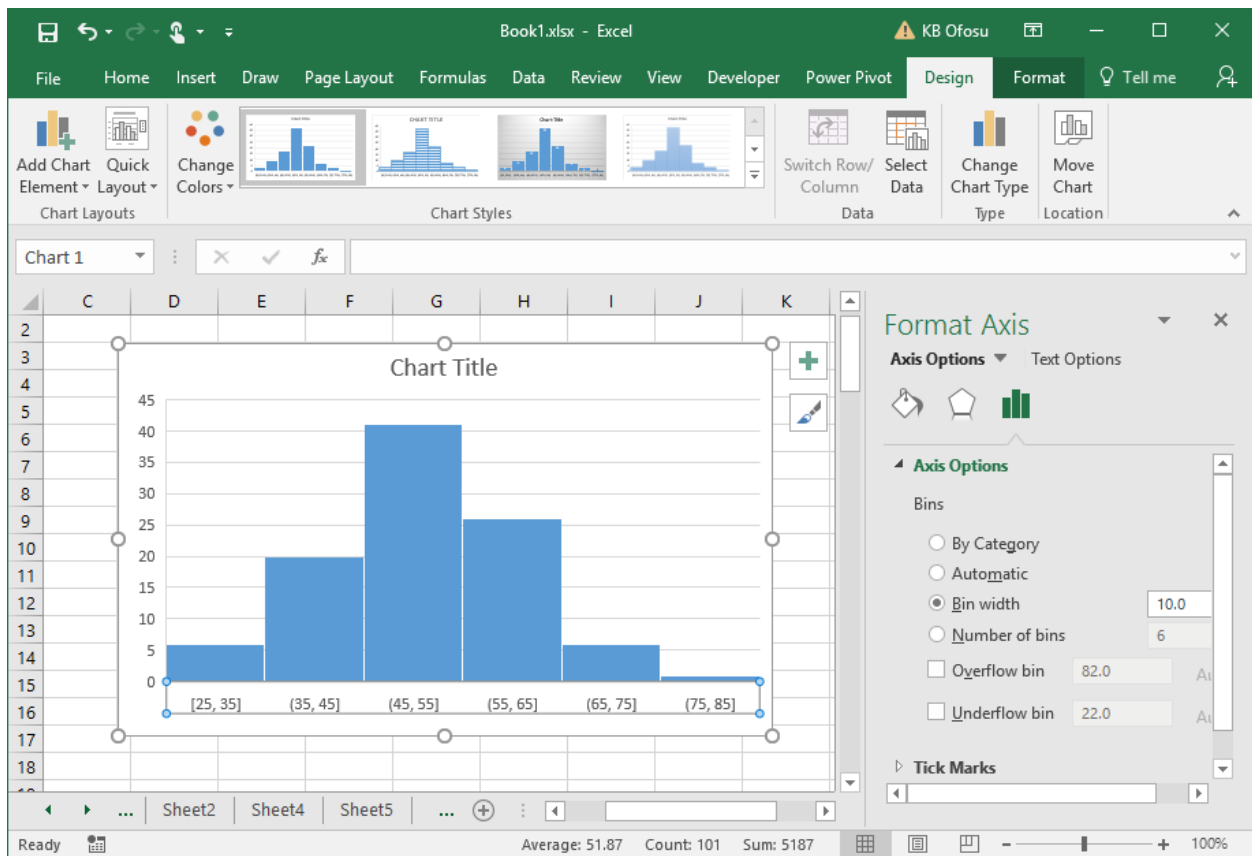


Engineering Methods in Excel  
A SunCam online continuing education course

Click on the **Bin width** option button, change the value to 10.  
Hit Enter on your keyboard.

Notice that the **Number of Bins** adjusts automatically.  
The histogram is reconfigured automatically.

Alternately, you may change the **Number of Bins**, and see the **Bin width** adjust automatically.



We shall see how to adjust other general features of charts as we proceed through this section.

This is the end of the Alternate Method for Histograms (*Excel 2016 and beyond users only*).

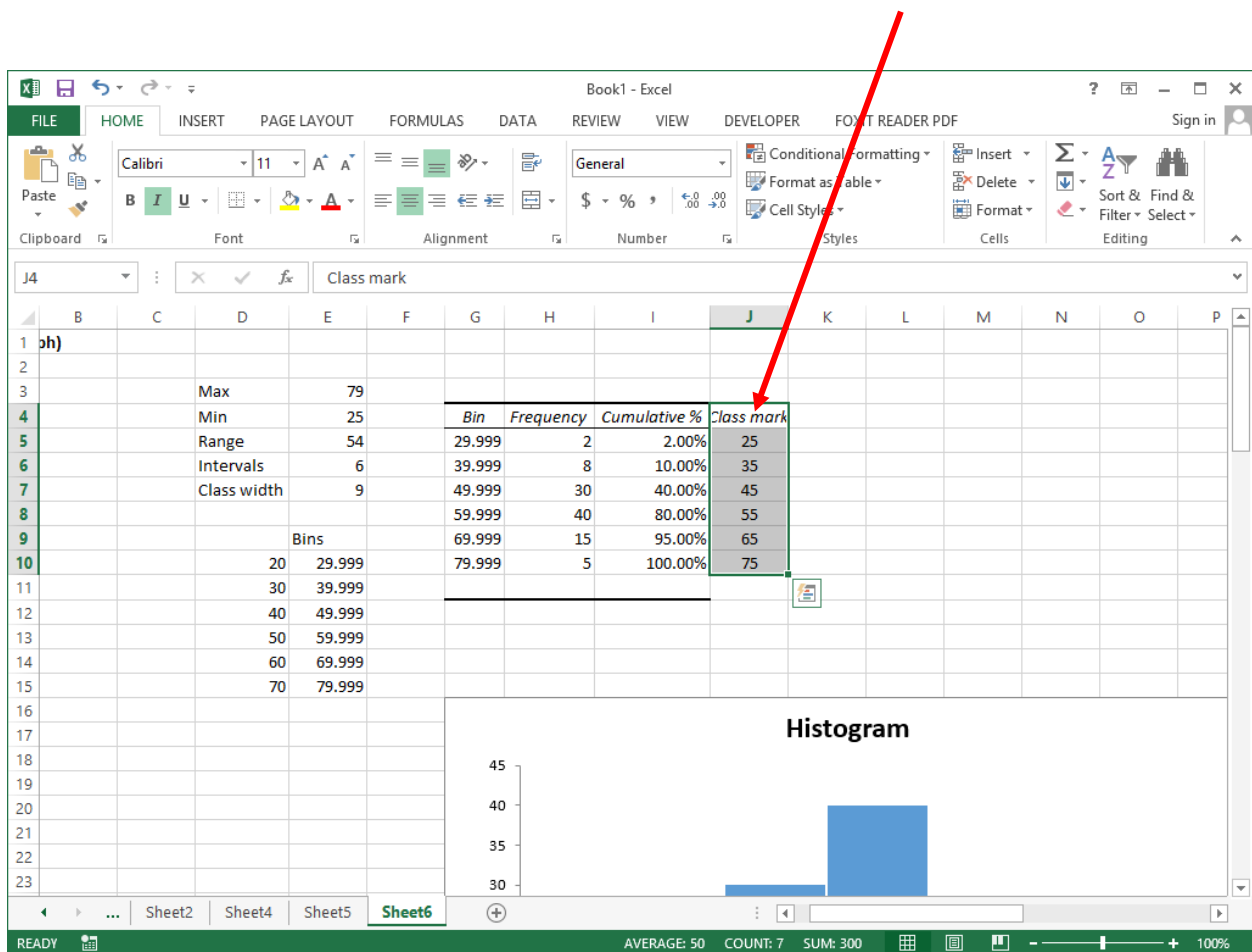


Engineering Methods in Excel  
A SunCam online continuing education course

Frequency Polygon

The frequency polygon can be plotted using the general chart tools available in *Excel*. It is simply the frequency versus the class mark (interval midpoint).

Update your frequency table or bins table to include the class mark for each interval.





Engineering Methods in Excel  
A SunCam online continuing education course

Click on the **INSERT** tab.

The screenshot shows the Microsoft Excel interface with the **INSERT** tab selected in the ribbon. A red arrow points to the **INSERT** tab. The spreadsheet data is as follows:

	Max	79			
4	Min	25	Bin	Frequency	Cumulative %
5	Range	54	29.999	2	2.00%
6	Intervals	6	39.999	8	10.00%
7	Class width	9	49.999	30	40.00%
8			59.999	40	80.00%
9			69.999	15	95.00%
10			79.999	5	100.00%

The histogram in the bottom right corner is titled "Histogram" and shows two bars. The first bar has a frequency of 2 and a class mark of 25. The second bar has a frequency of 8 and a class mark of 35.



Engineering Methods in Excel  
A SunCam online continuing education course

In the **Charts** palette, hover over the chart types and settle over the **Scatter or Bubble Chart** type.

The screenshot shows the Microsoft Excel interface with the **INSERT** tab selected. The **Charts** group is active, and a tooltip for **Insert Scatter (X, Y) or Bubble Chart** is displayed. The tooltip text reads: "Use this chart type to show the relationship between sets of values. Click the arrow to see the different types of scatter and bubble charts available and pause the pointer on the icons to see a preview in your document." A red arrow points to the Scatter chart icon in the Charts palette.

The spreadsheet data is as follows:

	B	C	D	E	F	G	H
1	ph)						
2							
3			Max	79			
4			Min	25	Bin	Frequ	
5			Range	54	29.999		
6			Intervals	6	39.999	8	10.00%
7			Class width	9	49.999	30	40.00%
8					59.999	40	80.00%
9			Bins		69.999	15	95.00%
10			20	29.999	79.999	5	100.00%
11			30	39.999			
12			40	49.999			
13			50	59.999			
14			60	69.999			
15			70	79.999			

The histogram chart, titled "Histogram", is located in the bottom right corner of the spreadsheet. It shows two bars: the first bar has a frequency of 5 (for bin 20-29.999) and the second bar has a frequency of 40 (for bin 30-39.999). The y-axis ranges from 30 to 45.



Engineering Methods in Excel  
A SunCam online continuing education course

Click on the **Scatter or Bubble Charts** drop down to view available options.

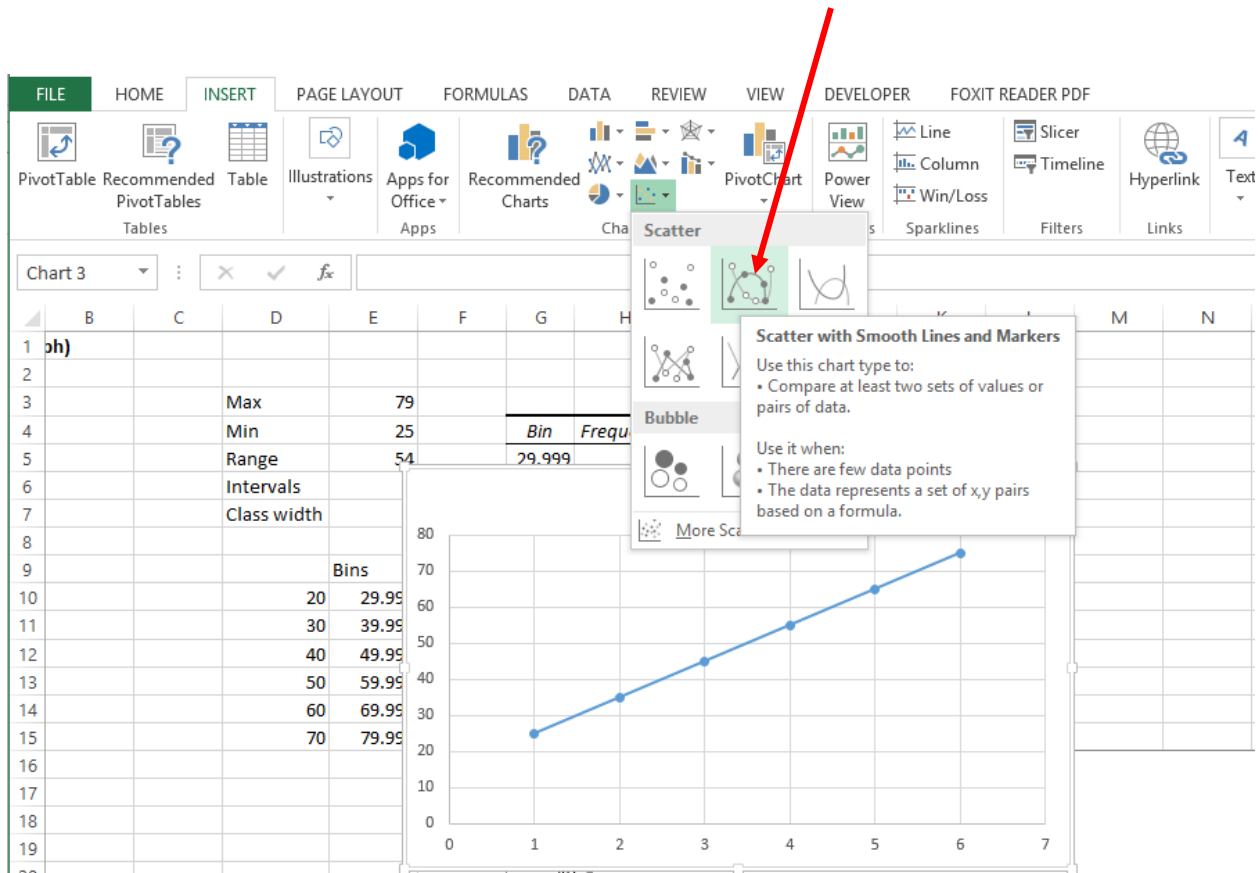
The screenshot shows the Microsoft Excel interface with the 'INSERT' tab selected. The 'Charts' group is expanded to show 'Scatter' and 'Bubble' options. The spreadsheet data is as follows:

	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ph)												
2													
3			Max		79								
4			Min		25	Bin	Frequ						
5			Range		54	29.999							
6			Intervals		6	39.999							
7			Class width		9	49.999							
8						59.999							
9				Bins		69.999	15	95.00%	65				
10			20	29.999		79.999	5	100.00%	75				
11			30	39.999									
12			40	49.999									
13			50	59.999									
14			60	69.999									
15			70	79.999									



Engineering Methods in Excel  
A SunCam online continuing education course

Select the option **Scatter with Smooth Lines and Markers**.

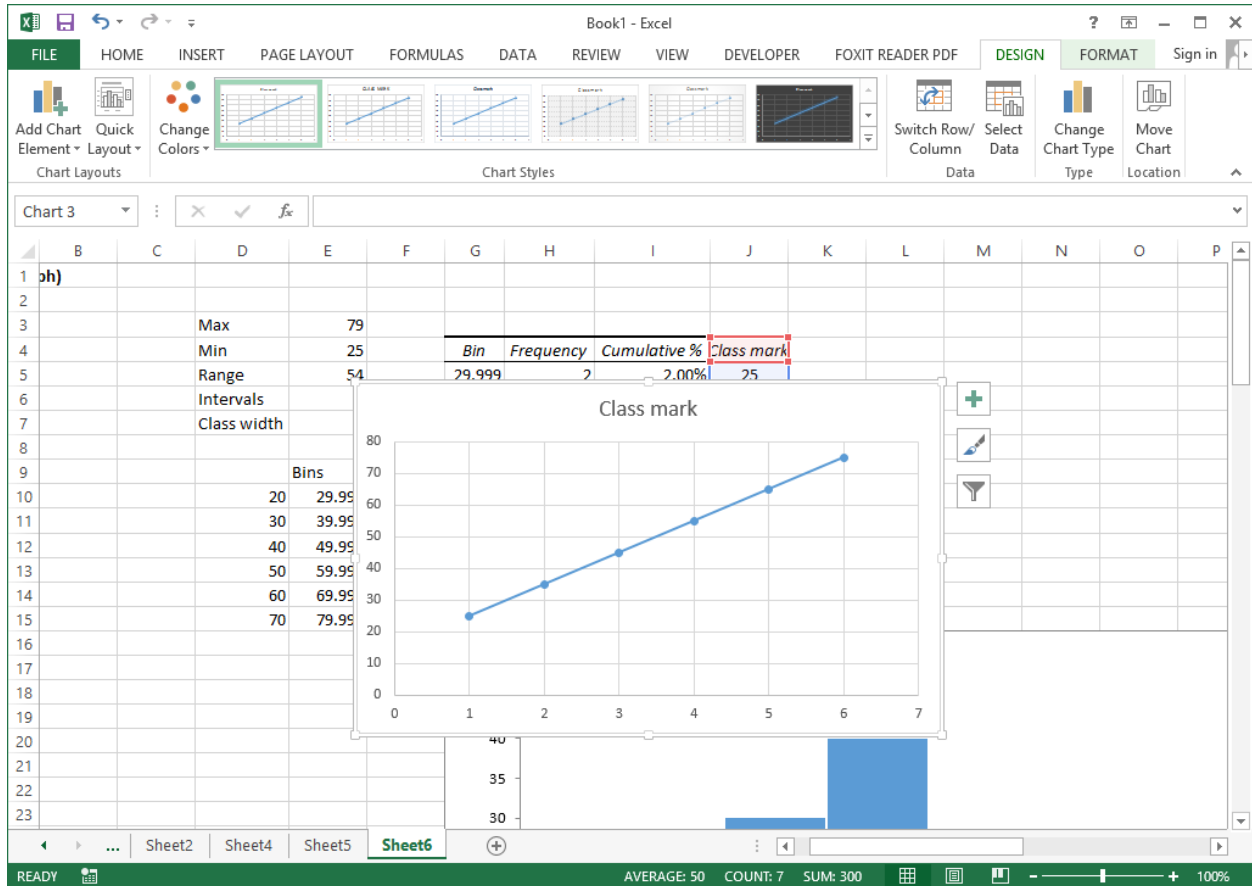






## Engineering Methods in Excel A SunCam online continuing education course

If any data was selected at the time the Scatter chart was created, that data will automatically be grabbed and plotted on the chart. We shall edit the chart to get it to show our correct data.





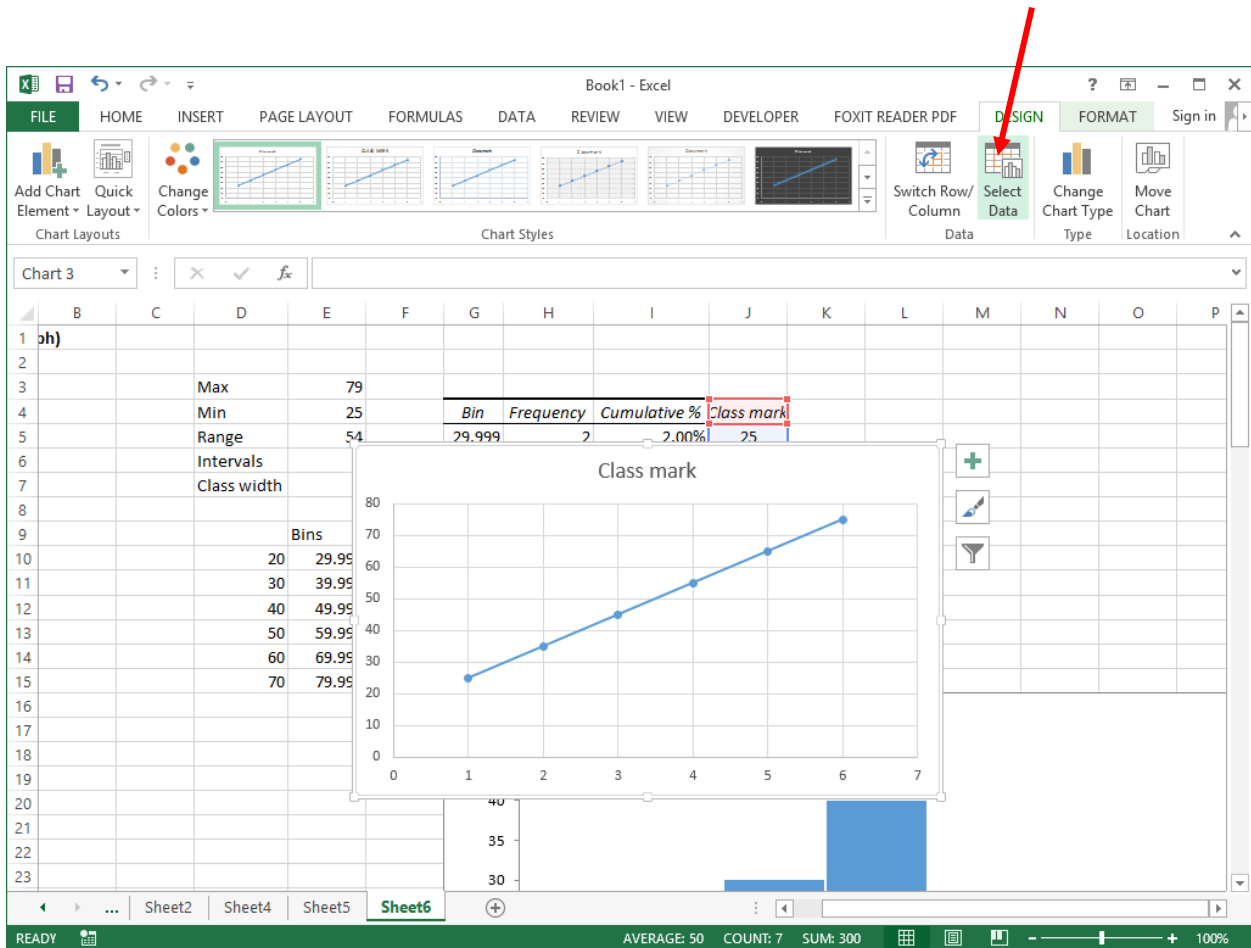
Engineering Methods in Excel  
A SunCam online continuing education course

Make sure the chart is selected, by clicking on it.

