

# Introduction to MS Excel 2016 for Epidemiologists

Self-Study Guide

CEPR Training and Development Unit  
Last reviewed: June 2022

## Learning objectives

After participating in this case study, learners will be able to:

1. Perform basic spreadsheet functions
  - a. Launch and exit Excel
  - b. Use the toolbars and move through cells
  - c. Freeze rows and columns for easy viewing
  - d. Modify a worksheet
  - e. Create new fields (i.e., variables) and format cells
  - f. Sort and filter data
  - g. Use auto fill and fill series functions
  - h. Change page setup (e.g., orientation, headers/footers)
  - i. Print directly from a worksheet
2. Perform basic data manipulation functions
  - a. Create data /line lists
  - b. Create a chart (i.e., graph)
  - c. Create a PivotTable
  - d. Create a Pivot chart and epidemic curve
  - e. Use basic formulas and functions

## Format and duration

This case study is intended to be used as self-study tool. Completion of this self-study will take 45 min to 1.5 hours depending on the user's familiarity with Excel.

## Materials required

Electronic files

- Word document: *Excel\_101\_Epis.doc*
- Excel file: *Excel\_101\_Epis.xls*

Computer with MS Office 2016

## Acknowledgements

This self-study guide was adapted by Rosalind Bihun, François William Tremblay and Phillippe Bélanger from the coursework originally designed in MS Excel 2007 for the Canadian Field Epidemiology Program's 'Epidemiology in Action 2011'. Additional resources used to develop this guide include:

- Liz Procter-Gray, Carol Bigelow and Yiwei Jiang. (2010). MS Excel version 2007 for epidemiologists. University of Massachusetts Amherst. Accessed July 29, 2011 from [http://courses.umass.edu/biep691f/pdf/3.%20%20MS%20Excel%202007%20for%20Epidemiology\\_2010.pdf](http://courses.umass.edu/biep691f/pdf/3.%20%20MS%20Excel%202007%20for%20Epidemiology_2010.pdf).
- Florida Gulf Cost University. (2007). Excel 2007 Tutorial. Accessed July 27, 2011 from <http://www.fgcu.edu/support/office2007/Excel/index.asp>.
- Training modules developed by Lisa Jensen for Epidemiology in Action 2010.
- Training modules developed by Joanne Tataryn, Louise Gagne, Jennifer Cutler, and Robyn Mitchell for the Advanced Outbreak Investigations Course from the Field Surveillance Training Institute 2011.
- Training modules developed by François-William Tremblay and Gillian Frost for Epidemiology in Action 2011.

## Additional Information

Microsoft Excel (MS Excel) is a standard program that epidemiologists use for creating spreadsheets, maintaining them, and producing tables and graphs. There are many different MS Excel resources and tools available for free online. If you are working in the field and have a question, remember that Google and YouTube are your friends! Type your question into the search bar and you will likely find an answer or helpful tutorial.

## Table of Contents

Learning objectives.....	2
Format and duration.....	2
Materials required.....	2
Acknowledgements .....	2
Additional Information .....	2
Introduction to MS Excel .....	5
What is MS Excel?.....	5
Advantages and disadvantages of MS Excel .....	6
Spreadsheet Basics .....	7
Getting Started with Excel .....	7
The Toolbars and Moving Through Cells.....	7
Modifying a Worksheet .....	9
Formatting Cells.....	11
Formulas and Functions.....	11
Tables.....	13
Creating a table.....	13
Sorting.....	15
Filtering.....	17
Auto-Filling and Fill Series.....	17
Page Setup and Printing.....	18
Creating a Data Set .....	19
Designing Your Database.....	20
Data Entry .....	21
Step 1: Create your data set .....	21
Step 2: Deal with missing data.....	22
Step 3: Format fields.....	22
Step 4: Create a Data Dictionary .....	23
Creating Charts and Epidemic Curves.....	24
Basics of creating a chart.....	24
Step 1: Creating your chart.....	24
Step 2: Basic formatting.....	25