Go to **CHART TOOLS**.

Click on **DESIGN**

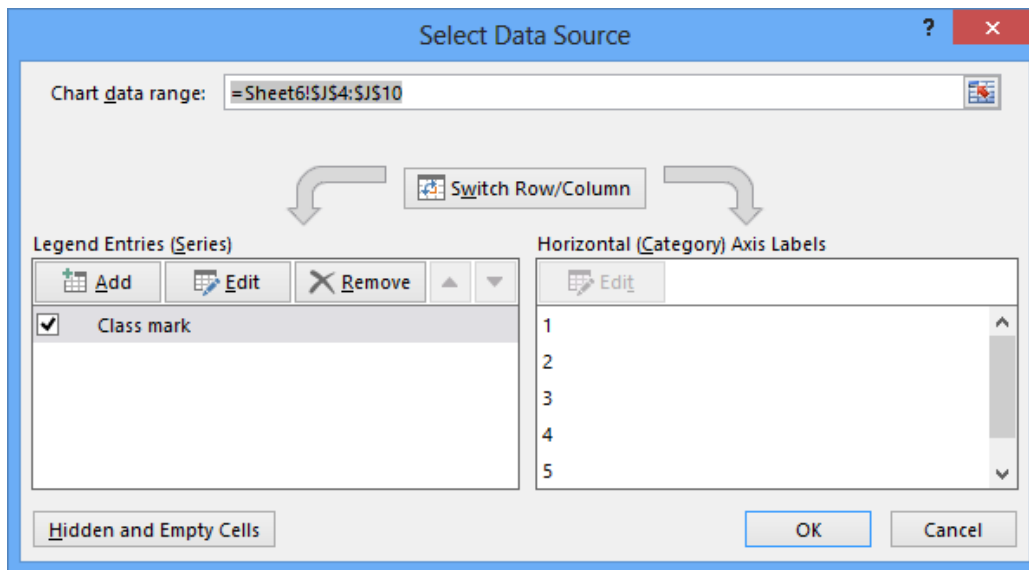
Click on **Select Data**.





Engineering Methods in Excel  
*A SunCam online continuing education course*

The **Select Data Source** window opens.

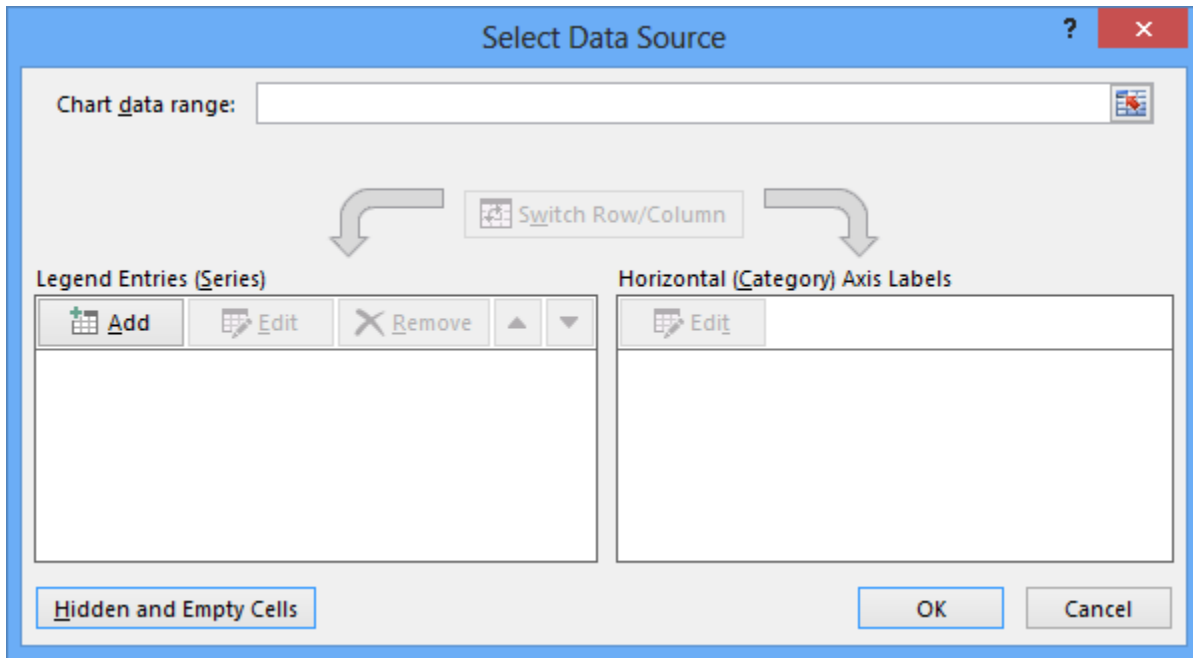


Delete the entry in the **Chart data range:** textbox.



Engineering Methods in Excel  
*A SunCam online continuing education course*

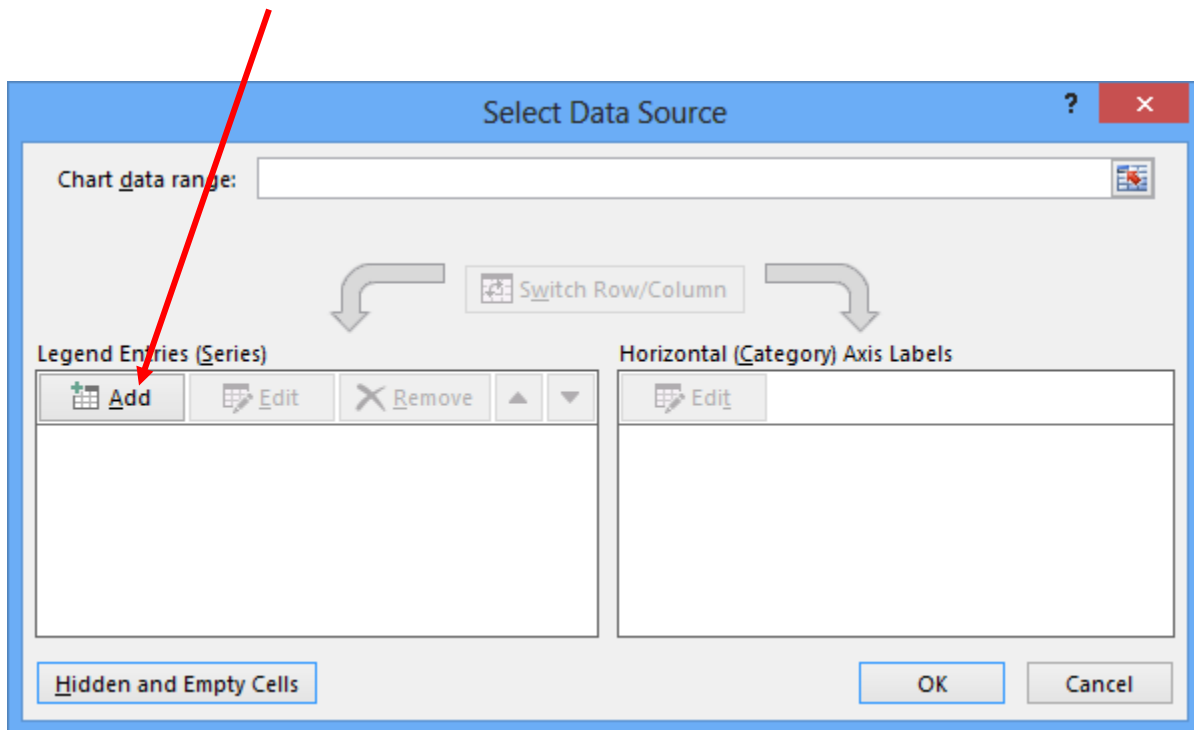
This clears the window. We shall now add our new data,





Engineering Methods in Excel  
*A SunCam online continuing education course*

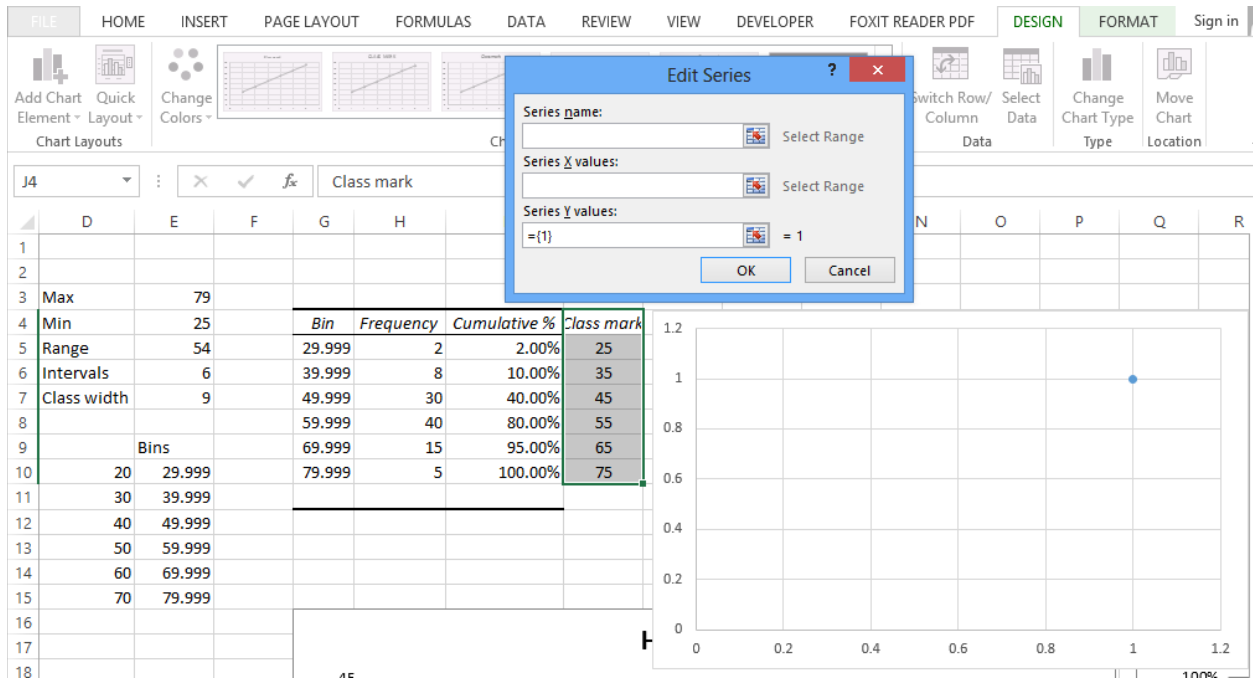
Click on **Add**.





Engineering Methods in Excel  
A SunCam online continuing education course

The **Edit Series** window opens.





Engineering Methods in Excel  
A SunCam online continuing education course

Click in the **Series X values:** textbox.

Go to the spreadsheet and select your X values; in our case, the class mark values.

The screenshot shows the Excel interface with the 'Edit Series' dialog box open. The dialog box has the following fields:

- Series name: (empty)
- Series X values: =Sheet6!\$J\$5:\$J\$10 = 25, 35, 45, 55...
- Series Y values: =1

The background spreadsheet shows a table with the following data:

			Bin	Frequency	Cumulative %	Class mark
3	Max	79				
4	Min	25				
5	Range	54	29.999	2	2.00%	25
6	Intervals	6	39.999	8	10.00%	35
7	Class width	9	49.999	30	40.00%	45
8			59.999	40	80.00%	55
9		Bins	69.999	15	95.00%	65
10		20	79.999	5	100.00%	75
11		30				
12		40				
13		50				
14		60				
15		70				

To the right of the table is a blank coordinate system with a vertical axis from 0 to 1.2 and a horizontal axis from 0 to 20.



Engineering Methods in Excel  
A SunCam online continuing education course

Click in the **Series Y values:** textbox.

Go to the spreadsheet and select your Y values; in our case, the frequency values.

The screenshot shows an Excel spreadsheet with a histogram chart. An "Edit Series" dialog box is open, showing the following settings:

- Series name: (empty)
- Series X values: =Sheet6!\$J\$5:\$J\$10 (range: = 25, 35, 45, 55, ...)
- Series Y values: =Sheet6!\$H\$5:\$H\$10 (range: = 2, 8, 30, 40, ...)

The spreadsheet data is as follows:

	D	E	F	G	H	I
1						
2						
3	Max		79			
4	Min		25			
5	Range		54			
6	Intervals		6			
7	Class width		9			
8						
9		Bins				
10	20	29.999				
11	30	39.999				
12	40	49.999				
13	50	59.999				
14	60	69.999				
15	70	79.999				
16						
17						
18						

The histogram chart shows a distribution with a peak at approximately 55. The x-axis ranges from 0 to 80, and the y-axis ranges from 0 to 45.



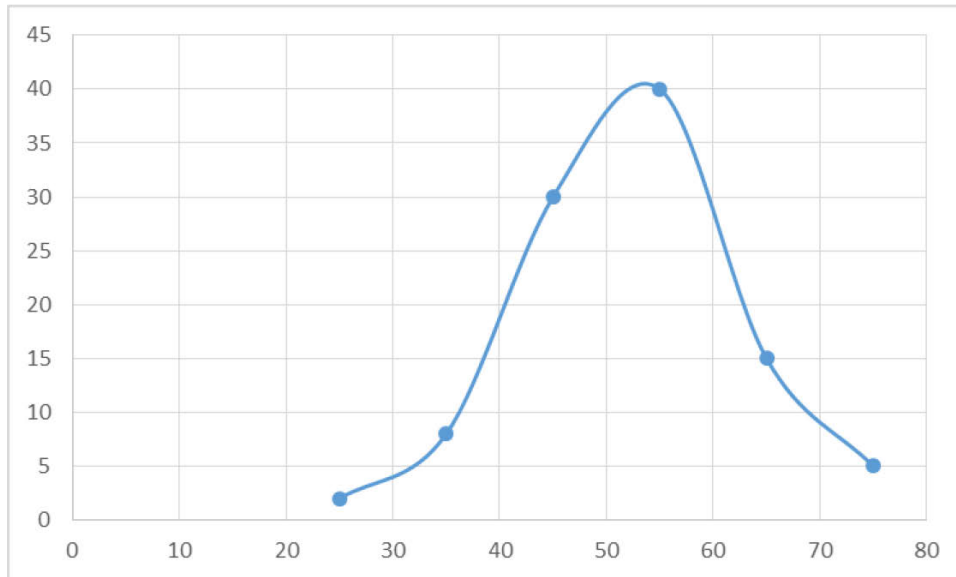


Engineering Methods in Excel  
*A SunCam online continuing education course*

Hit **OK** to exit out of the **Edit Series** window.

Hit **OK** to exit out of the **Select Data Source** window.

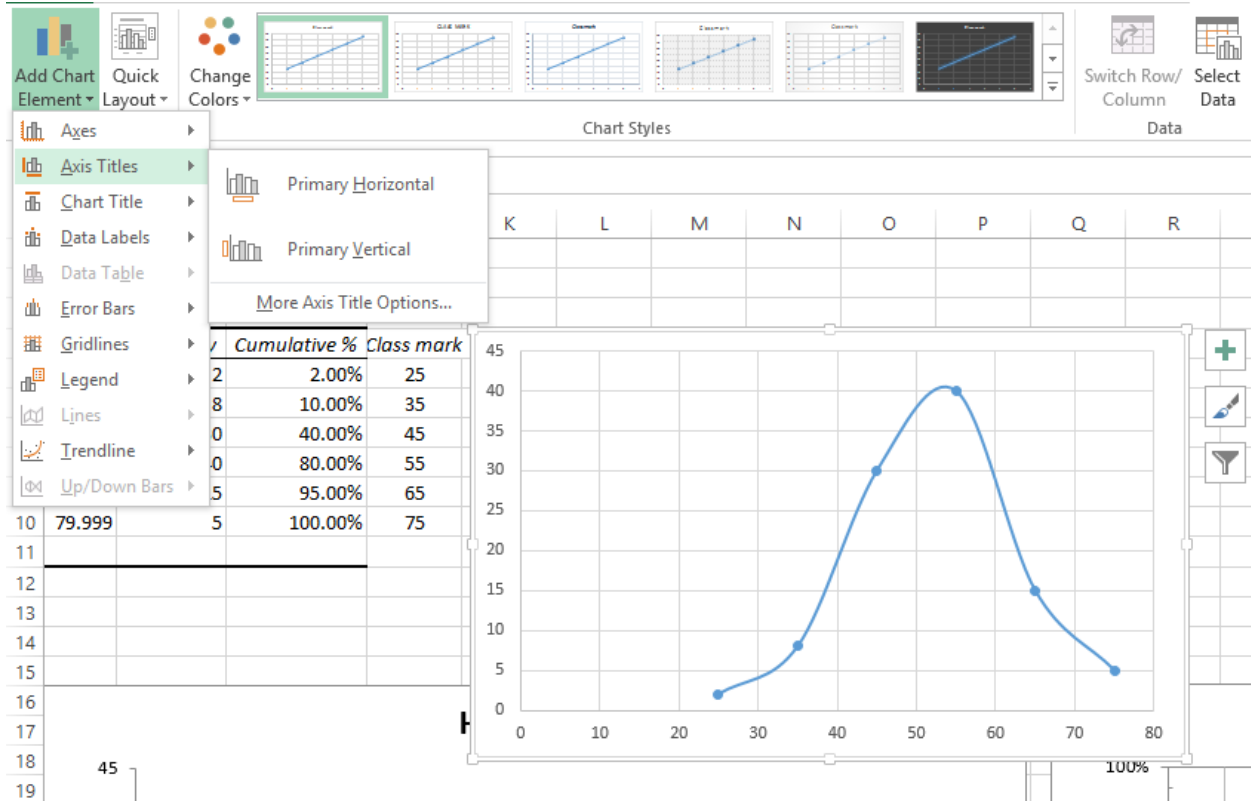
View your frequency polygon.





Engineering Methods in Excel  
A SunCam online continuing education course

On your own, use the **Add Chart Elements** tools to add labels on the vertical and horizontal axes.



On your own, if desired, you may change the axes formats, and gridlines. You may also add a title and a legend to the chart.

Reposition (and resize) your chart elsewhere on the spreadsheet.

Frequency polygon complete!



Engineering Methods in Excel  
A SunCam online continuing education course

Dotplot

A dotplot is usually drawn vertically. In this exercise, we shall draw one horizontally—but set it up such that it can be rotated to the vertical upon pasting in reports etc. In *Excel*, the dot plot can be drawn directly on the spreadsheet with the innovative use of built-in functions—without using any graphing or chart functionality.

Set up the frequency table for the speed data we have been working on as shown:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)	Freq													
2	20 - 29.99	2													
3	30 - 39.99	8													
4	40 - 49.99	30													
5	50 - 59.99	40													
6	60 - 69.99	15													
7	70 - 79.99	5													



## Engineering Methods in Excel A SunCam online continuing education course

We shall use the built-in REPT function to repeat the dots the requisite number of times. The REPT function repeats a text string a specified number of times. The dot (“.”) is *Excel* character number 149. We can specify this using the built-in CHAR function. So we shall repeat CHAR(149) a certain number of times based on the frequency.

For the first interval of the speed distribution type in the formula as follows:

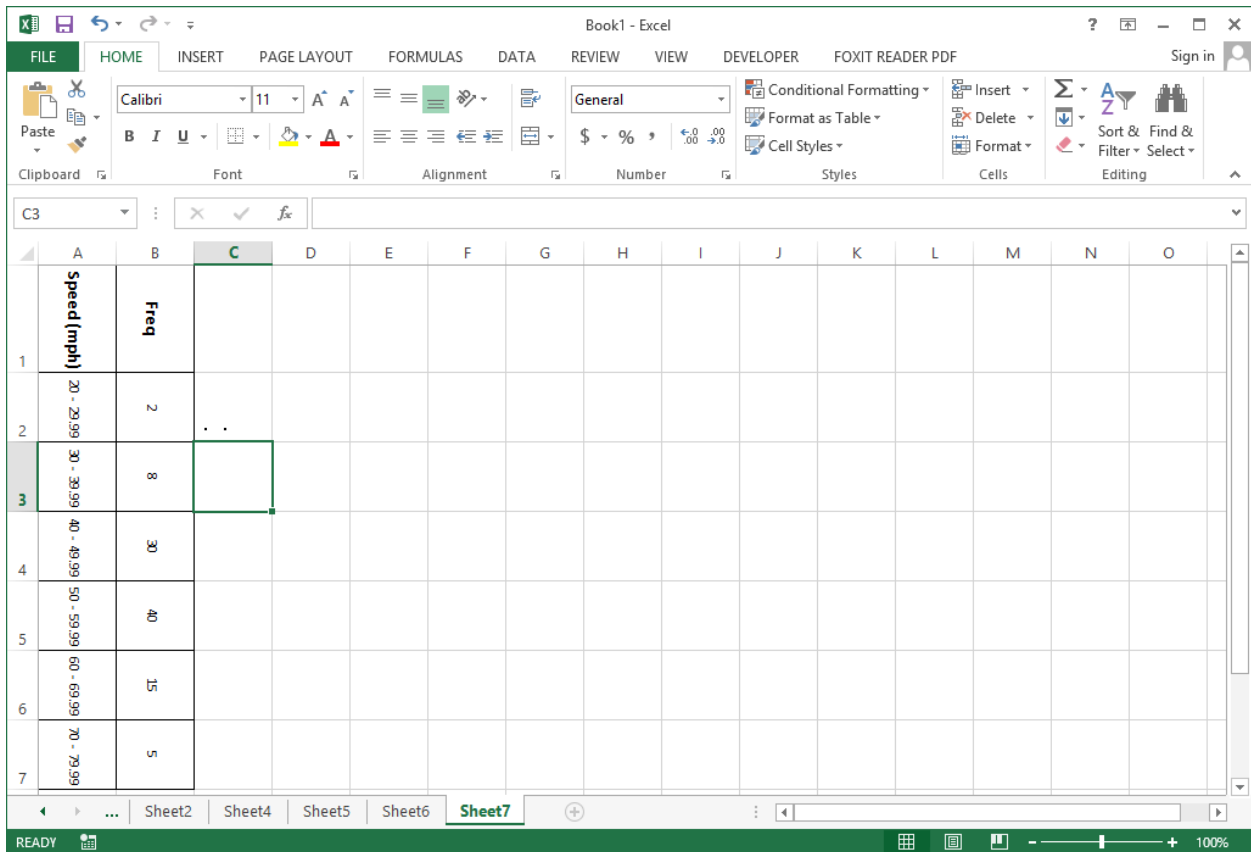
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)	Freq													
2	20 - 29.99	2	=REPT(CHAR(149),B2)												
3	30 - 39.99	8													
4	40 - 49.99	30													
5	50 - 59.99	40													
6	60 - 69.99	15													
7	70 - 79.99	5													



Engineering Methods in Excel  
A SunCam online continuing education course

Hit Enter.

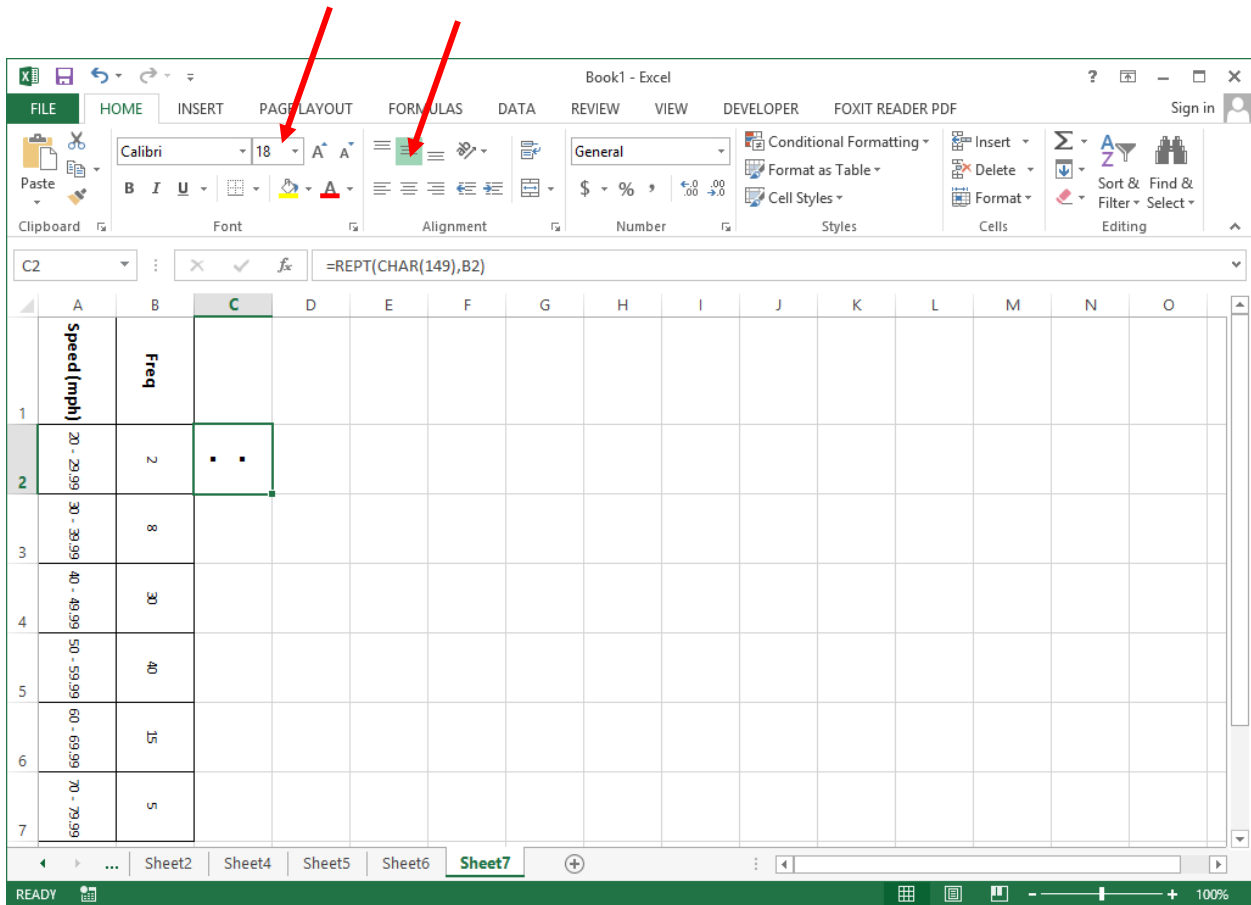
Two (2) dots are plotted representing the frequency of 2 for the interval.





Engineering Methods in Excel  
A SunCam online continuing education course

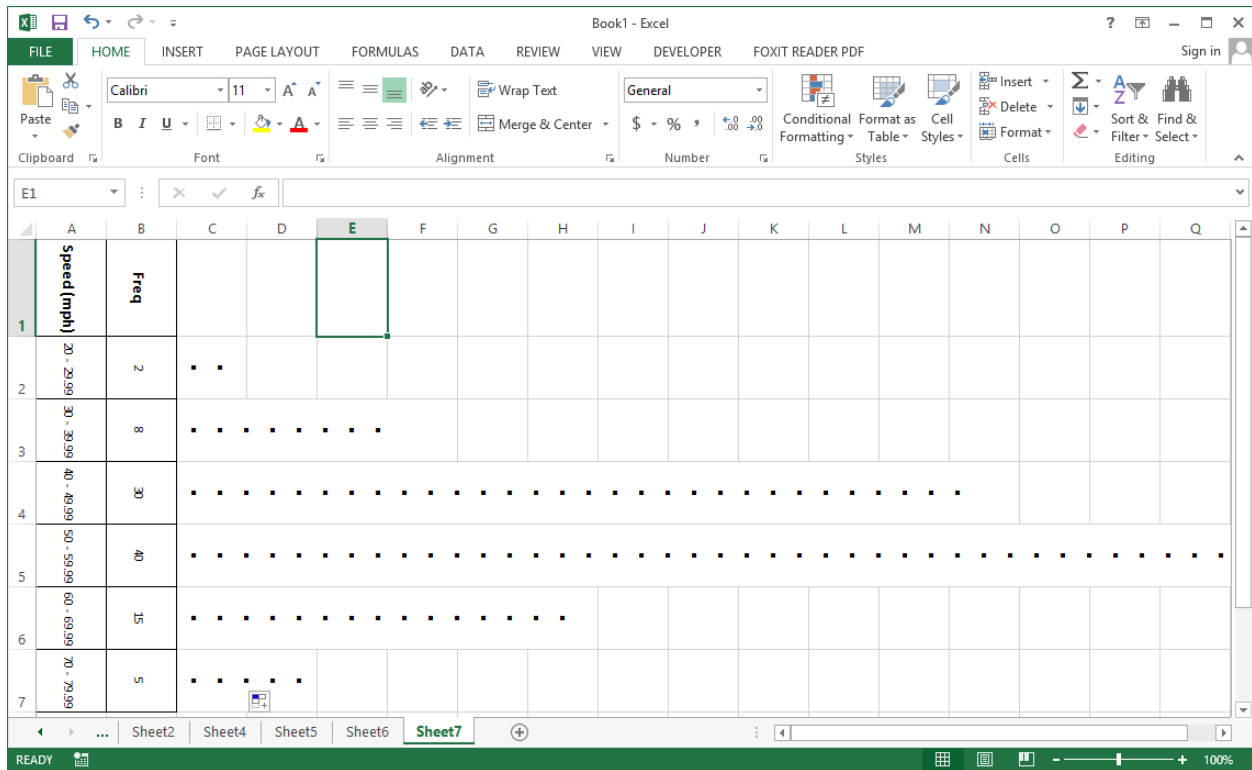
- Click on the **HOME** tab.
- Select the cell holding the dots.
- Increase font size to 18.
- Align the contents vertically to the center.





Engineering Methods in Excel  
A SunCam online continuing education course

Fill handle for the rest of the intervals of the frequency distribution.



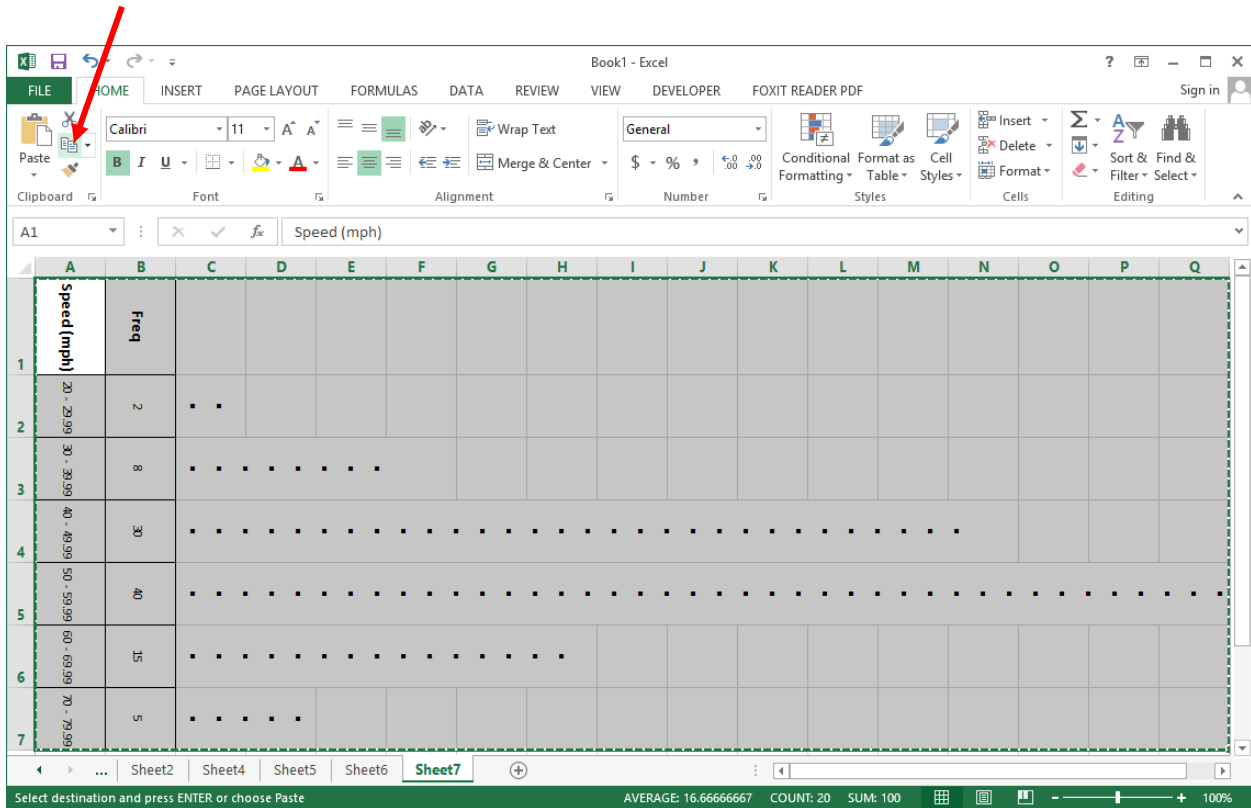
If necessary, increase the width of the column containing the dots. The reason we drew our dotplot with a horizontal orientation is because widening a column has more flexibility than increasing the height of a row—which has an upper limit of 409 units. For small datasets, this may not be relevant; however, for larger datasets, the horizontal orientation is advantageous.

We shall now copy the dotplot into a report.



Engineering Methods in Excel  
A SunCam online continuing education course

Select all cells of the dotplot.  
Click on the **HOME** tab.  
Click on **Copy**.

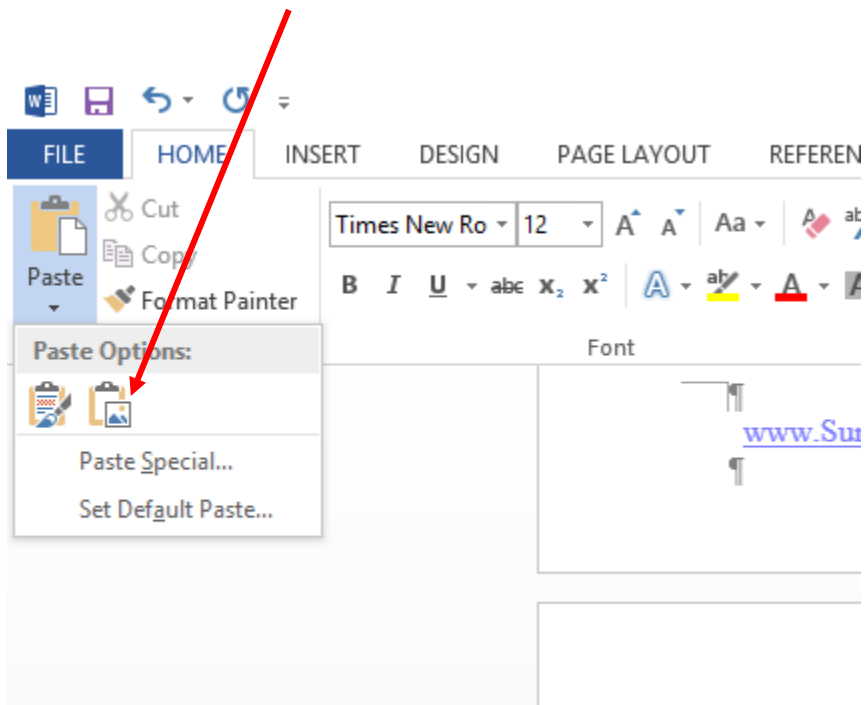






Engineering Methods in Excel  
*A SunCam online continuing education course*

- Open a *Word* document.
- Click on the **HOME** tab.
- Click on the **Paste** drop down.
- Click on **Paste Picture**.





Engineering Methods in Excel  
*A SunCam online continuing education course*

The dot plot is pasted over.

Speed (mph)	Freq																			
20 - 29.99	2	•	•																	
30 - 39.99	8	•	•	•	•	•	•	•	•											
40 - 49.99	30	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
50 - 59.99	40	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
60 - 69.99	15	•	•	•	•	•	•	•	•	•	•									
70 - 79.99	5	•	•	•	•	•														



Engineering Methods in Excel  
*A SunCam online continuing education course*

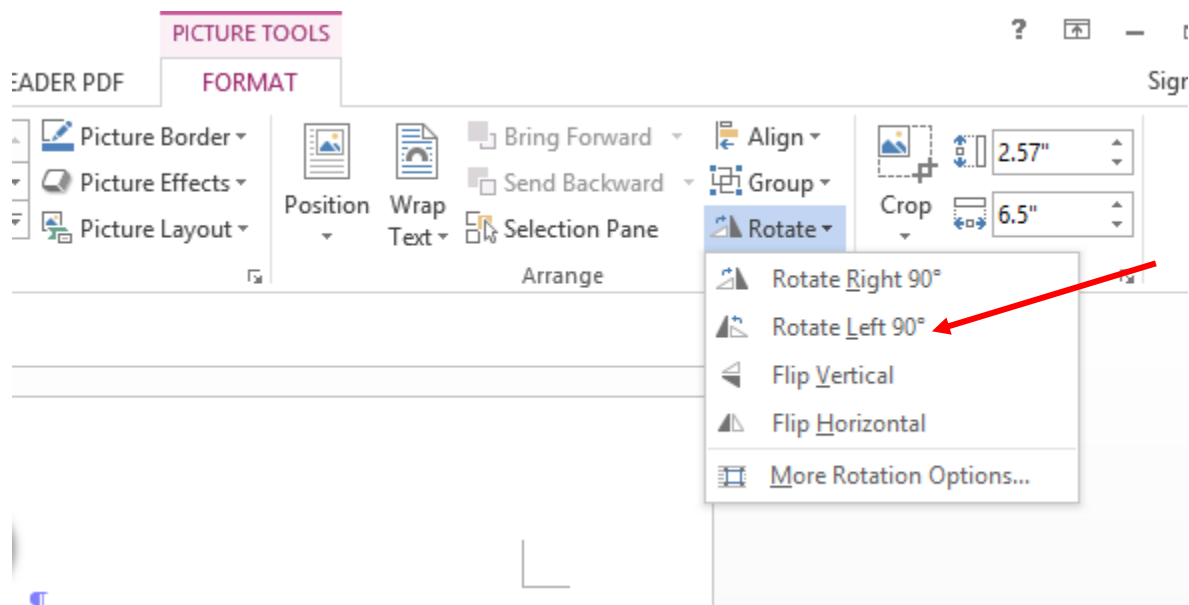
In the *Word* document, click on the dotplot to select it.

Got to **PICTURE TOOLS**.

Click on **FORMAT**.

Click on the **Rotate** drop down.

Click on **Rotate Left 90°**.



The diagram is rotated accordingly.





Engineering Methods in Excel  
*A SunCam online continuing education course*

Successful completion of the dotplot!

Stem-and-Leaf Diagram

We shall apply the innovative use of built-in spreadsheet functions to construct the stem-and-leaf diagram for our speed data on the spreadsheet—without using any chart or graphing functionality.

The COUNTIF function counts the number of cells that meet a given condition. The function is of the form:

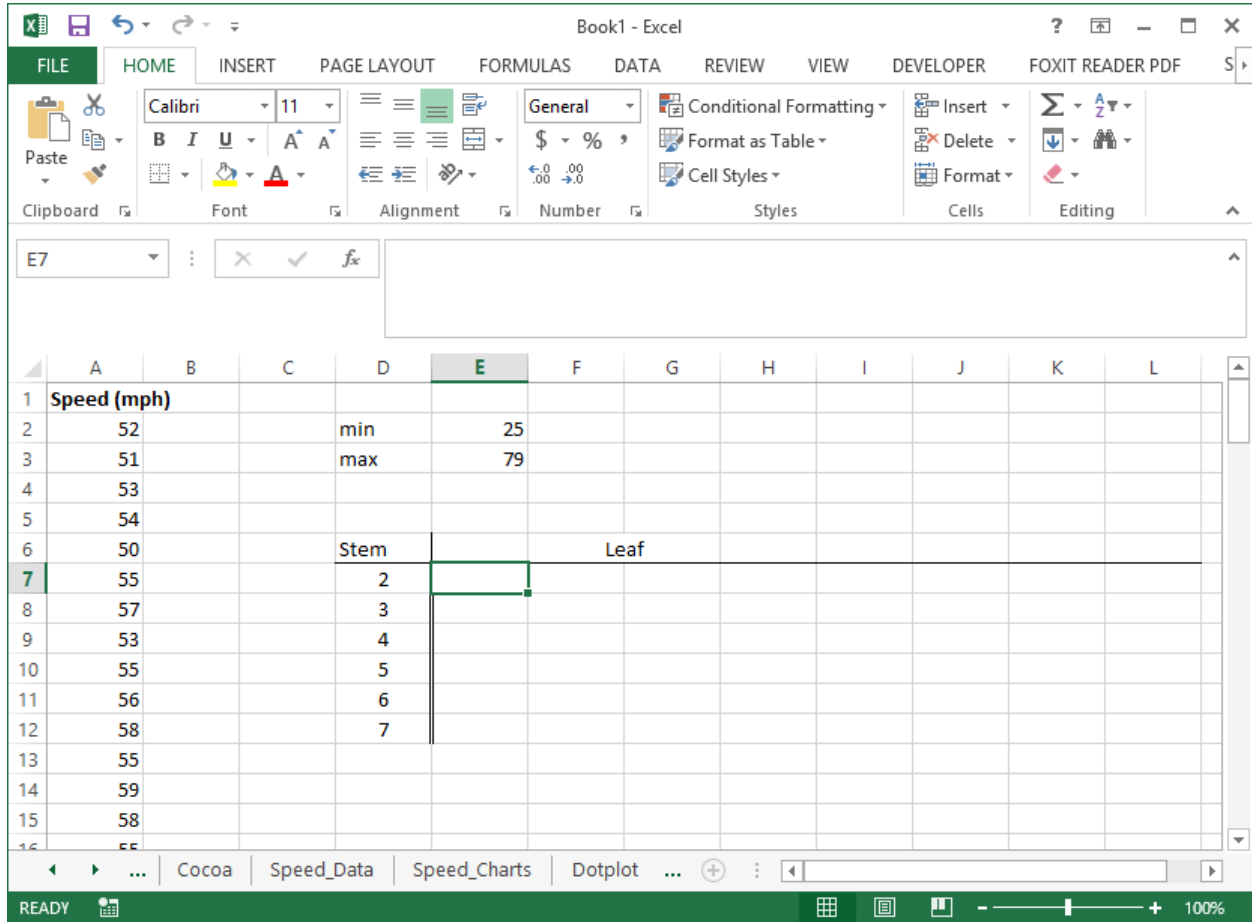
COUNTIF(*range of cells, condition*)

We shall use the REPT function to repeat a given string a specified number of times. The string to be repeated will be the digit representing a data value in the leaf, and the number of times it will be repeated will be obtained from the COUNTIF function.