Step 3: Aesthetics .....	26
Charting Do's and Don'ts .....	28
Creating PivotTables .....	29
Basics of creating a PivotTable .....	29
Step 1: Format the data to be used in the PivotTable.....	29
Step 2: Create a PivotTable.....	29
Some other tips and tricks for working with PivotTables .....	35
Creating multiple PivotTables.....	35
Refreshing your data.....	35
Adding rows or columns of data to your linelist.....	35
Generate a linelist of specific cases .....	35
Creating Charts with PivotTables.....	36
Creating Epidemic Curves with Pivot Tables .....	36
Step 1: Create a Pivot Table containing the data you want in your epi-curve .....	36
Step 2: Regroup the date.....	36
Step 3: Create the chart.....	37
Formatting Guidelines .....	38

# Introduction to MS Excel

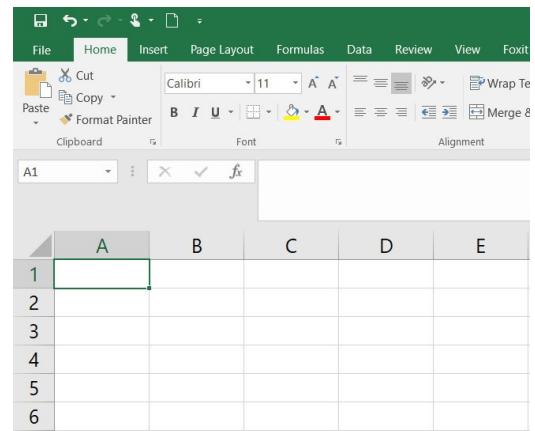
## What is MS Excel?

MS Excel is a widely-used software package used for creating spreadsheets. In Excel, spreadsheets are called "worksheets".

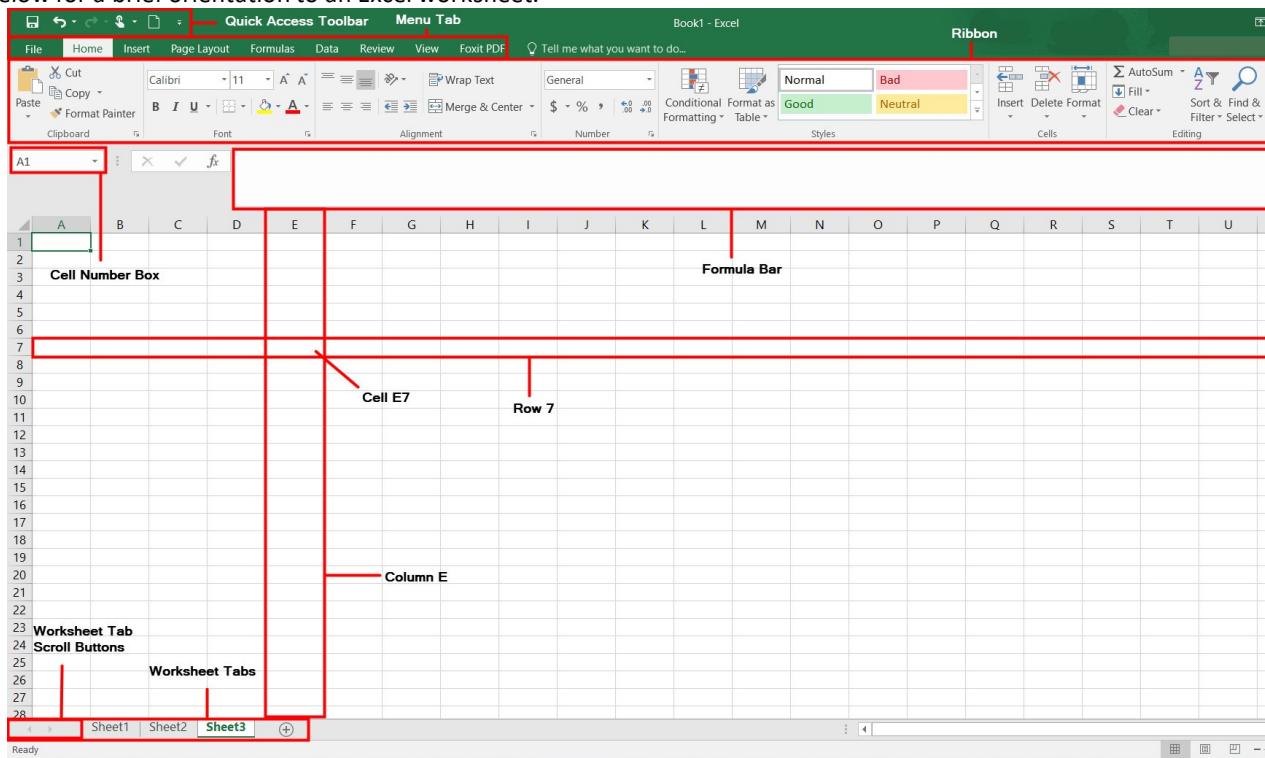
A spreadsheet is simply a grid of information that might include numbers or words or a mix of the two. Storage in a grid is a handy way of organizing all of the information. You can do all sorts of things with a spreadsheet, including making sorted lists, tables, and graphs (called "chart" in Excel).