Based on our speed data we shall adopt the following stem structure for our stem-and-leaf diagram:



Engineering Methods in Excel  
A SunCam online continuing education course





Engineering Methods in Excel  
A SunCam online continuing education course

Let us use the stem values (and their cell references) to define (or identify) a data value. For example, the value 25 can be represented as:

$(stem\ value)*10 + 5$ , or in our spreadsheet:

$D7*10 + 5$

For our speed data, the number of times this value occurs in the data is given by:

$COUNTIF(A2:A101, D7*10 + 5)$

We then use this number of occurrences to drive the number of times the digit “5” (representing 25 in the leaf) will be repeated in the leaf. Therefore:

$REPT(“5”, COUNTIF(A2:A101, D7*10 + 5))$

For the value 26, we would have:

$REPT(“6”, COUNTIF(A2:A101, D7*10 + 6))$ ,

and so on.

To set up any row of the stem-and-leaf diagram we will therefore “weld” together the (REPT) strings for each value in that range. The text function “&” (ampersand) can do this “welding” called concatenation of strings, for us.

So, the output string for the bracket 20 through 29 on the stem-and-leaf diagram will be:

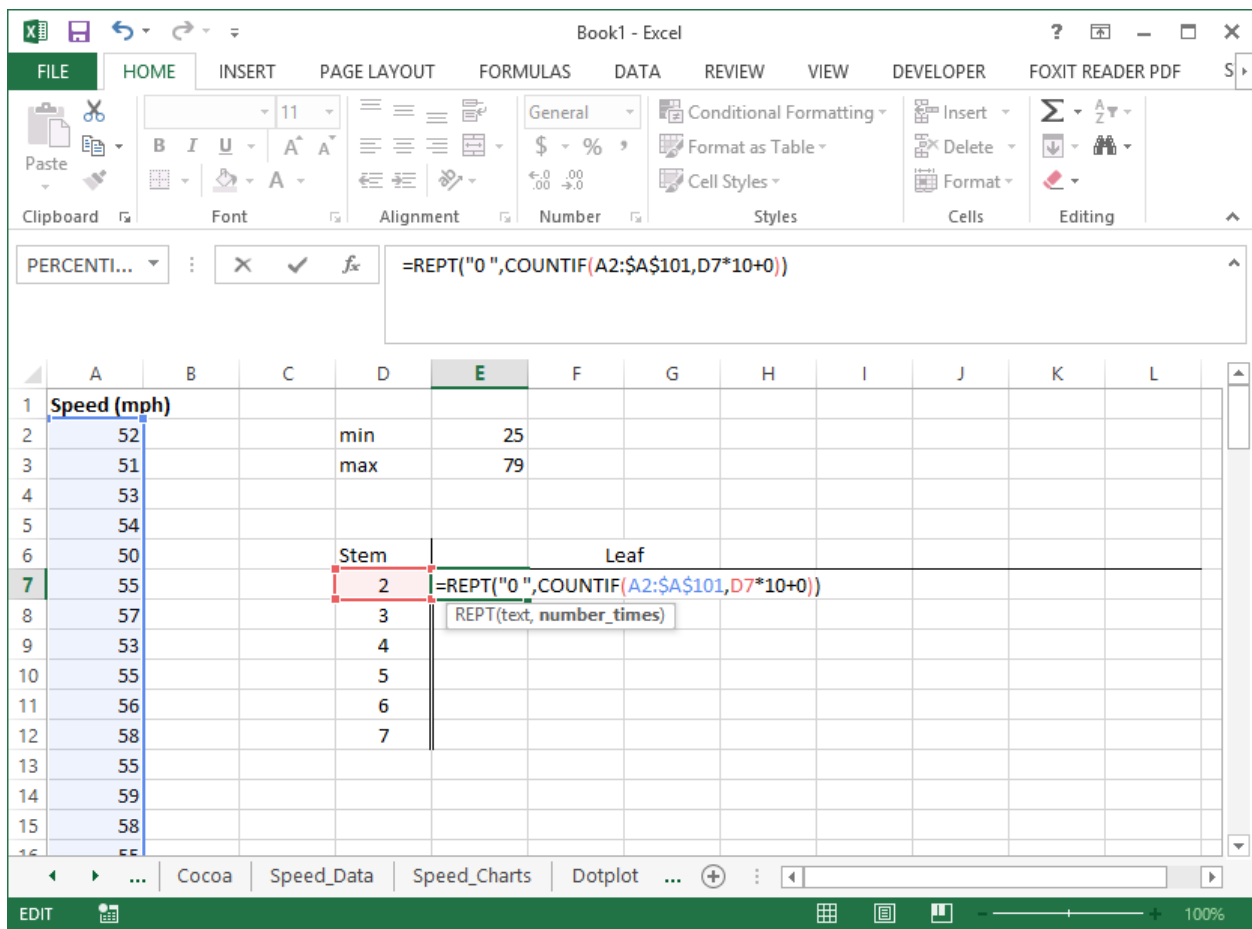
$REPT(“0”, COUNTIF(A2:A101, D7*10 + 0)) \& REPT(“1”, COUNTIF(A2:A101, D7*10 + 1))$   
 $\& REPT(“2”, COUNTIF(A2:A101, D7*10 + 2)) \& \dots \& REPT(“9”, COUNTIF(A2:A101,$   
 $D7*10 + 9))$



Engineering Methods in Excel  
A SunCam online continuing education course

To implement this on the spreadsheet, we may use point-and-click -or- type in the cell references. Be sure to apply absolute referencing (**F4**) to the speed data range of cells, so if we use fill handle later—the range of cells holding the original data stays fixed.

The value 20 is implemented as follows:



Note that when typing the “0”, and a space or two before the closing parenthesis (“0 “). This spaces out the digits and will improve the visual appearance of the digits in the leaf on the output diagram.





Engineering Methods in Excel  
A SunCam online continuing education course

Continuing with the other values in the range 20 through 29, we may continue typing -or- we may copy and paste, and then modify as we move through the range of values. Thus, we obtain:

The screenshot shows an Excel spreadsheet with a stem-and-leaf plot for speed data. The data is as follows:

Speed (mph)	Stem	Leaf
52	5	2
51	5	1
53	5	3
54	5	4
50	5	0
55	5	5
57	5	7
53	5	3
55	5	5
56	5	6
58	5	8
55	5	5
59	5	9
58	5	8

The formula bar for cell E7 shows the following formula:

```
=REPT("0 ",COUNTIF(A2:$A$101,D7*10+0))&REPT("1 ",COUNTIF($A$2:$A$101,D7*10+1))&  
REPT("2 ",COUNTIF($A$2:$A$101,D7*10+2))&REPT("3 ",COUNTIF($A$2:$A$101,D7*10+3))&  
REPT("4 ",COUNTIF($A$2:$A$101,D7*10+4))&REPT("5 ",COUNTIF($A$2:$A$101,D7*10+5))&
```



Engineering Methods in Excel  
A SunCam online continuing education course

Hit Enter.

We obtain the first leaf for the values 20 through 29 that are present in the data.

The screenshot shows an Excel spreadsheet with a stem-and-leaf plot. The data is organized as follows:

Row	Column A	Column B	Column C	Column D	Column E	Column F	Column G	Column H	Column I	Column J	Column K	Column L
1	Speed (mph)											
2	52			min	25							
3	51			max	79							
4	53											
5	54											
6	50			Stem	Leaf							
7	55			2	5 8							
8	57			3								
9	53			4								
10	55			5								
11	56			6								
12	58			7								
13	55											
14	59											
15	58											
16	55											



Engineering Methods in Excel  
A SunCam online continuing education course

Fill handle the first leaf to obtain the remaining leaves.

The screenshot shows an Excel spreadsheet with the following data and formulas:

Row	A	B	C	D	E	F	G	H	I	J	K	L
1	Speed (mph)											
2	52			min	25							
3	51			max	79							
4	53											
5	54											
6	50			Stem	Leaf							
7	55			2	5 8							
8	57			3	0 1 5 5 7 8 8 9							
9	53			4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9							
10	55			5	0 0 0 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 8 8 9 9 9							
11	56			6	0 0 0 1 1 2 2 2 2 3 3 3 3 4 6 9							
12	58			7	0 2 2 5 9							
13	55											
14	59											
15	58											

The formula bar for cell E7 contains the following formula:

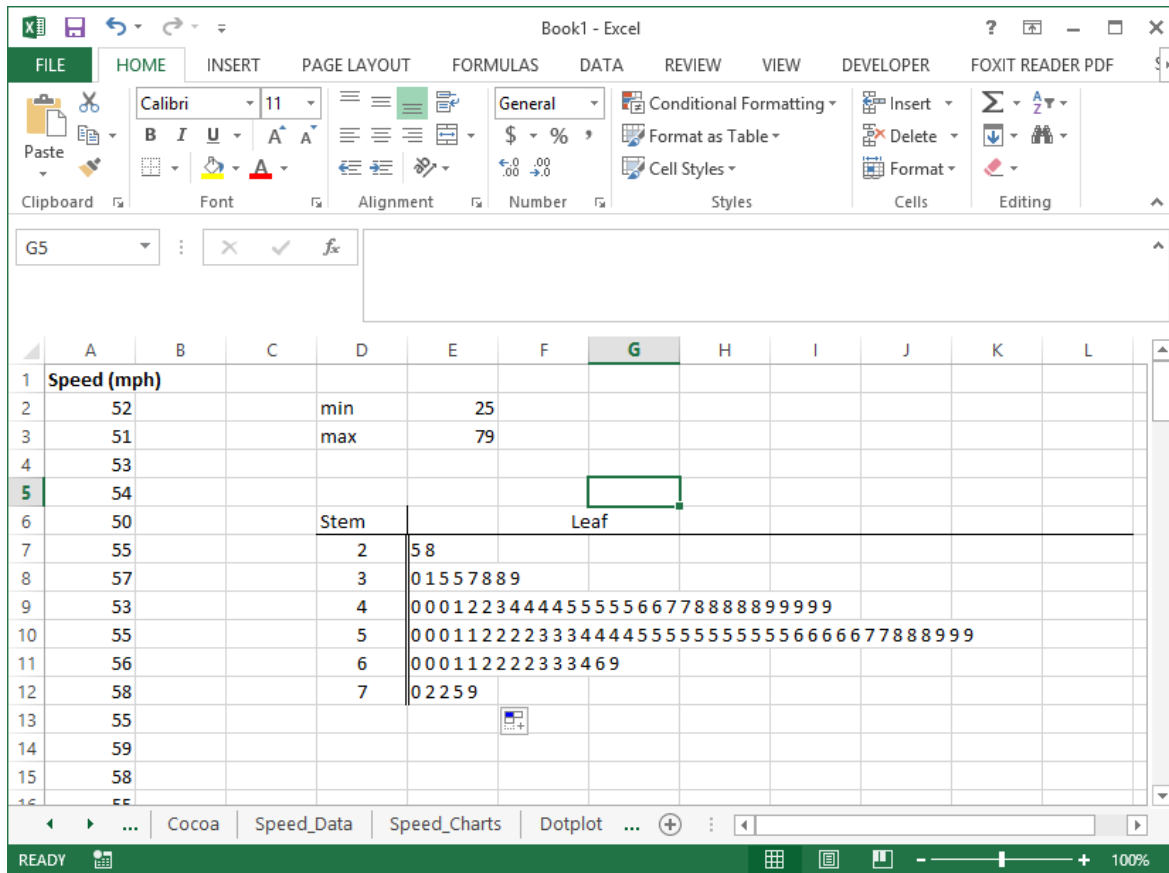
```
=REPT("0 ",COUNTIF(A2:$A$101,D7*10+0))&REPT("1 ",COUNTIF($A$2:$A$101,D7*10+1))&REPT("2 ",COUNTIF($A$2:$A$101,D7*10+2))&REPT("3 ",COUNTIF($A$2:$A$101,D7*10+3))&REPT("4 ",COUNTIF($A$2:$A$101,D7*10+4))&REPT("5 ",COUNTIF($A$2:$A$101,D7*10+5))
```



Engineering Methods in Excel  
A SunCam online continuing education course

Click away from the diagram.

Now we see that the stem-and-leaf diagram is essentially a depiction of the data sorted in numerical order. This is advantageous, and enables the stem-and-leaf diagram to assist us in conducting other statistical analyses we shall discuss later in this course. However, we must keep in mind that the stem-and-leaf diagram may become complicated and tedious to construct when the data involves decimal values.



Stem-and-leaf diagram complete!

In later sections of this course, we shall see some of the uses of some of the frequency diagrams and how engineers use them to make decisions.



Engineering Methods in Excel  
A SunCam online continuing education course

### 2.7 Descriptive Statistics

In the previous section, we saw how the frequency distribution (table) and the frequency diagrams provide a tabular summary and graphical summary of a dataset respectively. Another way of summarizing data is by using numerical measures called **descriptive statistics**. One group of such measures, represents the center or middle of the data (or distribution). These are called **measures of central tendency** or **measures of center**.

#### Mode

The first measure of center we shall look at is called the **mode**. The mode is the data value with the highest frequency.

Using our speed data as an example, we can determine the mode directly by looking at the stem-and-leaf diagram.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 <b>5 5 5 5 5 5 5 5 5</b> 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

The mode of this data is 55 mph, with the high frequency of 11.

In some cases, a dataset may have more than one value with the high frequency. If there is a two-way-tie, the data has two modes, and the distribution is described as being bimodal. If there are multiple values tied for the high frequency, then the data is said to be multimodal. However, if all values in the data have equal frequency, then the data has no mode.

#### Median

The **median** is the value at the middlemost position when the data are arranged in order from smallest to largest. In other words, the median value has an equal number of data values above it and below it.



Engineering Methods in Excel  
A SunCam online continuing education course

In general, if  $n$  is the number of observations (sample size),

Then if  $n$  is an odd number, then the median ( $\tilde{x}$ ) is the observation at position  $\frac{n+1}{2}$

If  $n$  is an even number, then the median is the average of the observations at positions  $\frac{n}{2}$   
and  $\frac{n+2}{2}$

For our speed data, we have  $n = 100$ , which is even.

$$\text{Position } \frac{n}{2} = \frac{100}{2} = 50$$

$$\text{Position } \frac{n+2}{2} = \frac{102}{2} = 51$$

Recall that the stem-and-leaf diagram shows the entire data arranged in increasing numerical order. We can therefore find the values at the relevant positions for determining the median by inspecting the stem-and-leaf diagram.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

Position 50  
Position 51

$$\tilde{x} = \frac{53+53}{2} = 53 \text{ mph}$$

### Mean

The **mean** or **arithmetic mean** is defined as the summation of the data values in the dataset, divided by the number of values in the dataset.



Engineering Methods in Excel  
A SunCam online continuing education course

Let  $x_i$  denote a data value in our dataset, then the mean ( $\bar{x}$ ) is:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots(2.1)$$

Thus, for our speed data, the mean speed:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{52+51+53+54+\dots}{100} = \mathbf{51.87mph}$$

Median versus Mean

In many applications, the question arises as to which measure of center is the better one to use. Unfortunately there is no simple answer to this question. It depends on your data and what you are doing with it. In many cases the values will coincide for a dataset. The median is more of a descriptive measure and is considerably less computationally intensive. The median is also less impacted by the presence of anomalous data, data errors etc. (commonly called outliers), than the mean.

Measures of Dispersion

Another group of descriptive statistics are called **measures of dispersion**, or **measures of variability** (or variation). Measures of dispersion measure the amount of variability (or variation) of the values in the data.

For a data value  $x_i$ , let us define the difference of the data value and the mean of the data as the **deviation** of the value  $x_i$  from the mean. So:

$$\text{Deviation} = (x_i - \bar{x}) \dots\dots\dots(2.2)$$



Engineering Methods in Excel  
A SunCam online continuing education course

Let us define the **sample variance** ( $s^2$ ) as:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots\dots\dots(2.3)$$

If we take the square root of the sample variance, we obtain the **sample standard deviation** ( $s$ ):

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \dots\dots\dots(2.4)$$

In other words, the sample standard deviation is the “average” of the deviations of all of the values in the data. The sample variance and the sample standard deviation involve some computational effort, even for small samples, and they cannot be pulled from any chart or graph as we saw with measures of center and frequency diagrams.

It is pertinent to note that the formula for the sample variance has a denominator of  $(n - 1)$ . This is used when the variance of a sample is being computed. If the data is for a population then the denominator used is  $n$ , resulting in a population variance being computed, and hence a population standard deviation.

### 2.8 Quartiles and Percentiles

Quartiles are data values that split the distribution into four quarters—when the data values are arranged in increasing order. The **1<sup>st</sup> quartile** (also called the **lower quartile**) ( $Q_1$ ) has one-fourth or 25% of the data below it, and three quarters or 75% of the data above it—when the data is arranged in ascending order. The **2<sup>nd</sup> quartile** (also called the **middle quartile**) ( $Q_2$ ) has two-fourths or 50% of the data below it, and two-fourths or 50% of the data above it—when the





Engineering Methods in Excel  
*A SunCam online continuing education course*

data is arranged in ascending order. It therefore follows that that the 2<sup>nd</sup> quartile is the median of the data. The **3<sup>rd</sup> quartile** (also called the **upper quartile**) ( $Q_3$ ) corresponds to 75% of the data below it and 25% of the data above it—when the data is arranged in ascending order.

As the name suggests, **percentiles** divide the data into hundredths. The 90<sup>th</sup> percentile ( $P_{90}$ ) has 90% of the distribution below it and 10% above it—when the data is arranged in increasing order, and so on and so forth. Therefore the 75<sup>th</sup> percentile is the same as the upper quartile, the 50<sup>th</sup> percentile is the same as the second quartile which is the same as the median, and so on.

The procedure for determining quartiles and percentiles is the same.

Step1: Arrange data from smallest to largest

Step 2: Determine the product  $np$  where  $n$  is sample size of the data and  $p$  the  $p^{\text{th}}$  percentile (or  $p^{\text{th}}$  quartile) expressed as a decimal.

Step 3: Round off the  $np$  value to the next integer if it is not an integer, and find the data value at that position.

Step 4: If  $np$  is an integer, say  $k$ , calculate the mean of the values at the  $k^{\text{th}}$  and the  $(k+1)^{\text{th}}$  positions.

Example: Let us use our speed data to illustrate the calculation of quartiles and percentiles.

Step1: We know that the stem-and-leaf diagram depicts the distribution arranged in increasing order. We shall use the stem-and-leaf diagram to assist us in the computations.

Step 2:

Lower quartile (which is also the 25<sup>th</sup> percentile):  $np = 100 * 0.25 = 25$ , which is an integer, so we go to Step 4.

Step 4: We need the average of the values at the 25<sup>th</sup> and 26<sup>th</sup> positions.



Engineering Methods in Excel  
A SunCam online continuing education course

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

$$Q_1 = \frac{45 + 45}{2} = 45 \text{ mph}$$

Upper quartile (which is also the 75<sup>th</sup> percentile):  $100 \times 0.75 = 75$ , so we average the values in the 75<sup>th</sup> and 76<sup>th</sup> positions.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

$$Q_3 = \frac{58 + 58}{2} = 58 \text{ mph}$$



Engineering Methods in Excel  
A SunCam online continuing education course

85<sup>th</sup> percentile:  $100 \times 0.85 = 85$ , so we average the values in the 85<sup>th</sup> and 86<sup>th</sup> positions.

Stem	Leaf
2	5 8
3	0 1 5 5 7 8 8 9
4	0 0 0 1 2 2 3 4 4 4 4 5 5 5 5 6 6 7 7 8 8 8 8 8 9 9 9 9 9
5	0 0 0 1 1 2 2 2 2 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 7 7 8 8 8 9 9 9
6	0 0 0 1 1 2 2 2 2 3 3 3 4 6 9
7	0 2 2 5 9

*Note: In the original image, red arrows point to the 85<sup>th</sup> and 86<sup>th</sup> positions in the 6th stem row. The 85<sup>th</sup> position is the first '1' and the 86<sup>th</sup> position is the second '2' in the sequence '0 0 0 1 1 2 2 2 2 3 3 3 4 6 9'.*

$$P_{85} = \frac{61 + 62}{2} = 61.5 \text{ mph}$$

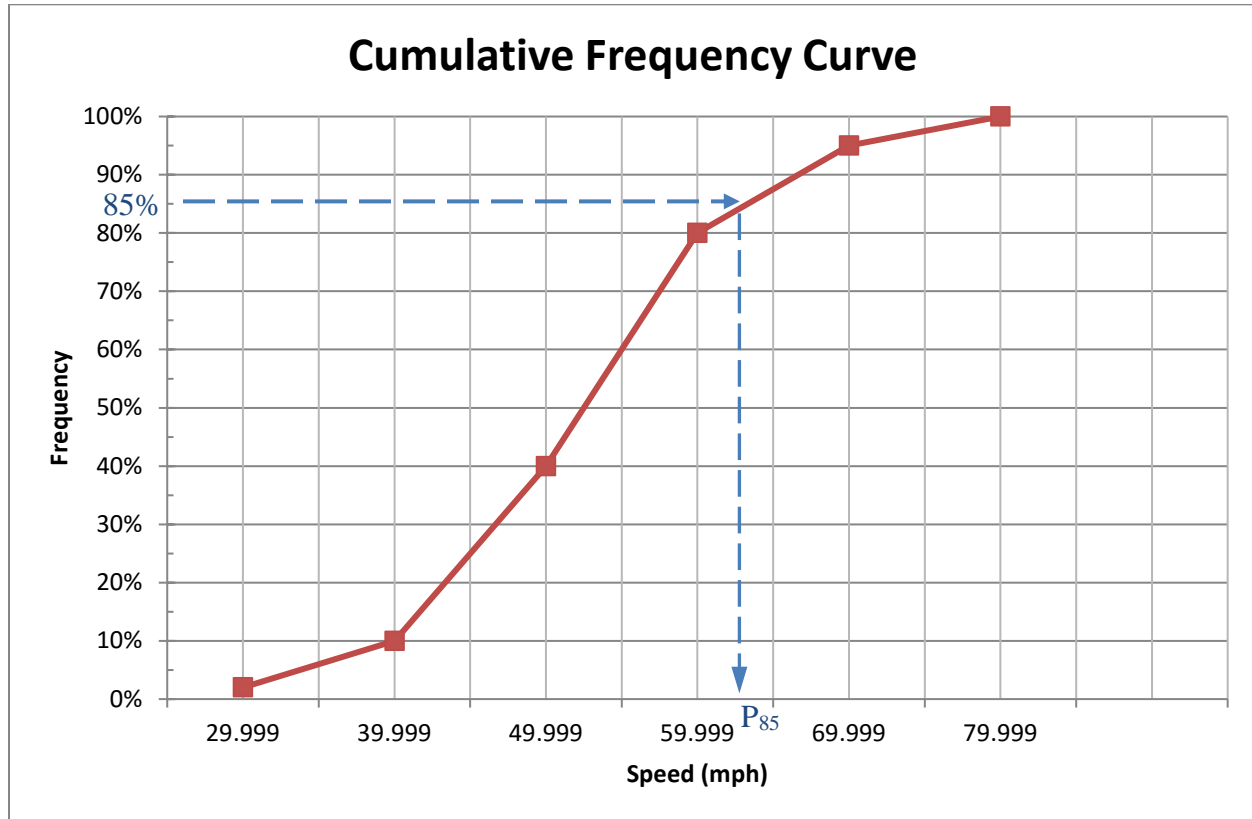
In the field of traffic engineering, the 85<sup>th</sup> percentile speed is of importance. The 85<sup>th</sup> percentile speed is the baseline used to set the posted speed limit on a roadway. For example, in Florida, the 85<sup>th</sup> percentile must be obtained for a sample of at least 100 vehicles on the roadway. Vehicle speeds recorded during traffic congestion, inclement weather, traffic incidents, and nighttime conditions must not be included in the data. The 85<sup>th</sup> percentile value obtained is rounded up or down to the nearest 5 mph based on engineering judgment and knowledge of local conditions. For example, the nature of the abutting land use – schools, hospitals etc., or for example, heavy pedestrian traffic, may warrant rounding down versus rounding up. The posted speed limit, however, cannot exceed the geometric design speed of the roadway.

The quartiles and percentiles may be estimated from the ogive, if the vertical axis (cumulative frequency) is expressed in percent. A horizontal line is drawn from the 85% point on the vertical axis to cut the ogive. The speed value at the point the horizontal line cuts the ogive is the 85<sup>th</sup> percentile speed.

For our speed data example:



Engineering Methods in Excel  
A SunCam online continuing education course



Comparing the estimation of the percentile from the ogive with the 4-step process for determining quartiles and percentiles, we notice that estimating the percentile from the ogive is equivalent to completing Step 1 and Step 2 of the 4-step procedure. In other words, it is the  $np$  position that is being used to determine the value of the quartile or percentile—without interpolating as required in Step 3 or Step 4. Therefore, using the ogive is an approximate method. However, for many engineering purposes, particularly with larger samples, any discrepancies introduced by this approximation are generally negligible for all intents and purposes.

#### Outliers

An **outlier** is a data value that is distant from other data points. An outlier may be due to variability in the measurement process or it may be due to some experimental error or anomaly. Outliers have the potential to bias the results, and should be identified and removed from the data before moving on to conduct any further statistical analysis with the data.



Engineering Methods in Excel  
*A SunCam online continuing education course*

By convention, an outlier is any data value  $x$  such that:

$$\begin{aligned}x &> Q_3 + 1.5*IR, \text{ or} \\x &< Q_1 - 1.5*IR\end{aligned}\quad (2.3)$$

where

$Q_1$  is the lower quartile

$Q_3$  is the upper quartile

$IR$  is the interquartile range, which is defined as

$$IR = Q_3 - Q_1 \quad (2.4)$$

Example: Let us conduct the outlier test on our data.

We previously determined  $Q_1 = 45$  mph, and  $Q_3 = 58$  mph.

From Equation (2.4),  $IR = 58 - 45 = 13$  mph

From Equation (2.3)

$$Q_3 + 1.5*IR = 58 + 1.5*13 = 77.5$$

And

$$Q_1 - 1.5*IR = 45 - 1.5*13 = 25.5$$

Thus, for any value in the data ( $x$ ), if  $x > 77.5$  mph, or if  $x < 25.5$  mph, then  $x$  is an outlier.

Therefore, the speed data contains two outliers, 25 mph, and 79 mph.

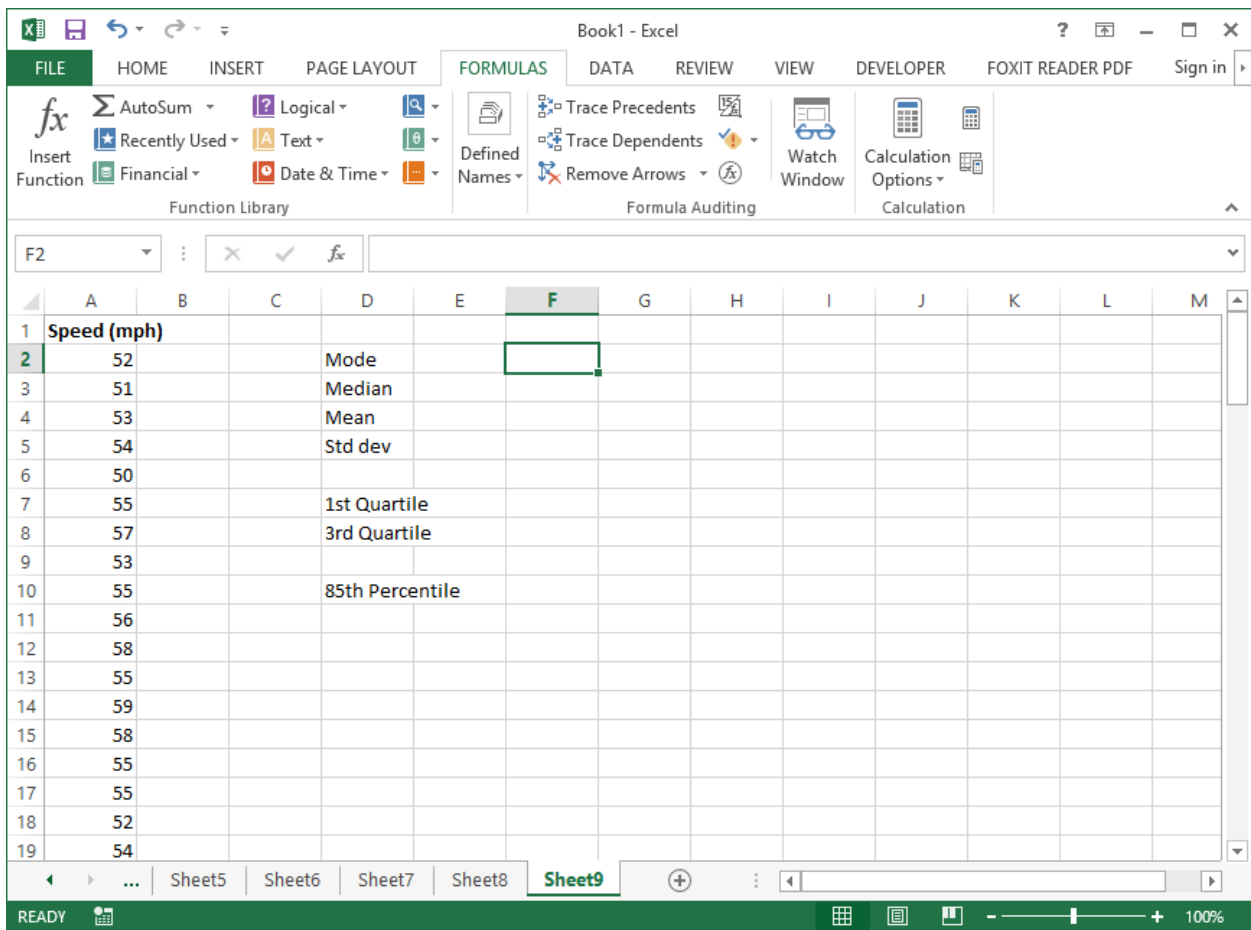
The outliers should be removed from the data. The outlier test is then conducted again on the new data set. If there are new outliers, they shall be removed, and the test repeated as many times as necessary until there are no outliers.



Engineering Methods in Excel  
A SunCam online continuing education course

### 2.9 Descriptive Statistics in Excel

Copy your speed data over to a new spreadsheet,



Let us calculate some of the descriptive statistics in Excel.

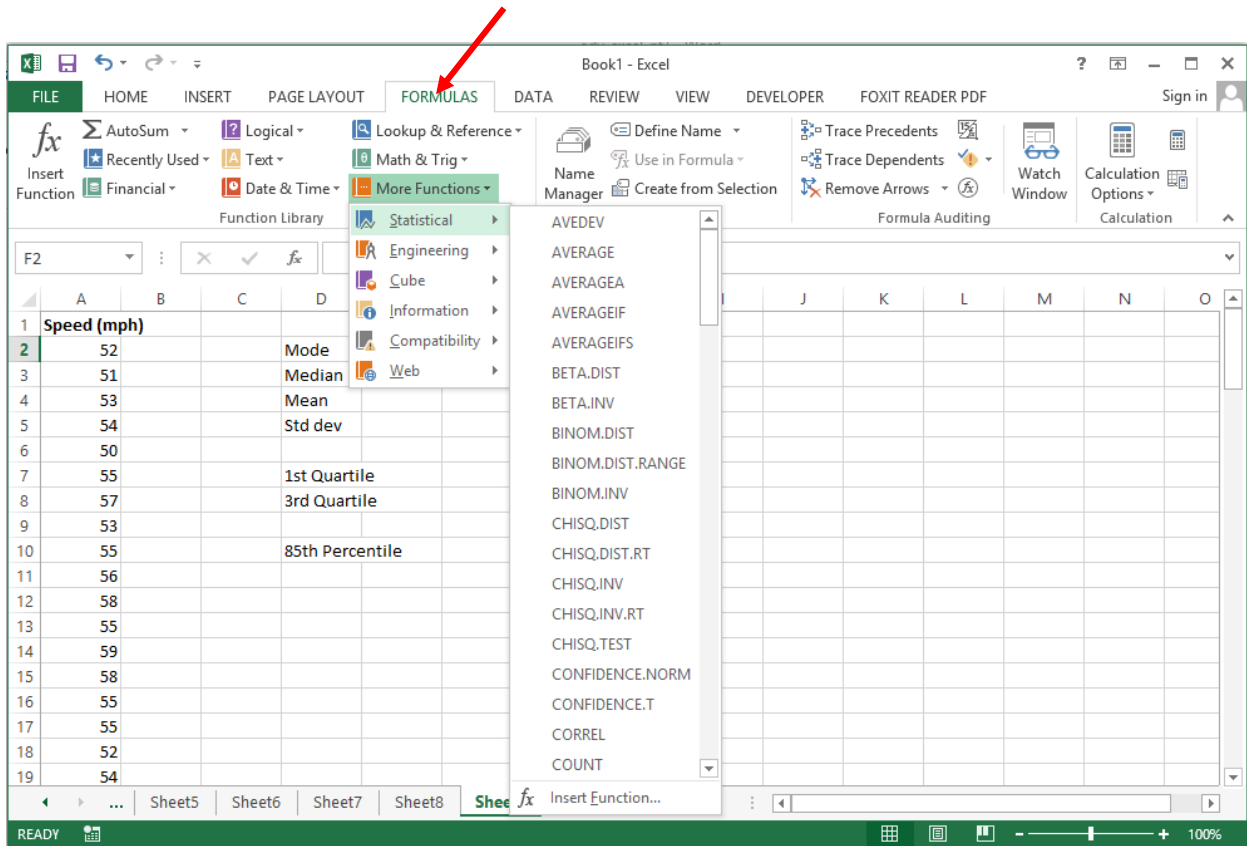


Engineering Methods in Excel  
A SunCam online continuing education course

Click on the **FORMULAS** tab.

Click on **More Functions**.

Click on **Statistical**.





Engineering Methods in Excel  
A SunCam online continuing education course

Scroll down the alphabetical list of statistical functions.

Click on **MODE.SINGL**

The screenshot shows the Microsoft Excel interface. The 'Formulas' ribbon is active, and the 'More Functions' dropdown menu is open, displaying a list of statistical functions. 'MODE.SINGL' is highlighted. A tooltip for 'MODE.SINGL(number1,number2,)' is displayed, explaining its function. The spreadsheet data is as follows:

	A	B	C	D
1	Speed (mph)			
2	52			Mode
3	51			Median
4	53			Mean
5	54			Std dev
6	50			
7	55			1st Quartile
8	57			3rd Quartile
9	53			
10	55			85th Percentile
11	56			
12	58			
13	55			
14	59			
15	58			
16	55			
17	55			
18	52			
19	54			





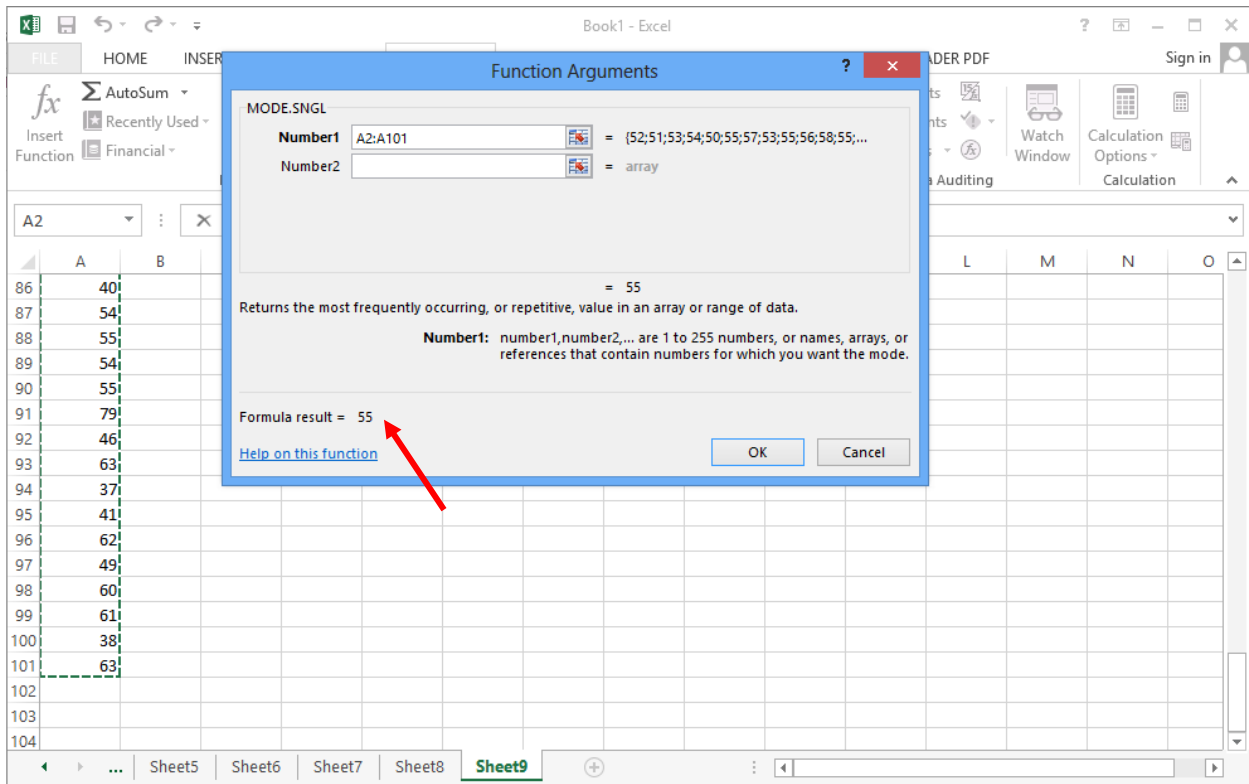
Engineering Methods in Excel  
A SunCam online continuing education course

The **Function Arguments** window opens.

Click in the **Number1** textbox.

Point-and-click to select the speed data on the spreadsheet.

Note that a preview of the formula result is displayed.



Click **OK** to dismiss the window.



## Engineering Methods in Excel A SunCam online continuing education course

The mode of the data is displayed on the spreadsheet at the selected cell location.

The screenshot shows the Microsoft Excel interface. The active cell is F2, containing the formula `=MODE.SNGL(A2:A101)`. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)														
2	52			Mode		55									
3	51			Median											
4	53			Mean											
5	54			Std dev											
6	50														
7	55			1st Quartile											
8	57			3rd Quartile											
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														

Incidentally, the function **MODE.MULT** function will produce an array of the multiple modes for multimodal data,



Engineering Methods in Excel  
A SunCam online continuing education course

Repeat the procedure to look up the function call for the Median.

Or, as with any other function begin typing “=” followed by the first few letters of the name of the function in the destination cell on the spreadsheet to pull up a shortlist of candidate functions.

The screenshot shows the Microsoft Excel interface with the FORMULAS ribbon selected. The spreadsheet contains a list of speeds in column A (rows 1-19) and corresponding statistical values in column D (rows 2-10). Cell F2 contains the value 55, which is the Mode. Cell F3 contains the formula =MEDIAN(A2:A19), which is being typed into the cell. A tooltip for the MEDIAN function is visible, stating: "Returns the median, or the number in the middle of the set of given numbers".

Row	Speed (mph)	Statistical Value
1		
2	52	Mode
3	51	Median
4	53	Mean
5	54	Std dev
6	50	
7	55	1st Quartile
8	57	3rd Quartile
9	53	
10	55	85th Percentile
11	56	
12	58	
13	55	
14	59	
15	58	
16	55	
17	55	
18	52	
19	54	



Engineering Methods in Excel  
A SunCam online continuing education course

Double click on **MEDIAN**.  
Point-and-click to select the data.

The screenshot shows the Microsoft Excel interface with the 'FORMULAS' ribbon selected. The formula bar displays the formula `=MEDIAN(A2:A101)`. The worksheet shows a column of data in column A, labeled 'Speed (mph)', with values ranging from 52 to 54. The 'Median' row in column D shows the result of the formula as 55. A tooltip for the MEDIAN function is visible, showing the syntax: `MEDIAN(number1, [number2], ...)`.

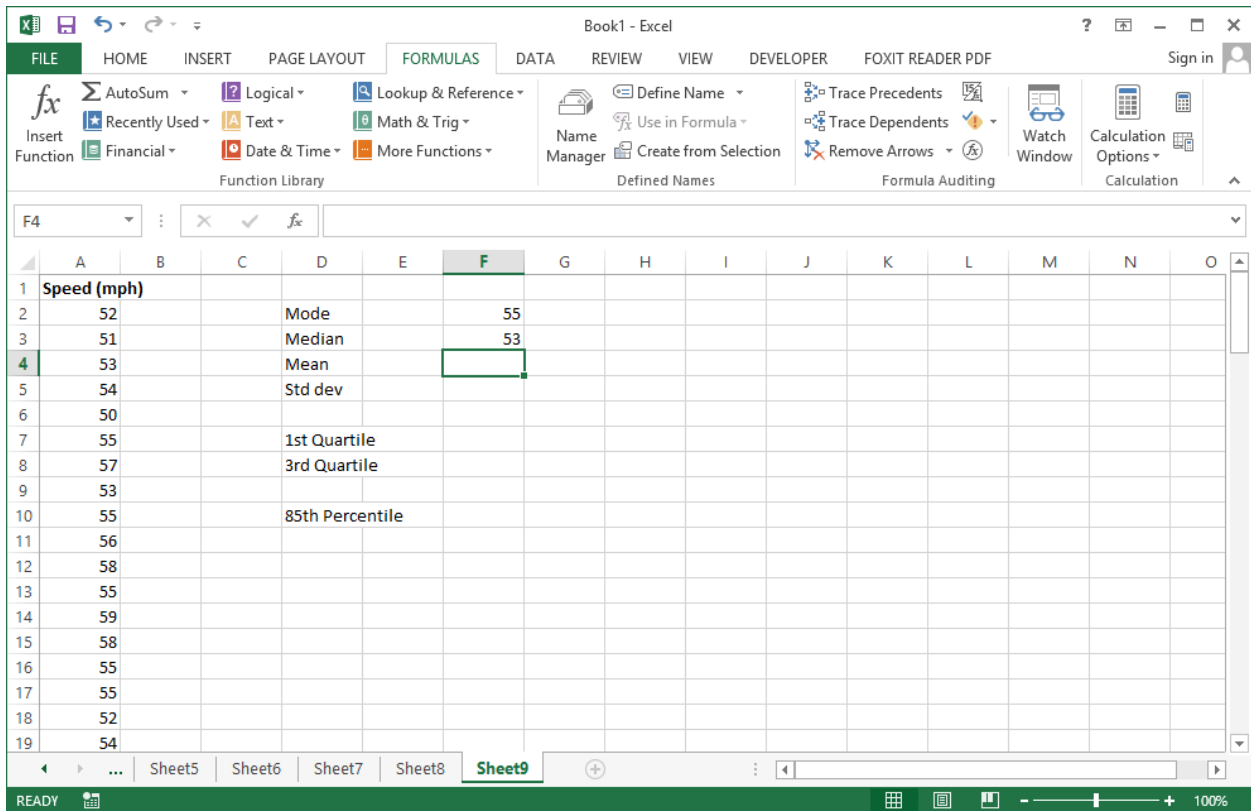
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)														
2	52			Mode		55									
3	51			Median		=MEDIAN(A2:A101)									
4	53			Mean											
5	54			Std dev											
6	50														
7	55			1st Quartile											
8	57			3rd Quartile											
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														



Engineering Methods in Excel  
A SunCam online continuing education course

Hit **Enter** on the keyboard.