Each worksheet is a grid of columns (indicated by letters across the top) and rows (indicated by numbers along the left-hand side). Thus, each space (or cell) on the worksheet is identified by a row-column designation, such as A1 (see right). Each cell can hold a number, text or the result or calculation of a mathematical formula.

In Excel, records are normally organized in rows with each cell in the row containing one item of information or value. Each individual item of information is referred to as a field.



See below for a brief orientation to an Excel worksheet.



## **Advantages and disadvantages of MS Excel**

### Advantages

- It is very commonly used and very easily shared.
- Many statistical analysis packages permit the direct import of Excel data or export into Excel format.
- Data can be sorted by any column while still retaining the integrity of each record.
- It is easy to create new variables that are mathematical functions of variables. For example, you can tell Excel to calculate the mean of the fields in columns A, B, and C and store the result as a new field, such as column D.
- Blocks of data can be copied and moved from one part of the worksheet to another or from worksheet to worksheet.
- Excel offers lots of formats for displaying data (e.g., number of significant digits, display of dates as month-day-year) with no loss of information.

### Disadvantages

- Excel can make mistakes in mathematical formulae if you inadvertently mix character (i.e., string or text) and numeric fields. For example, if column A is character and column B is numeric, the addition of entries in column A and column B may be incorrect.
- Only one user at a time can enter data / work on the spreadsheet
- Excel is prone to human error when entering and manipulating data
- Excel is unable to query data
- Excel is not relational
- Excel is not a substitute for database software

# Spreadsheet Basics

## Getting Started with Excel



To launch Excel, navigate to the application on your computer and double-click on this icon.

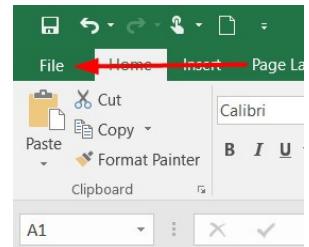
If you do not have a shortcut on your desktop, you can find the application by clicking on the **Start** button (bottom left-hand corner of the screen) **Microsoft Office 2016 > Excel 2016**.

## The Toolbars and Moving Through Cells

Excel 2016 is quite similar to Excel 2010 and you will find only few new features. There are three features that you're going to use most as you work with Excel 2010: the **File menu**, the **Ribbon**, and the **Quick Access Toolbar**.

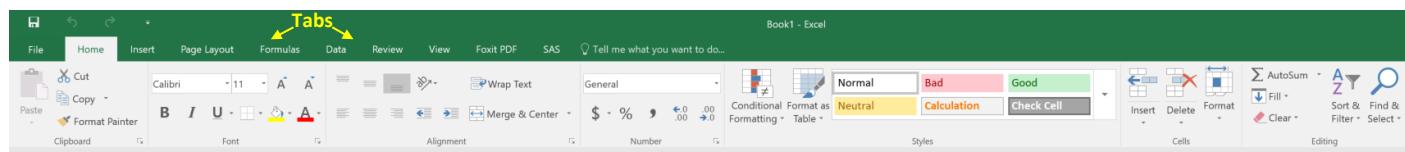
### File menu

From the **File menu** you can create a new workbook, open an existing workbook, save and save as, print, share, and close, etc. using the functions located in the drop down menu.



### Ribbon

The **Ribbon** is a panel beside the File menu which has several tabs. The seven common ones are: Home, Insert, Page Layouts, Formulas, Data, Review, and View, but there might be more depending on the specific application installed on your computer. Each tab is divided into groups which are logical connections of features.



**Home:** Clipboard, Font, Alignment, Number, Styles, Cells, Editing, Webex

**Insert:** Tables, Illustrations, Add-ins, Charts, Tours, Sparklines, Filters, Links, Text, Symbols

**Page Layouts:** Themes, Page Setup, Scale to Fit, Sheet Options, Arrange

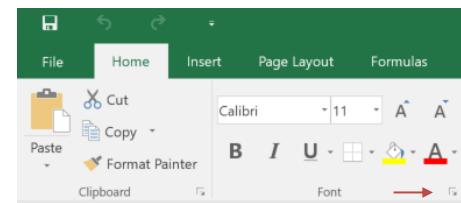
**Formulas:** Function Library, Defined Names, Formula Auditing, Calculation

**Data:** Get External Data, Get & Transform, Connections, Sort & Filter, Data

Tools, Forecast, Outline

**Review:** Proofing, Insights, Language, Comments, Changes, Ink

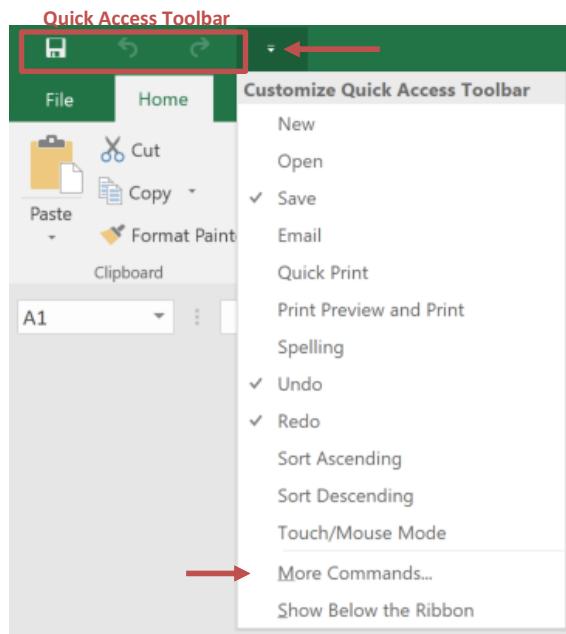
**View:** Workbook Views, Show, Zoom, Window, Macros



Besides commonly utilized features that are displayed on the **Ribbon**, you could see additional features within each group by clicking the arrow at the bottom right corner of each group.

## Quick Access Toolbar

The **Quick Access Toolbar** is a customizable toolbar that contains useful commands. You can customize the buttons shown in the toolbar by clicking on them in the drop down menu and changing the position of the toolbar.



## Mini Toolbar

The **Mini Toolbar** is a floating toolbar that is displayed when you select text or right-click text. It displays common formatting tools, such as Bold, Italics, Fonts, Font Size and Font Color.



## Moving through cells

You now know how to move from one cell to another using the mouse. You can also move within the worksheet using the keyboard.

Movement	Key stroke
One cell up	up arrow key
One cell down	down arrow key or <b>ENTER</b>
One cell left	left arrow key
One cell right	right arrow key or <b>TAB</b>
Top of the worksheet (cell A1)	<b>CTRL+HOME</b>
End of the worksheet (last cell containing data)	<b>CTRL+END</b>
End of the row	<b>CTRL+right arrow key</b>
End of the column	<b>CTRL+down arrow key</b>
Any cell	<b>Home &gt; Editing &gt; Find&amp;Select &gt; Go To</b>

## Modifying a Worksheet

As you add data to your worksheet, you may find you need to modify the layout in various ways:

### Widen or shrink rows or columns

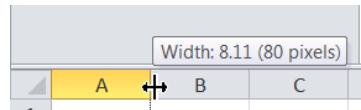
To resize a row:

1. Position your cursor over the boundary line between two rows at the far left of the worksheet. The appearance of your cursor will change from a little arrow to a cross.
2. Left-click and drag to obtain the row size you want and then release.

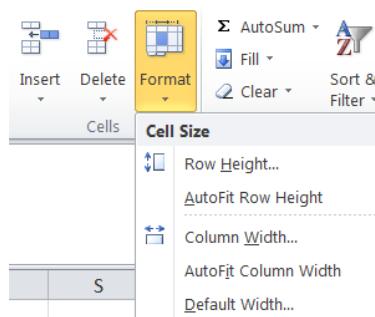


To resize a column:

1. Position your cursor over the boundary line between two columns at the top of the worksheet. The appearance of your cursor will change from a little arrow to a cross.
2. Left click and drag to obtain the column size you want. Release.