The result is displayed.





Engineering Methods in Excel  
A SunCam online continuing education course

The *Excel* function for the mean is **AVERAGE**.  
Apply the **AVERAGE** function to the data.

The screenshot shows the Microsoft Excel interface with the 'FORMULAS' tab selected. The formula bar displays the formula `=AVERAGE(A2:A101)`. The worksheet contains a column of speed data in cells A2 through A19. The following table represents the data shown in the spreadsheet:

Row	Speed (mph)	Statistical Measure	Value
2	52	Mode	55
3	51	Median	53
4	53	Mean	=AVERAGE(A2:A101)
5	54	Std dev	AVERAGE(number1, [number2], ...)
6	50		
7	55	1st Quartile	
8	57	3rd Quartile	
9	53		
10	55	85th Percentile	
11	56		
12	58		
13	55		
14	59		
15	58		
16	55		
17	55		
18	52		
19	54		



Engineering Methods in Excel  
A SunCam online continuing education course

Hit **OK**.

The average is computed and displayed.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Speed (mph)</b>														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev											
6	50														
7	55			1st Quartile											
8	57			3rd Quartile											
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														



Engineering Methods in Excel  
A SunCam online continuing education course

The *Excel* function for the sample standard deviation is **STDEV.S**.  
Apply the **STDEV.S** function to the data.

The screenshot shows the Microsoft Excel interface with the 'FORMULAS' ribbon selected. The formula bar displays the formula `=STDEV.S(A2:A10)` for cell A2. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev		<code>=STDEV.S(A2:A10)</code>									
6	50														
7	55			1st Quartile											
8	57			3rd Quartile											
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														





Engineering Methods in Excel  
A SunCam online continuing education course

Hit OK.

The sample standard deviation is computed, which yields:

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev		9.959164									
6	50														
7	55			1st Quartile											
8	57			3rd Quartile											
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														



## Engineering Methods in Excel A SunCam online continuing education course

We shall now work on quartiles and percentiles.

To find the quartiles, look up “Quartiles” in the statistical functions list, or type “=quartile”.

The screenshot shows an Excel spreadsheet with the following data:

Speed (mph)					
52			Mode		55
51			Median		53
53			Mean		51.87
54			Std dev		9.959164
50					
55			1st Quartile		=quartile
57			3rd Quartile		
53					
55			85th Percentile		
56					
58					
55					
59					
58					
55					
55					
55					
52					
54					

The formula bar shows the entry `=quartile`. A tooltip is visible for the `QUARTILE` function, which states: "Returns the quartile of a data set, based on percentile values from 0.1, exclusive".



Engineering Methods in Excel  
A SunCam online continuing education course

Double click on the option **QUARTILE.EXC**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Speed (mph)</b>														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev		9.959164									
6	50														
7	55			1st Quartile		=QUARTILE.EXC(									
8	57			3rd Quartile		QUARTILE.EXC(array, quart)									
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														



Engineering Methods in Excel  
A SunCam online continuing education course

Click on the function (fx) button to review further details of this function.

The screenshot shows the Microsoft Excel interface with the FORMULAS ribbon selected. The function library is open, and the QUARTILE.EXC function is highlighted. A red arrow points to the fx button in the ribbon. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	<b>Speed (mph)</b>														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev		9.959164									
6	50														
7	55			1st Quartile		=QUARTILE.EXC(									
8	57			3rd Quartile		QUARTILE.EXC(array, quart)									
9	53														
10	55			85th Percentile											
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														



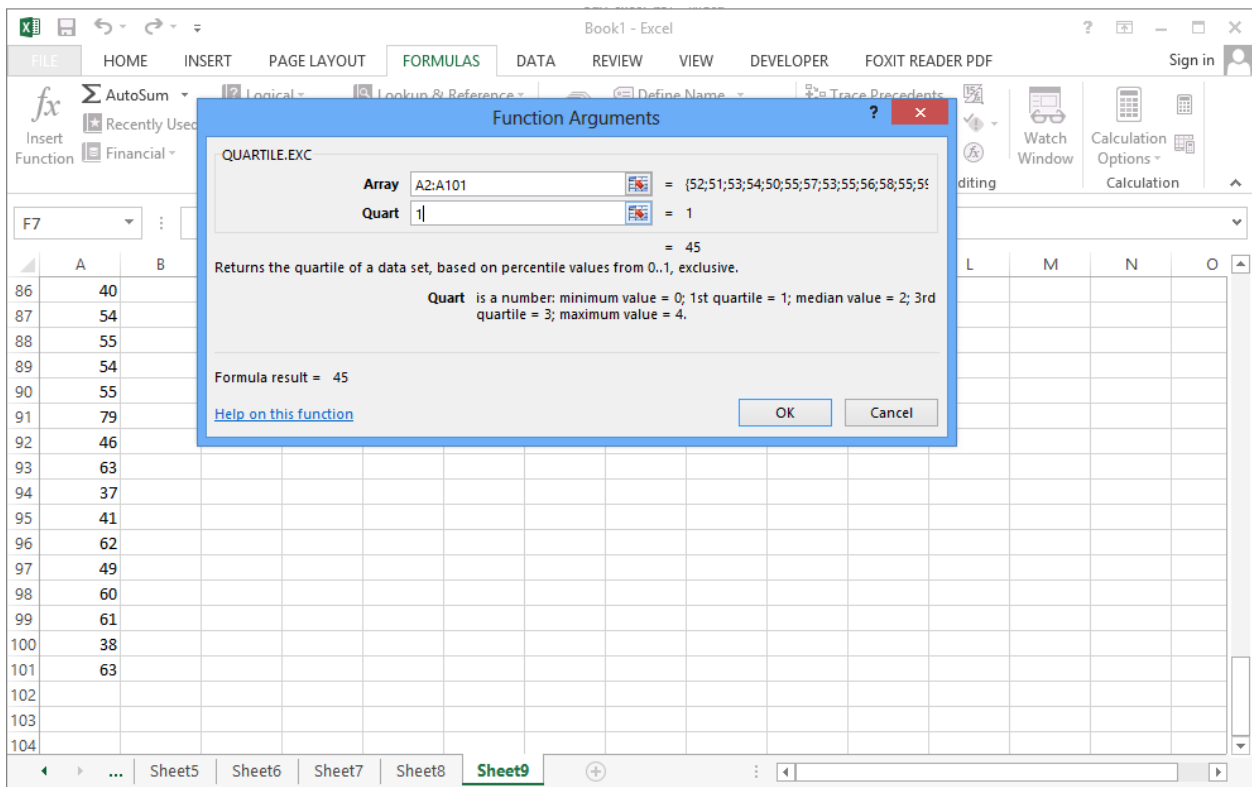
Engineering Methods in Excel  
A SunCam online continuing education course

The **Function Arguments** window opens.

Click in the **Array** textbox

Point-and-click to select your data on the spreadsheet.

For **Quart**, enter “1” to compute the 1<sup>st</sup> quartile (“3” to compute the 3<sup>rd</sup> quartile, and so on and so forth).



Review the result preview.



Engineering Methods in Excel  
A SunCam online continuing education course

Hit **OK** to complete the computation.

The screenshot shows an Excel spreadsheet with the following data:

Row	Speed (mph)	Statistical Measure	Value
2	52	Mode	55
3	51	Median	53
4	53	Mean	51.87
5	54	Std dev	9.959164
6	50		
7	55	1st Quartile	45
8	57	3rd Quartile	
9	53		
10	55	85th Percentile	
11	56		
12	58		
13	55		
14	59		
15	58		
16	55		
17	55		
18	52		
19	54		

The formula bar shows the formula: `=QUARTILE.EXC(A2:A101,1)`



Engineering Methods in Excel  
A SunCam online continuing education course

Repeat the process to compute the 3<sup>rd</sup> quartile, or simply fill handle the 1<sup>st</sup> quartile formula and then update it to ensure the correct **Array** is selected (use of absolute reference for the 1<sup>st</sup> quartile computation would eliminate the need for this update step), and update the **Quart** argument to “3”.

The screenshot shows an Excel spreadsheet with the following data in column A:

Speed (mph)
52
51
53
54
50
55
57
53
55
56
58
55
59
58
55
55
52
54

The formula bar shows the formula: `=QUARTILE.EXC(A2:A101,3)`. The spreadsheet shows the following results:

Statistic	Value
Mode	55
Median	53
Mean	51.87
Std dev	9.959164
1st Quartile	45
3rd Quartile	:A101,3
85th Percentile	

A tooltip for the QUARTILE.EXC function is visible, showing the following options:

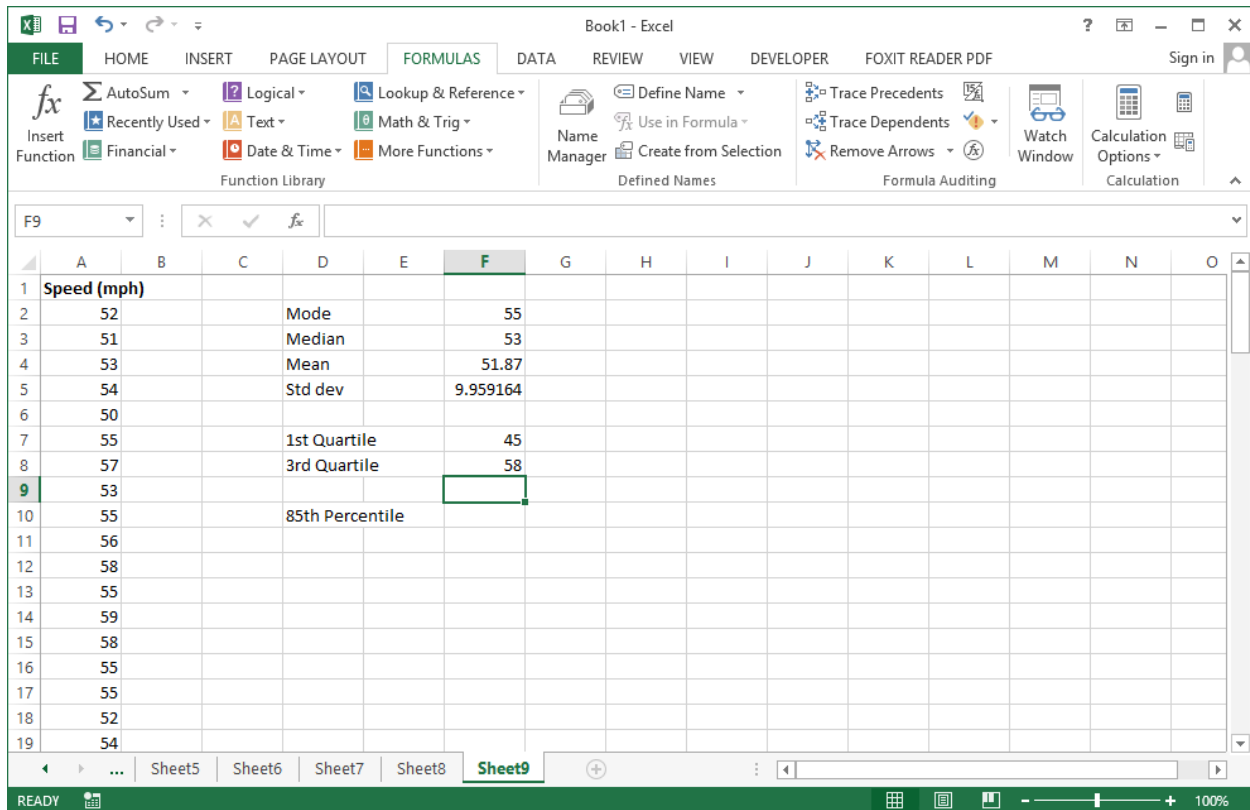
- 1 - First quartile (25th percentile)
- 2 - Median value (50th percentile)
- 3 - Third quartile (75th percentile)

The tooltip also states: "QUARTILE.EXC returns the first quartile (25th percentile)".



Engineering Methods in Excel  
A SunCam online continuing education course

Hit the **Enter** key to complete the calculation.



Note that in the quartile computations we saw three available functions in the *Excel* statistical functions library, namely: QUARTILE.EXC, QUARTILE.INC, AND QUARTILE.

QUARTILE is the function call for pre-2010 versions of the *Excel* program and has been maintained to enable backwards compatibility. QUARTILE.INC uses a zero-based position numbering system to perform the calculations, e.g. the lowest value in the data is position zero, the next highest value is position 1 and so on and so forth. QUARTILE.EXC uses a one-based position numbering system to perform the calculations, e.g. the lowest value in the data is position 1, the next highest value is position 2 and so on and so forth.





Engineering Methods in Excel  
A SunCam online continuing education course

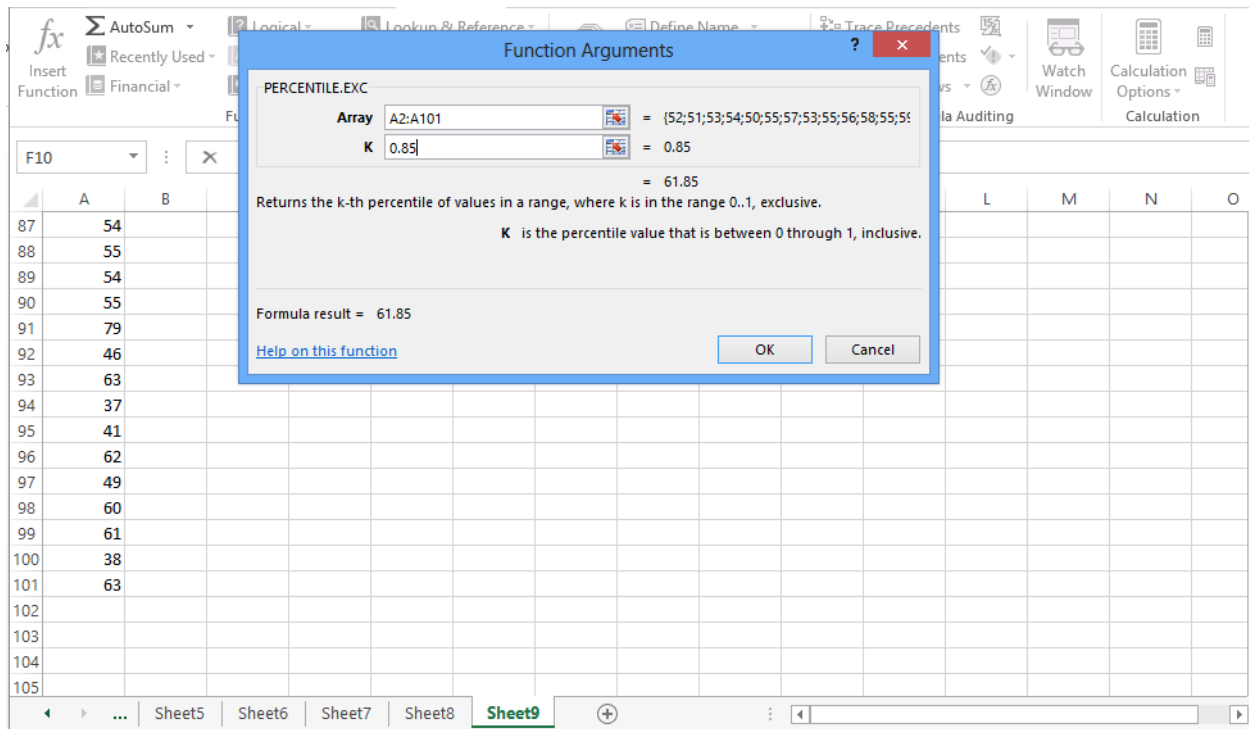
QUARTILE.INC and QUARTILE.EXC also differ in the way interpolations are conducted. Participants in this course who are interested in these computational nuances are referred to the *Microsoft* website for further information on this discussion.

The computation of percentiles follows that of quartiles.

Select the **PERCENTILE.EXC** function from the statistical functions library, or simply type it in the destination cell (with the “=” prefix).

Point-and-click to select the data.

For the **K** value, enter the percentile you are seeking, as a decimal.





## Engineering Methods in Excel A SunCam online continuing education course

Hit **OK** to complete the computations.

The screenshot shows an Excel spreadsheet with the following data and calculations:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Speed (mph)														
2	52			Mode		55									
3	51			Median		53									
4	53			Mean		51.87									
5	54			Std dev		9.959164									
6	50														
7	55			1st Quartile		45									
8	57			3rd Quartile		58									
9	53														
10	55			85th Percentile		61.85									
11	56														
12	58														
13	55														
14	59														
15	58														
16	55														
17	55														
18	52														
19	54														

We notice that we get a slightly different value from what we calculated manually earlier in this course. Also (and you may test this on your own), the PERCENTILE.INC function yields yet another slightly different result. As we discussed under quartiles, these discrepancies are due to the position numbering systems and/ or the different interpolation algorithms applied for each function in *Excel*. However, from fundamental principles, recall that any percentile (or quartile) must have a certain number of the data values below it and above it. So long as this is satisfied, the numerical discrepancies in the percentile (or quartile) value—due to these different interpolation strategies built into these *Excel* functions—may be considered negligible.



Engineering Methods in Excel  
A SunCam online continuing education course

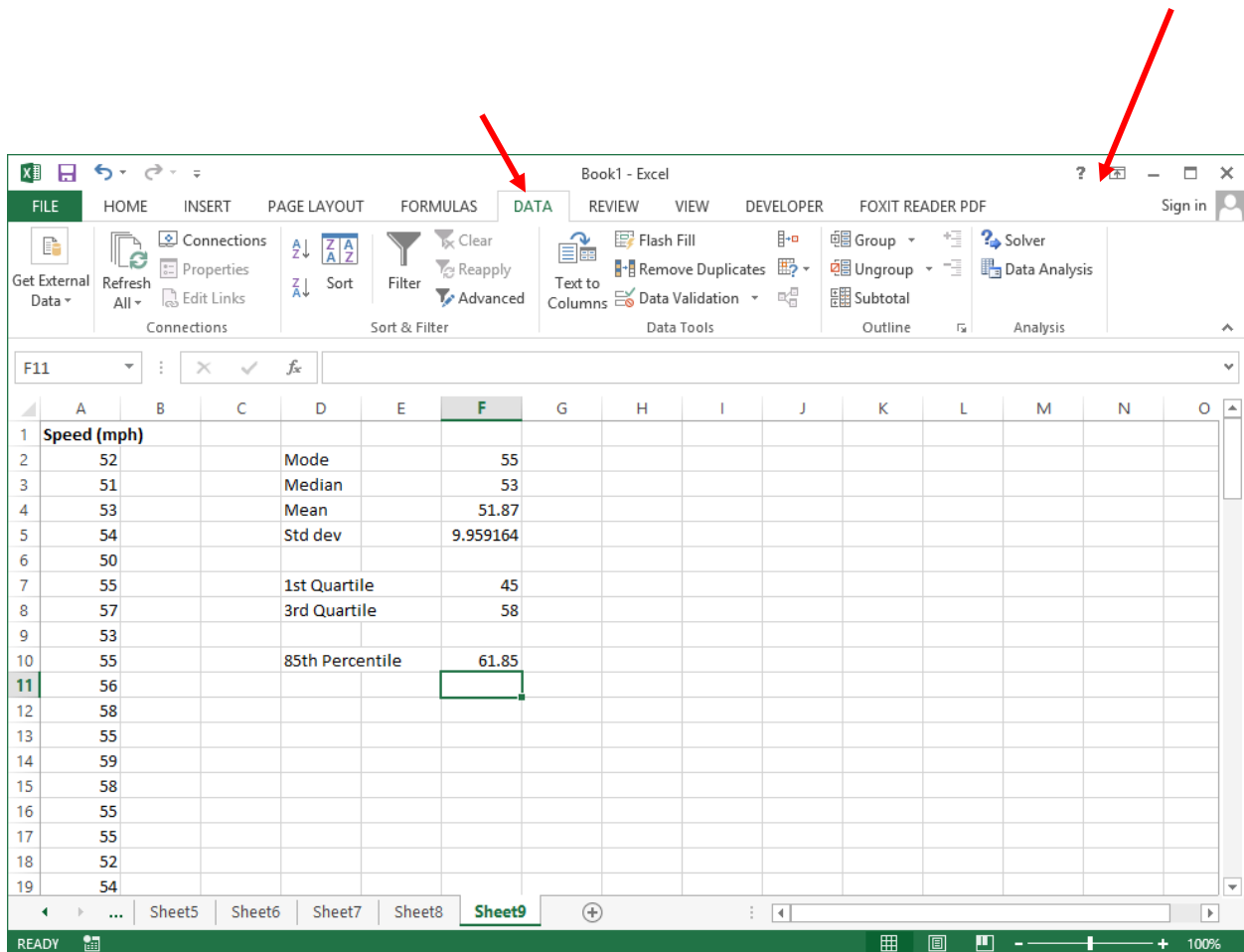
Descriptive Statistics from the Data Analysis Tools

The *Excel Data Analysis Tools* Add-in presents a one-button-click function that produces a summary of some descriptive statistics in a tabular output.