*Note: another way to resize a row or a column is by using options in the drop down menu of the Format button in the Cells group of the Home tab.*



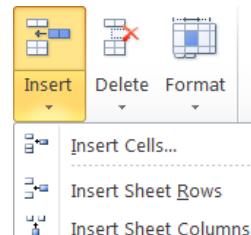
### Highlight a cell or cells

Excel offers some shortcuts for selecting cells:

Cells to select	Mouse action
One cell	click once in the cell
Entire row	click the row label (row number at far left)
Entire column	click the column label (column letter at top)
Entire worksheet	click the whole sheet button (cell just to the left of label "A")
Cluster of cells	drag mouse over the cells holding the left button of the mouse or hold down the SHIFT key while using the arrow keys

### Insert a row above

1. Highlight the row that you would like to be below the new row.
2. From the **Cells** group of the **Home** tab (or right click the highlighted row), choose **Insert > Insert Sheet Rows** (or just **Insert**).
3. Insert a column in a similar manner.



### Move or copy cells

1. Highlight the cells to be moved or copied using the **Cut** or **Copy** buttons from **Clipboard** group of the **Home** tab (or the CTRL-X or CTRL-C keys).
2. Then click the upper left-hand cell of the area you want to move them to and click the **Paste** button (or CTRL-V).

## Freeze panes

This is a wonderful feature! It allows you to retain specific rows for viewing as you scroll through your worksheet (such as row 1 which might contain your variable names) and important columns (such as column 1 which might contain study id).

1. To retain one or more rows:
  - a. Highlight the row that is below the row(s) you want to freeze.
2. To retain one or more columns:
  - a. Highlight the column that is to the right of the column(s) you want to freeze.
3. To retain row(s) and column(s) simultaneously:

	A	B	C	D	E	F	G	H	I
1	Date	Province	Unit	Count	Rate				
2	2019-09-01	1 A		24	0.234275				
3	2019-09-02	1 B		19	0.44462				
4	2019-09-03	1 C		66	0.455987				
5	2019-09-04	1 D		76	0.433866				
6	2019-09-05	1 E		20	0.427256				

a cell that is below the row(s) you want to freeze and to the right of the column(s) you want to freeze.

4. From the **Window** group of the **View** tab, choose **Freeze Panes** > **Freeze Panes** from the menu. (You can also just select **Freeze First Row** or **Freeze First Column** if you only want to freeze one row or column.)
5. You can undo any of these operations by selecting **Freeze Panes** > **Unfreeze Panes**.

*Note: the freeze panes feature is for viewing only. Formatting the printing of a worksheet so that a selection of top rows and left hand columns appears on every page is done using additional features in the **Sheet Options** group of the **Page Layout** tab.*

In the **Page Layout** tab, click on the arrow at the bottom right corner of the **Sheet Options** group to see additional features. You will be able to select which rows you want repeated at the top and which column you want repeated at the left in the **Print titles** options of the **Sheet** tab.

## Add, Remove, and Rename Worksheets

At the bottom of your excel file are tabs for each worksheet in the file. The default is 3 worksheets named "Sheet1", "Sheet2", and "Sheet3".

To add a worksheet:

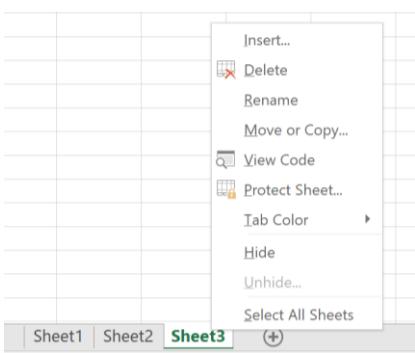
1. Click on the "Insert worksheet" icon  or position your cursor on the tab for one of the worksheets.
2. Right click.
3. From the drop down menu, choose **Insert**.

To remove a worksheet:

1. Position your cursor on the tab for the worksheet you wish to remove.
2. Right click.
3. From the drop down menu, choose **Delete**.

To rename a worksheet:

1. Position your cursor on the tab for the worksheet you wish to rename.
2. Right click.
3. From the drop down menu, choose **Rename**.