Click on the **DATA** tab.

Go to the **Analysis** palette.

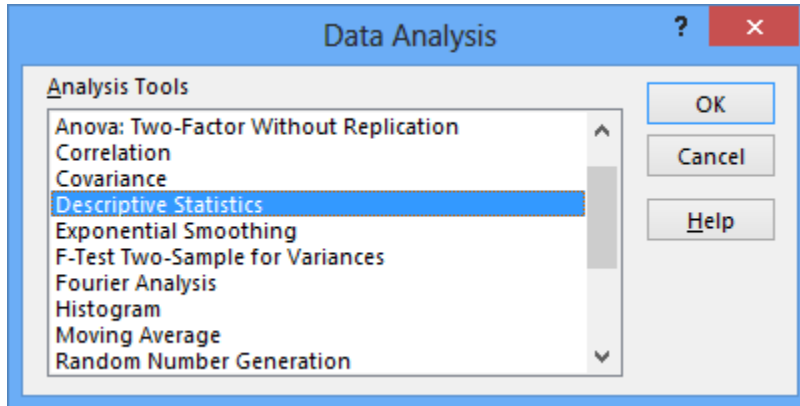
Click on **Data Analysis**.





Engineering Methods in Excel  
*A SunCam online continuing education course*

In the **Data Analysis** window, select **Descriptive Statistics**.  
Hit **OK**.



The **Descriptive Statistics** window opens.

Under **Input**, select the data,

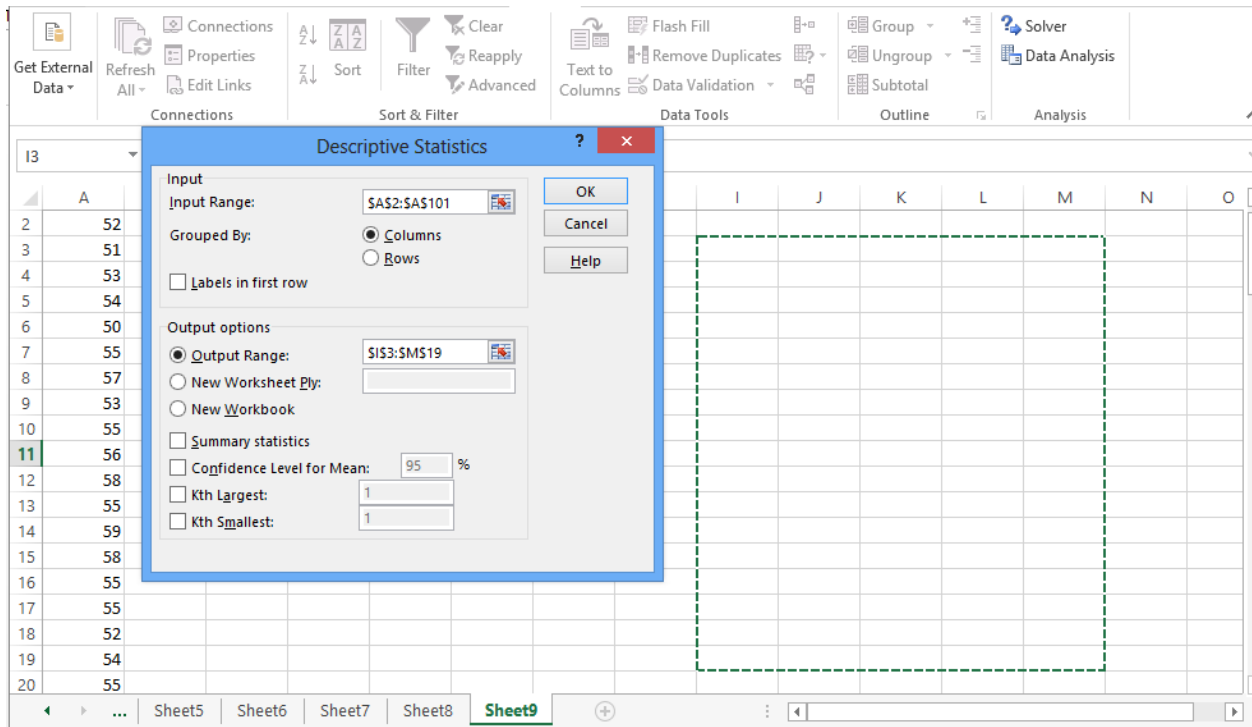
This data is grouped Columns.

The selection of the data that we made does not include any labels.



Engineering Methods in Excel  
A SunCam online continuing education course

For the **Output options**: select the **Output Range** option.  
Click in the **Output Range** textbox.  
Select a range of cells on the current spreadsheet.



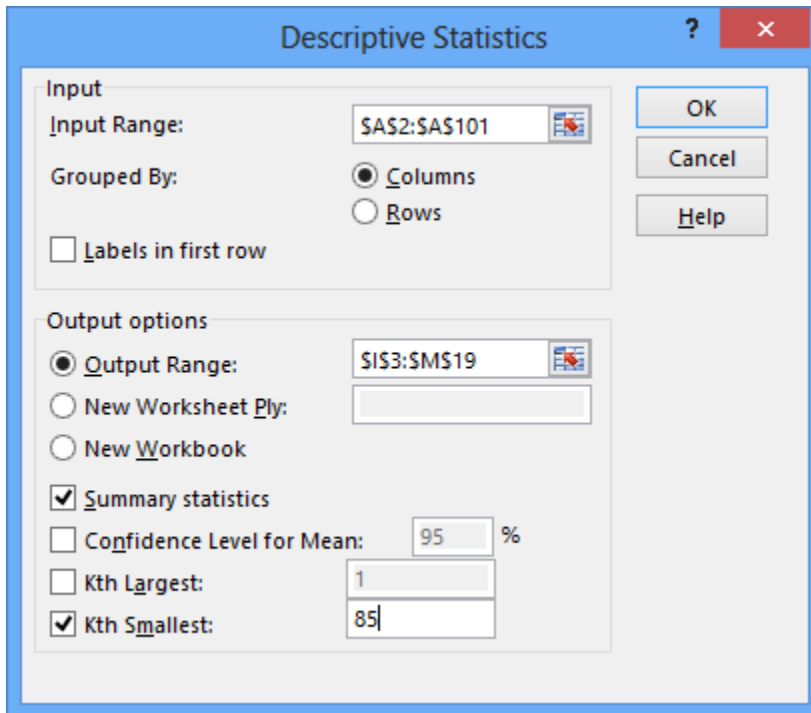


Engineering Methods in Excel  
A SunCam online continuing education course

Check **Summary statistics**.

Check **kth Smallest**, and enter 85, to display the value at the 85<sup>th</sup> position when the data is arranged in increasing order.

In other words, a rounded estimate of the 85<sup>th</sup> percentile value. (Note that 85 was obtained from  $n*p = 100*0.85$ , for other sample sizes, the  $np$  value shall be computed accordingly).



Hit **OK**.



Engineering Methods in Excel  
A SunCam online continuing education course

The result is displayed on the spreadsheet at the designated range of cells location(s).

The screenshot shows the Microsoft Excel interface with the Data Analysis Toolpak summary for a column of data. The data is located in column I, rows 2 through 20. The summary statistics are displayed in column J, rows 3 through 18. The summary includes the following values:

Statistic	Value
Mean	51.87
Standard Error	0.995916
Median	53
Mode	55
Standard Deviation	9.959164
Sample Variance	99.18495
Kurtosis	0.452502
Skewness	-0.08357
Range	54
Minimum	25
Maximum	79
Sum	5187
Count	100
Smallest (n)	61



Engineering Methods in Excel  
*A SunCam online continuing education course*

Review the results and compare to those obtained by previous methods.

<i>Column1</i>	
Mean	51.87
Standard Error	0.995916
Median	53
Mode	55
Standard Deviation	9.959164
Sample Variance	99.18495
Kurtosis	0.452502
Skewness	-0.08357
Range	54
Minimum	25
Maximum	79
Sum	5187
Count	100
Smallest(85)	61





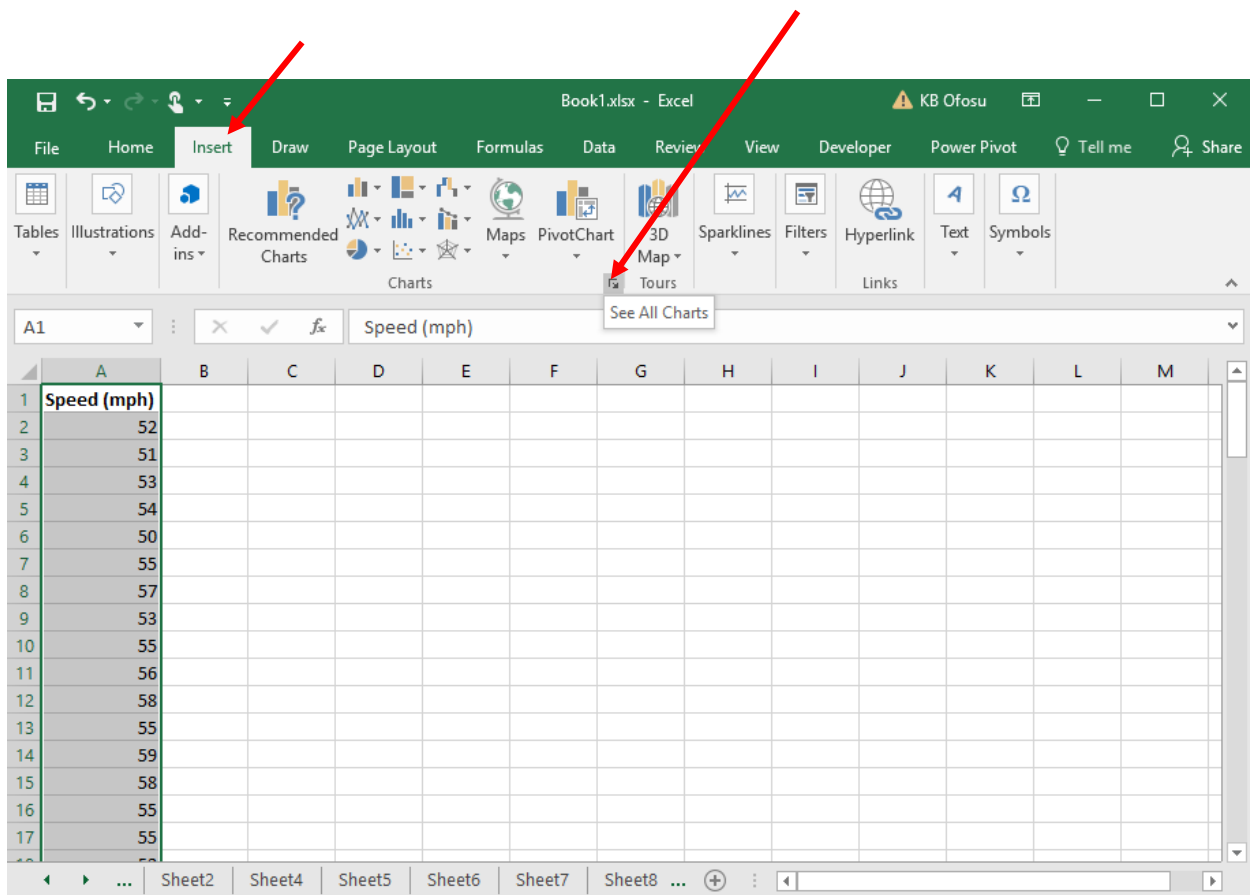
Engineering Methods in Excel  
A SunCam online continuing education course

Box and Whisker Plot (Excel 2016 and beyond users only)

A Box and Whisker Plot also called simply a **box plot** provides a graphical summary of a distribution in terms of its quartiles.

In *Excel*, a box plot can be created by the innovative use of spreadsheet functions. However, starting with the 2016 release, *Excel* provides a built-in functionality for plotting box plots directly from the general tools for creating charts.

- Select your speed data.
- Click on the **Insert** tab.
- Go the **Charts** palette.
- Click on **See All Charts**.



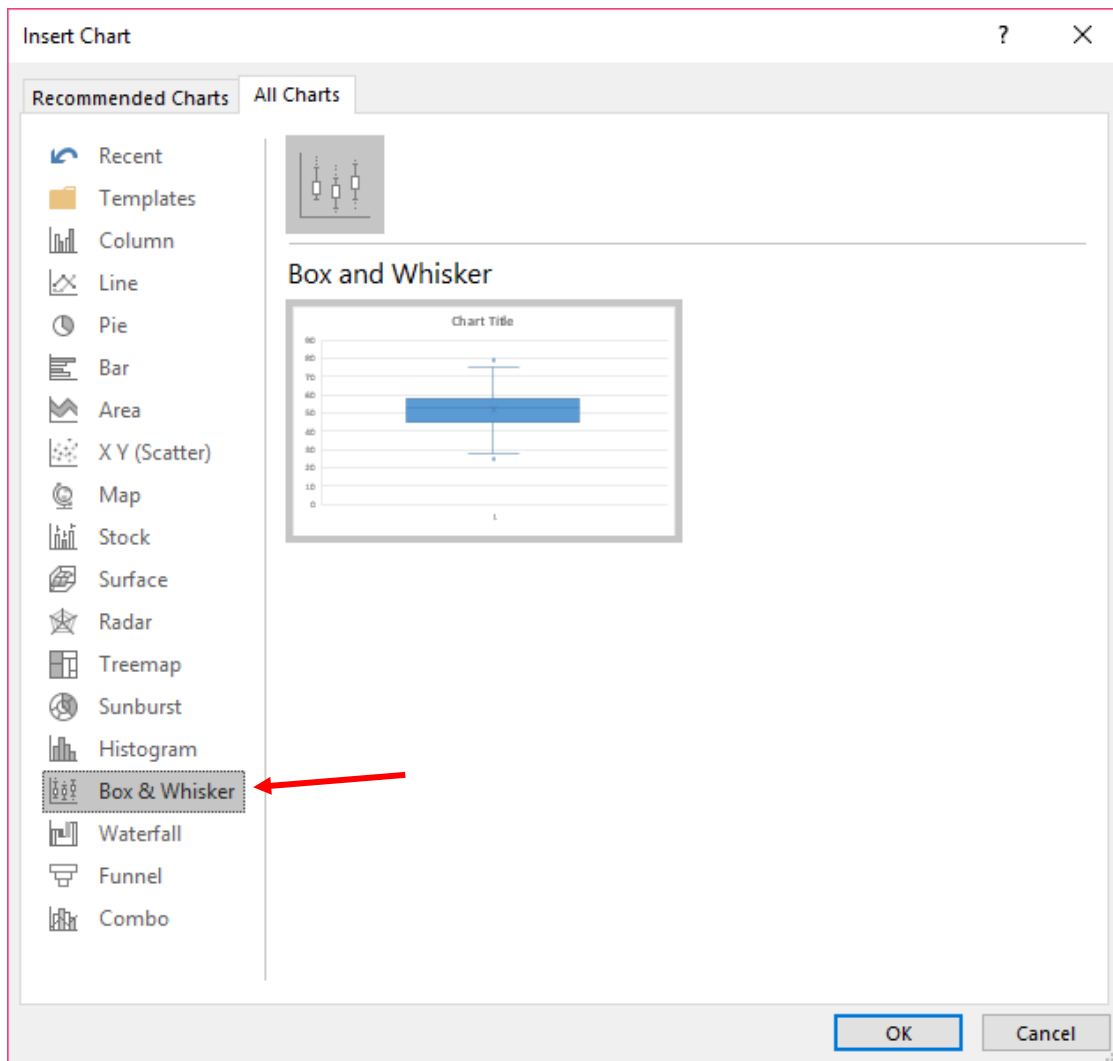


Engineering Methods in Excel  
*A SunCam online continuing education course*

The **Insert Chart** window opens.

Click on the tab **All Charts**.

Click on **Box and Whisker**.

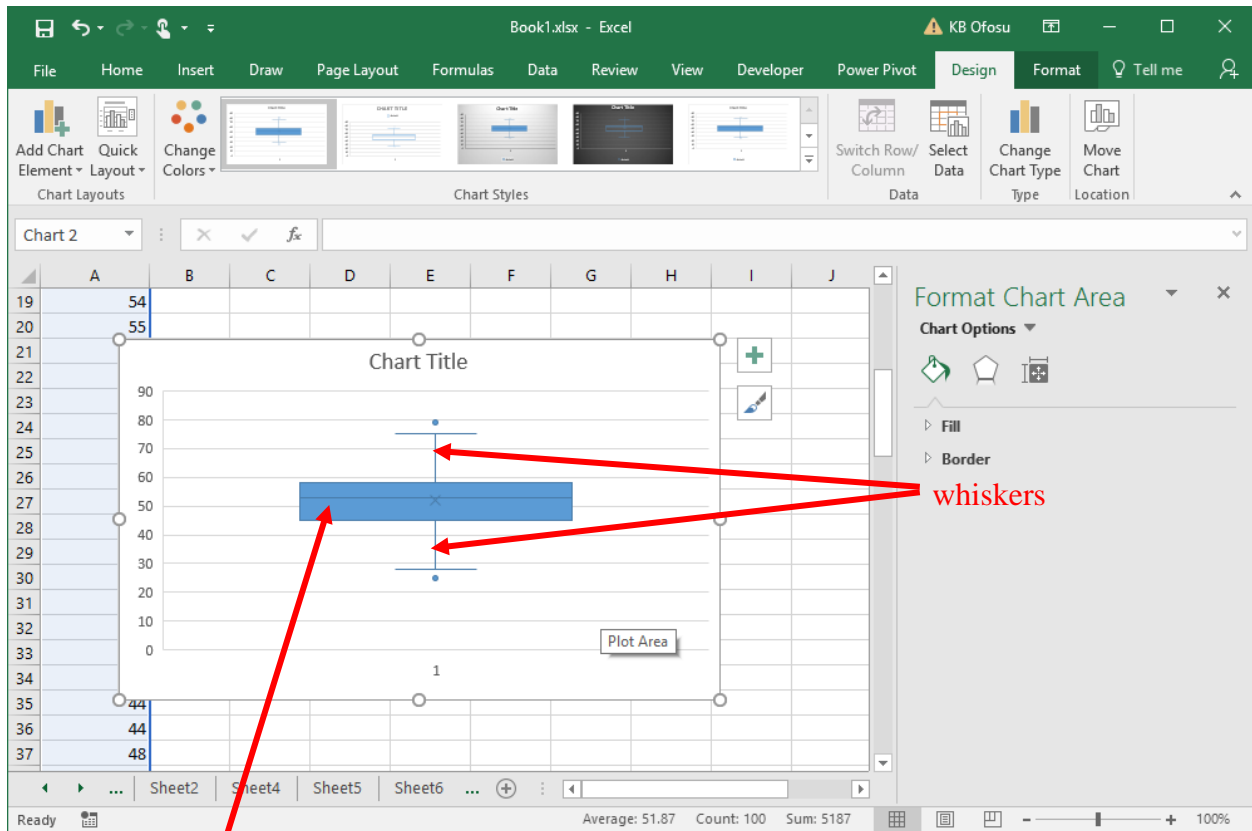


Click **OK**.