## Moving/Copying Worksheets

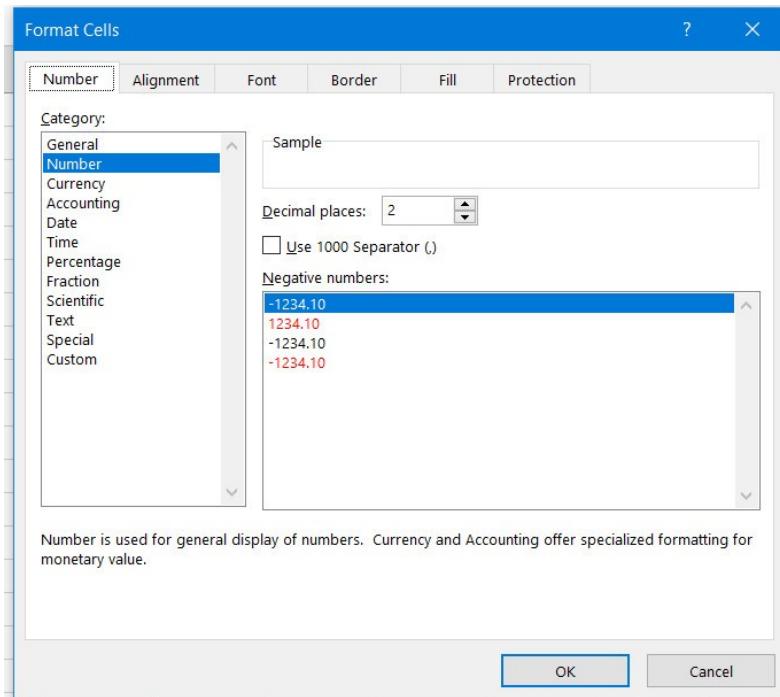
As your worksheets accumulate, you might want to rearrange their order. Alternatively, you might want to work with a copy of a worksheet.

1. Position your cursor on the tab for the worksheet to be moved or copied.
2. Right click.
3. From the drop down menu, choose **Move or Copy**.
4. To move the worksheet, select the **Book** (i.e. name of the spreadsheet) and location (**Before sheet** menu) where you would like it moved.
5. To copy a worksheet, do the same as (4) and select **Create a copy**.

## Formatting Cells

As previously noted, Excel has formatting options for the display of spreadsheet information so that it is readable to us! (e.g., dates, times, percentages or dollars) To format cells, columns of cells, or multiple columns of cells:

1. Highlight the cell you wish to format.
2. Click the **Format** button of **Cells** group of the **Home** tab.
3. Click **Format Cells** from the drop down menu. This will open the dialog box below. It has several tabs. You will be positioned in the **Number** tab.



4. Choose the category that you want to format. Then select from the drop down menus that are provided.
5. The other tabs (Alignment, Font, Border, Fill, and Protection) can be accessed to change the font of text entries, to align entries on the right, left, or center of cells, etc.

## Formulas and Functions

### Creating new variables using calculations

The creation of new variables (or fields) that are the result of calculations is easy.

1. Highlight the cell where you would like the result to be stored.
2. Position your cursor in the **fx** dialog box (see arrow below).
3. When entering a formula, always begin your entry with an equal sign “=”.

	A	B	C	D	E
1	STUDENT	GRADE 1	GRADE 2	GRADE 3	
2	Jane	88	95	79	=B2+C2+D2)/3
3	Robert	79	90	77	
4	Kamil	92	88	83	
5	Linda	90	84	65	
6	Hua	87	89	76	
7	Zeke	95	94	88	
8	Joshua	82	83	75	

Try it: – Calculate the average grade across Grade 1, Grade 2, and Grade 3 for each student.

1. Open file Excel\_101\_Epis.xlsx/ worksheet EX1.
2. We wish to create a new variable that is the average of the values in columns B, C, and D, and we want the result to be stored in column E (outlined by red box in your example spreadsheet).
  - a. Select cell E2 = (note the bold border of cell E2 in the above figure).
  - b. Enter “=(B2+C2+D2)/3” into the **fx** dialog box. (Alternative: enter “=sum(B2:D2)/3”. In doing this you could also highlight cells B2, C2, and D2.)
3. At this point, you have done the calculation of the new variable for just one record. To repeat this calculation for the other rows in your spreadsheet (Robert, Kamil, etc.):
  - a. Highlight the cell that contains the first result; this is cell E2 in this example.
  - b. From the **Clipboard** group of the **Home** tab, choose **Copy** (or CTRL-C).
  - c. Highlight all the destination cells; these will be cells E3, E4, and so on down to the last row in your data set.
  - d. From **Clipboard** group of the **Home** tab, choose **Paste** (or CTRL-V).