Engineering Methods in Excel  
A SunCam online continuing education course

The box plot is inserted on the spreadsheet.



box

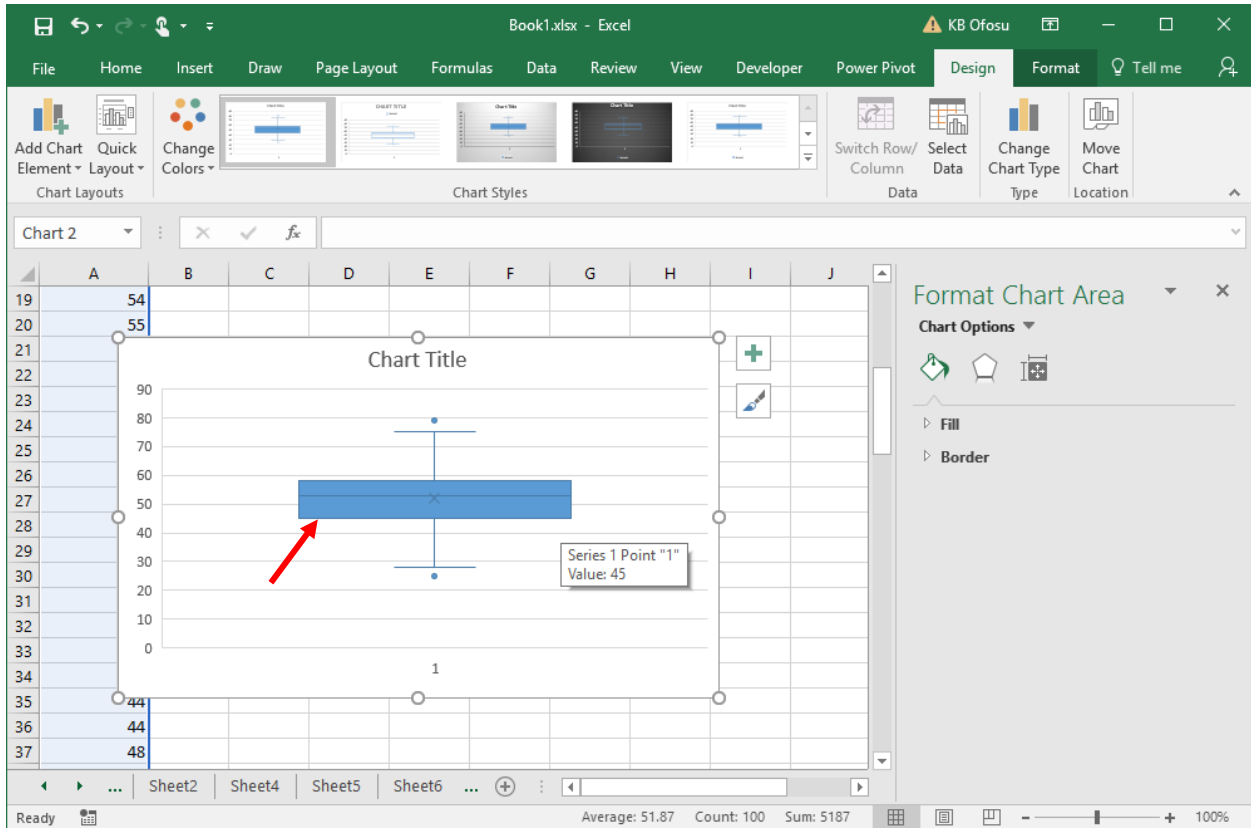
whiskers



Engineering Methods in Excel  
A SunCam online continuing education course

Hover over the line features and the point features of the chart to identify,

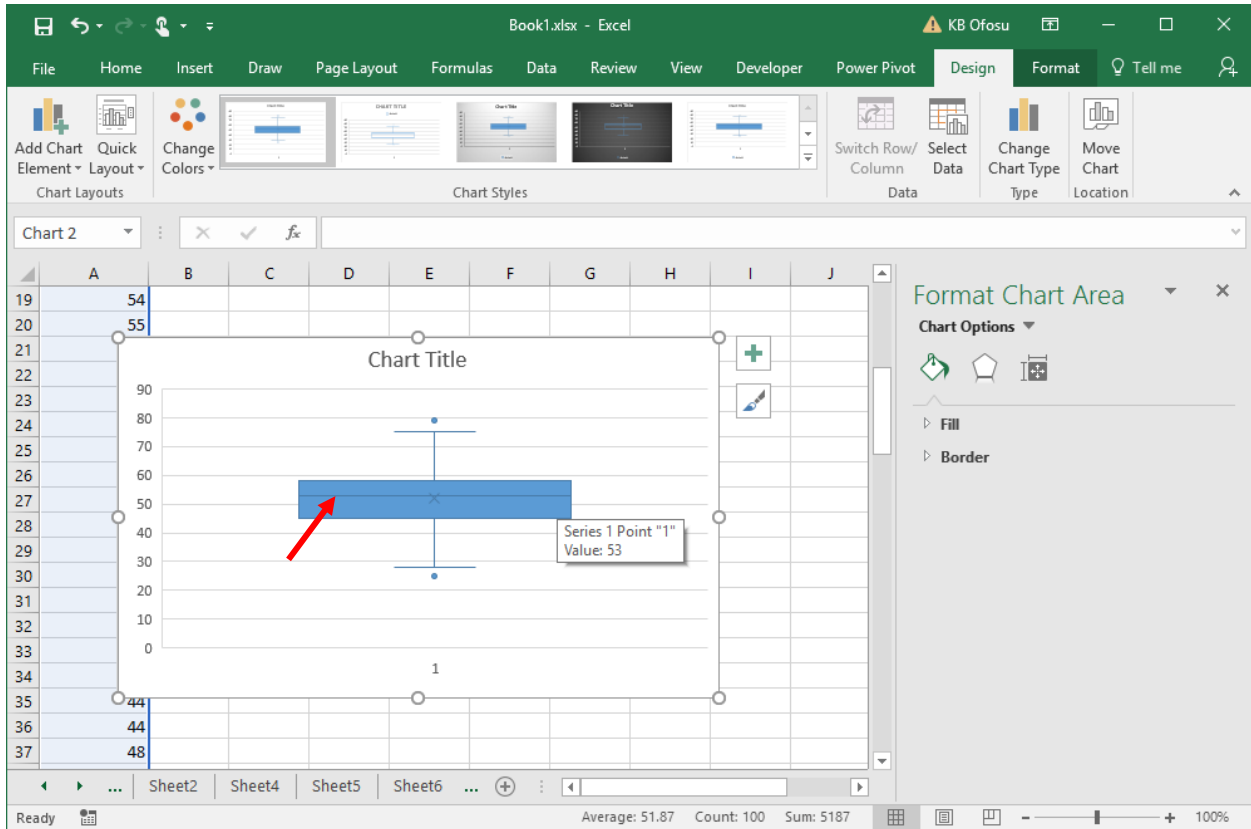
1. the lower quartile





Engineering Methods in Excel  
A SunCam online continuing education course

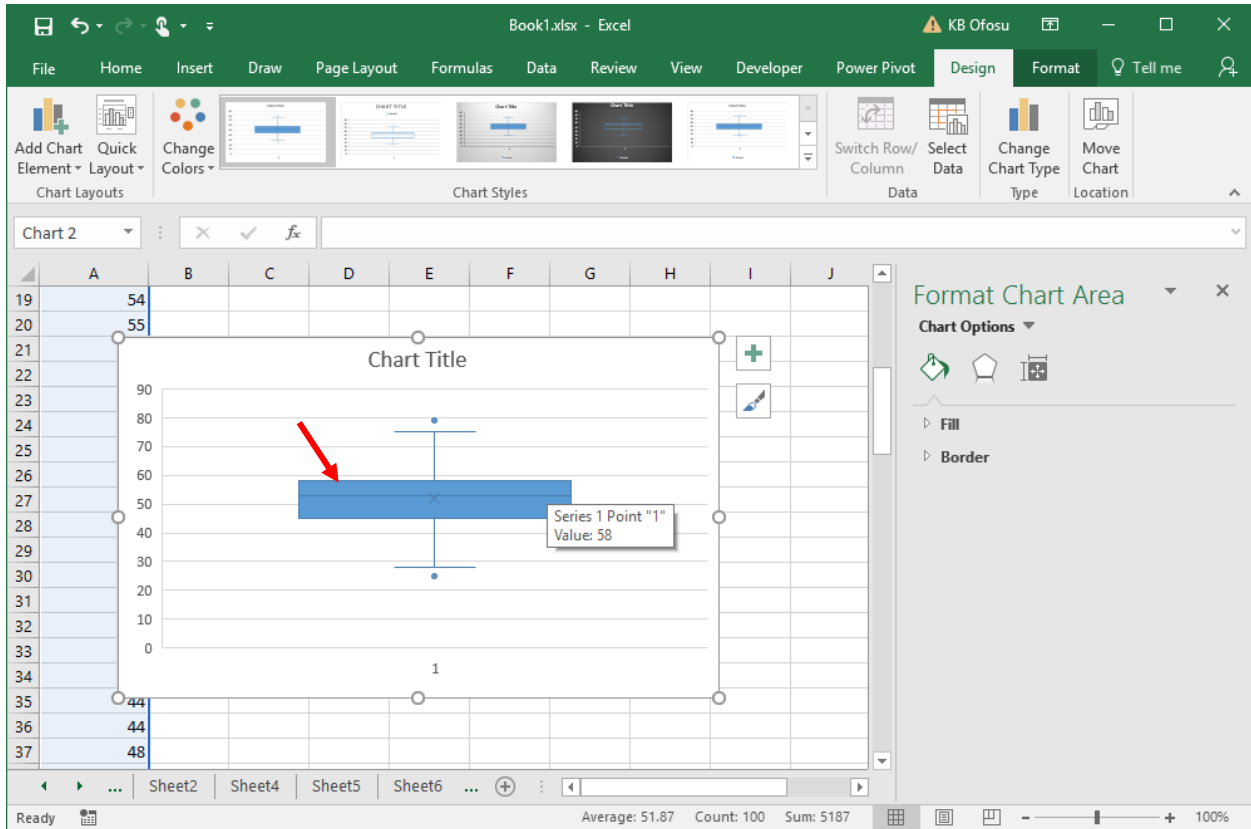
2. the median





Engineering Methods in Excel  
A SunCam online continuing education course

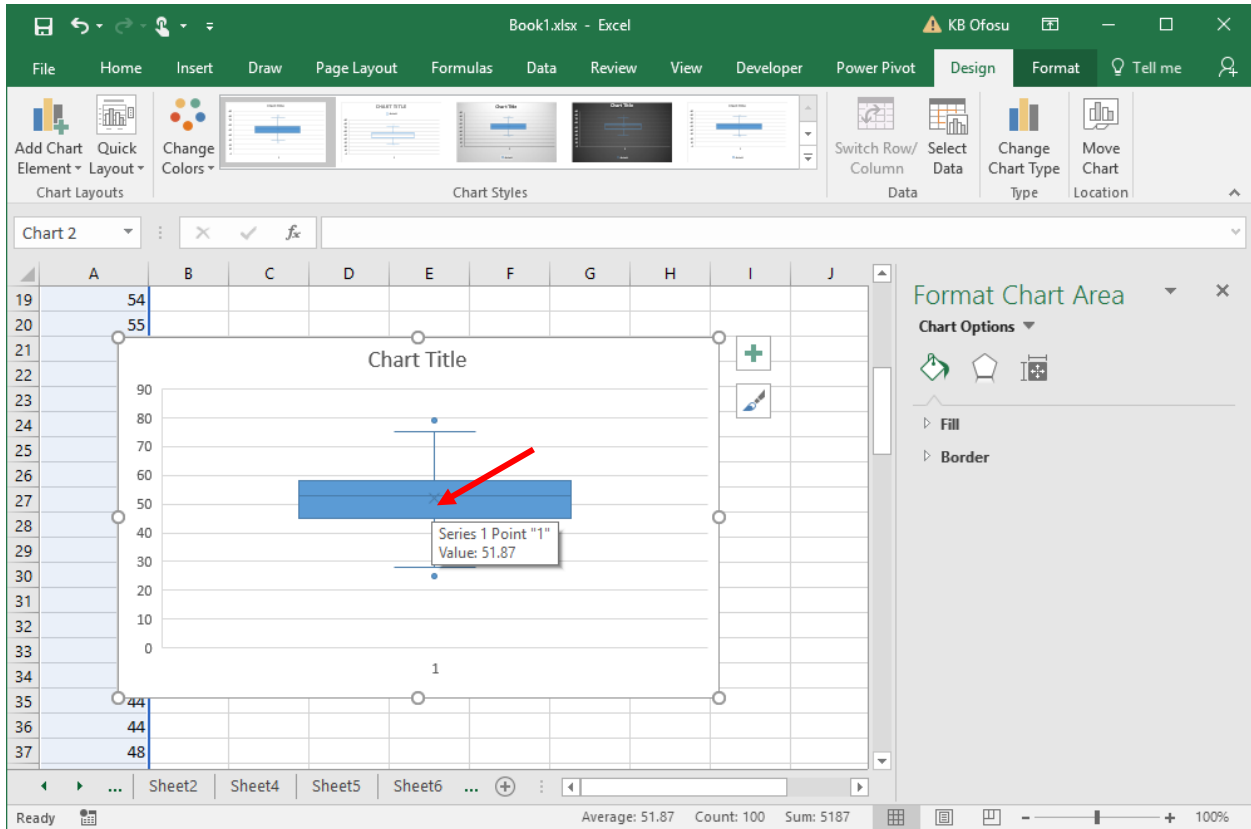
3. the upper quartile





Engineering Methods in Excel  
A SunCam online continuing education course

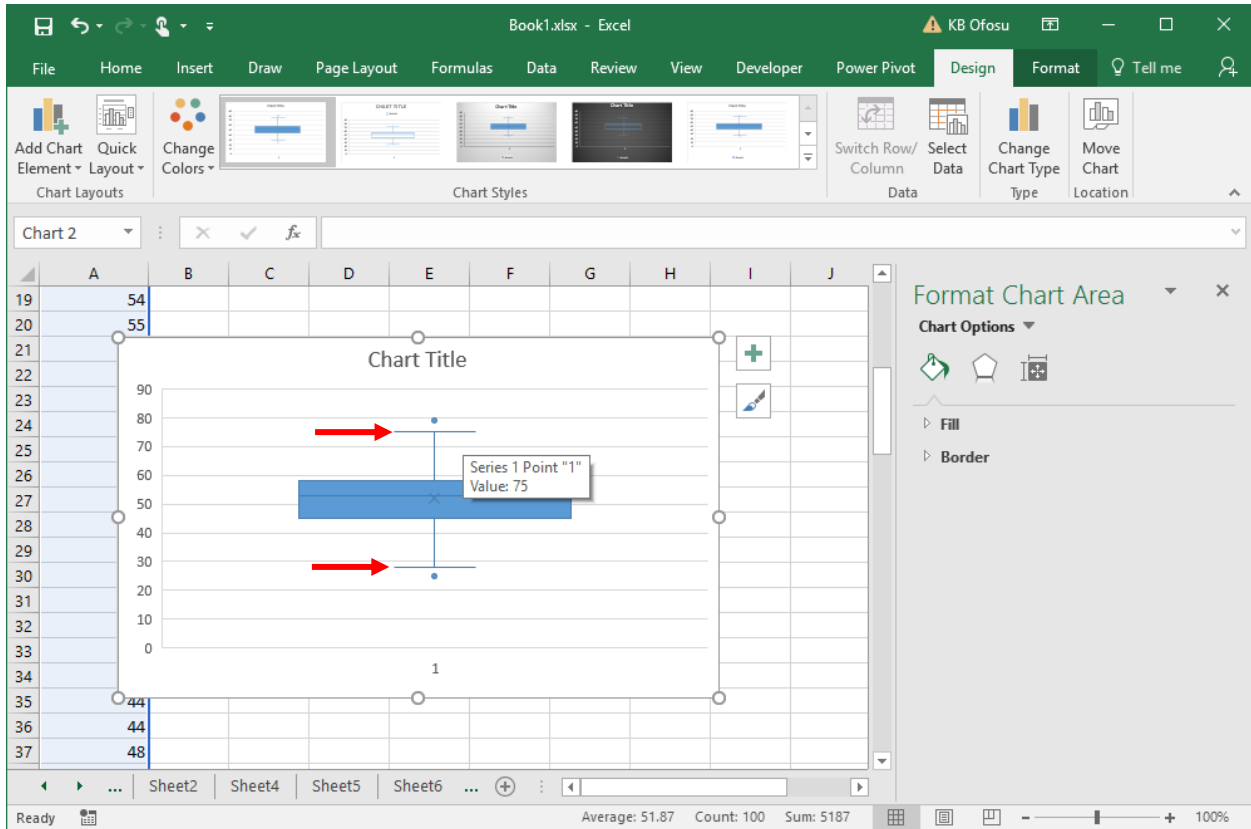
4. the mean (an “X” marker)





Engineering Methods in Excel  
A SunCam online continuing education course

5. the threshold values for outliers

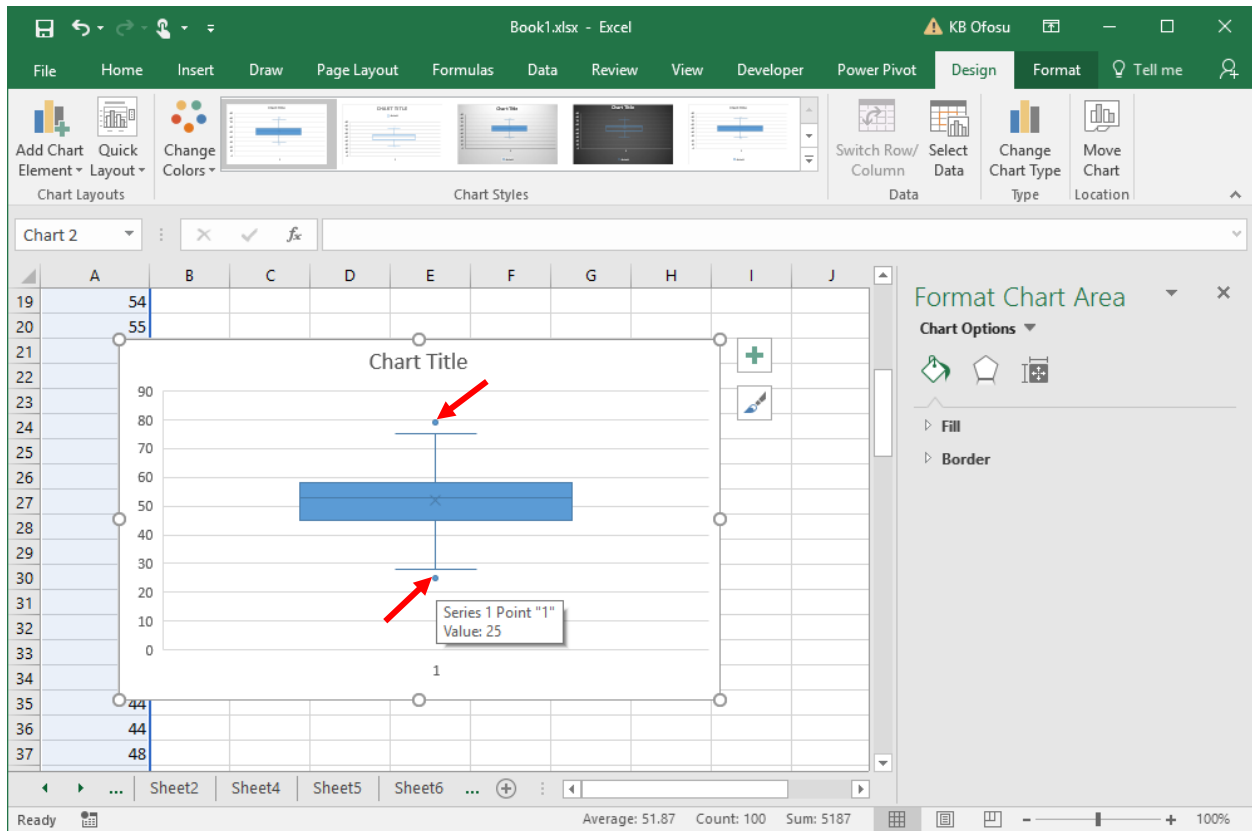






Engineering Methods in Excel  
A SunCam online continuing education course

6. The outliers in the data (if any) (circular point markers).



The box plot enables us to see the spread of the data in terms of the quartiles. It enables us to visually see the difference between the mean and median, hence the skew of the distribution. We are also able to immediately see any outliers.

Descriptive Statistics complete! Congratulations.



Engineering Methods in Excel  
*A SunCam online continuing education course*

### 3. CONCLUSION

This course is part of a series on engineering methods in *Excel*. This course presented a broad overview of fundamental principles and engineering applications of engineering data analysis, and how to use *Excel* tools to analyze and solve these types of problems.

In this course, we looked at statistical methods used to organize and summarize data sets—such as frequency distributions, frequency diagrams and descriptive statistics. We learned how to use the built-in functions of *Excel* to implement these methods, strategies and computations.

This course has enabled participants to identify professional situations where the innovative application of techniques learned in this course are relevant and of benefit to the productivity, efficiency, and quality of the work output of a practitioner.

Practitioners are strongly encouraged to look out for situations in their domains of expertise where engineering methods in *Excel* are applicable and will be of benefit to their work and their organization.

A successful application of engineering methods in *Excel* requires a careful and meticulous approach; and can only be mastered and retained by practice and repetition.

It has been my utmost pleasure presenting this topic to you. Thank you.



Engineering Methods in Excel  
*A SunCam online continuing education course*

**REFERENCES**

- Institute of Transportation Engineers. (2010). *Traffic Engineering Handbook* (6th ed.). Washington, DC: Institute of Transportation Engineers.
- Johnson, R. A. (2011). *Miller & Freund's Probability and Statistics for Engineers* (Eighth ed.). Prentice Hall.
- Microsoft. (2017). *Microsoft Office help and training - Office Support*. Retrieved March 13, 2017, from Microsoft Office Home: <https://support.office.com/en-us/article/Define-and-solve-a-problem-by-using-Solver-9ed03c9f-7caf-4d99-bb6d-078f96d1652c>

Images were all drawn/ prepared by Kwabena. Ofosu