OR

- a. Highlight the cell that has the first result; this is cell E2 in this example.
- b. Click the bottom right corner of this cell.
- c. Now drag down through E3, E4, etc. to the last row in your data set.

#### “Built in” Functions

MS Excel has a selection of “Built in” Functions that save typing by hand. In our example above, instead of typing the formula “=(B2+C2+D2)/3” in cell E2, we could have accomplished the same operation by typing “=AVERAGE(B2:D2)”.

The following is a partial listing of the available functions. You could find more in the **Function Library** group of the **Formulas** tab.

Function	Example	Description
SUM	=SUM(A1:100)	calculate the sum of cells A1 through A100
AVERAGE	=AVERAGE(B1:B10)	calculate the average of cells B1 through B10
MEDIAN	=MEDIAN(B1:B10)	calculate the media of cells B1 through B10
MAX	=MAX(C1:C100)	returns the highest value from cells C1 through C100
MIN	=MIN(D1:D100)	returns the lowest value from cells D1 through D100
SQRT	=SQRT(D10)	calculate the square root of the value in cell D10
TODAY	=TODAY()	returns the current date if the cell is formatted to Date (leave the parentheses empty)

Note: If you want to use the value of a certain cell as a fixed parameter in your function, add a dollar sign before both the column character and the row number (e.g., \$A\$1).

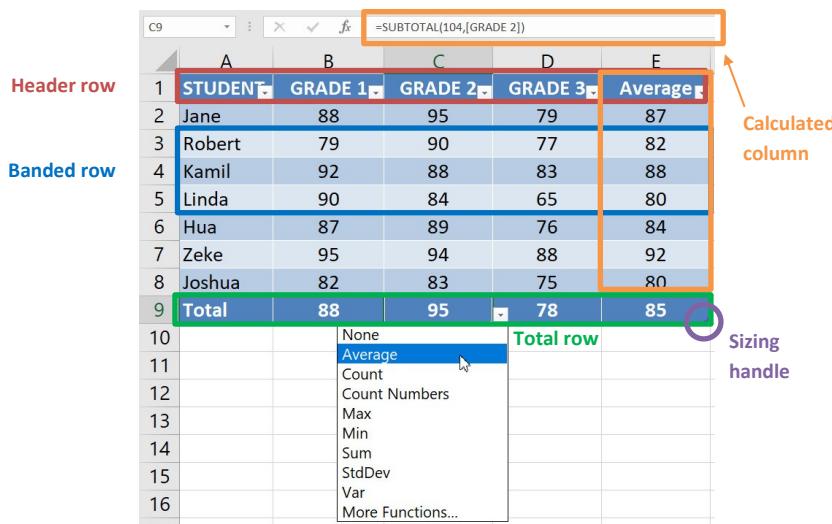
Try it: - Use file Excel\_101\_Epis.xlsx/ worksheet EX2 and try the functions listed above using the AGE variable.

# Tables

To make managing and analyzing a group of related data easier, you can turn a range of cells into a table. A table typically contains related data in a series of worksheet rows and columns that have been formatted as a table. By using the table features, you can then manage the data in the table rows and columns independently from the data in other rows and columns on the worksheet.

A table can include the following elements:

- **Header row:** By default, a table has a header row. Every table column has filtering enabled in the header row so that you can filter or sort your table data quickly.
- **Banded rows:** By default, alternate shading or banding has been applied to the rows in a table to better distinguish the data.
- **Calculated columns:** By entering a formula in one cell in a table column, you can create a calculated column in which that formula is instantly applied to all other cells in that table column.
- **Total row:** You can add a total row to your table that provides access to summary functions (such as the AVERAGE, COUNT, or SUM function). A drop-down list appears in each total row cell so that you can quickly calculate the totals that you want.
- **Sizing handle:** A sizing handle in the lower-right corner of the table allows you to drag the table to the size that you want by adding columns or rows to the table.

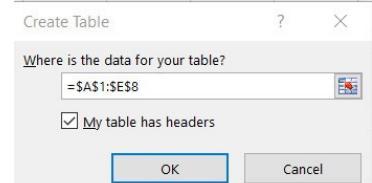
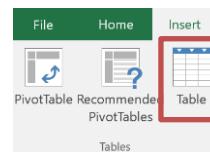


	A	B	C	D	E
1	STUDENT	GRADE 1	GRADE 2	GRADE 3	Average
2	Jane	88	95	79	87
3	Robert	79	90	77	82
4	Kamil	92	88	83	88
5	Linda	90	84	65	80
6	Hua	87	89	76	84
7	Zeke	95	94	88	92
8	Joshua	82	83	75	80
9	Total	88	95	78	85
10		None			
11		Average	Total row		
12		Count			
13		Count Numbers			
14		Max			
15		Min			
16		Sum			
--		StdDev			
--		Var			
--		More Functions...			

## Creating a table

When you create a table in an Excel worksheet, it's not only easier for you to manage and analyze its data, but you also get built-in filtering, sorting, and row shading.

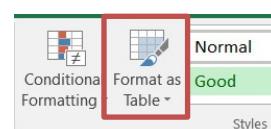
1. Select the range of cells you want included in the table.
2. The cells can be empty or can contain data.
3. In the **Insert** tab, select **Table**
4. If the top row of the selected range contains data you want to use as table headers, check the **My table has headers** box. If you don't check the **My table has headers** box, table headers with default names like "Column1" and "Column2" are added to your table above its data. You can modify default header names at any time.



## Table Style options

Create a table in the style you want:

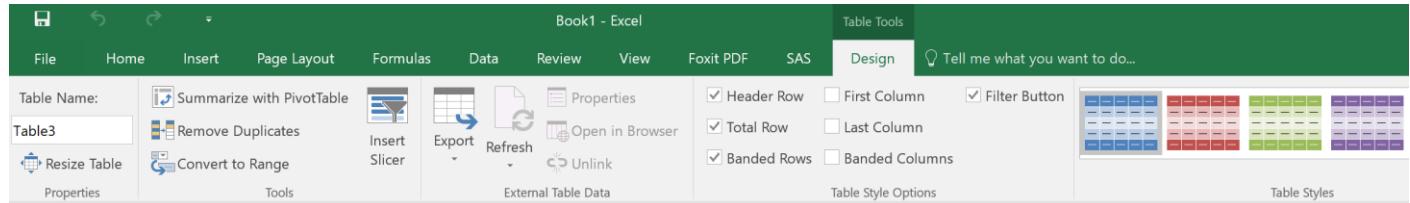
1. Select the range of cells you want included in the table.
2. On the **Home** tab, click **Styles** and then select **Format as Table**.
3. When you use **Format as Table**, Excel automatically inserts a table.



You can create your own table style to use in the current workbook. After you create a custom table style, it's available for the current workbook when you use Format as Table, under Custom. For more information, see Format an Excel table.

### More about tables

After you create a table, the **Table Tools** become available in the ribbon, and you'll see a **Design** tab. You can use the tools on the **Design** tab to customize or edit the table. Note that the **Design** tab is visible only when at least one or more cells in the table are selected.



You can add a Total Row by checking off that option in the **Table Style Options** menu of the **Design** tab. The Total Row provides access to summary functions (such as the AVERAGE, COUNT, or SUM function). A drop-down list appears in each total row cell so that you can quickly calculate the totals that you want.

To add a row, select the last cell in the last row of the table (not the total row), and press Tab.

In the example shown below, pressing Tab with cell D8 selected expands the table to a 9th row, and moves the selection to the first column in the new row, cell A9

	A	B	C	D
1	STUDENT	GRADE 1	GRADE 2	GRADE 3
2	Jane	88	95	79
3	Robert	79	90	77
4	Kamil	92	88	83
5		90	84	65
6	Tab	87	89	76
7	Zeke	95	94	88
8	Joshua	82	83	75
9	Total	88	89	78

	A	B	C	D
1	STUDENT	GRADE 1	GRADE 2	GRADE 3
2	Jane	88	95	79
3	Robert	79	90	77
4	Kamil	92	88	83
5	Linda	90	84	65
6	Hua	87	89	76
7	Zeke	95	94	88
8	Joshua	82	83	75
9				
10	Total	88	89	78

To add a column, put your cursor in the title of a column and select the right arrow. Select the insert option and decide on which side of the selected column you want the new column to be. To change the title of your new column, select the title of the column and enter its name in the **fx** dialog box.

**Try it:** - Use file **Excel\_101\_Epis.xlsx/worksheet EX3** and create a table

## Sorting

Sorting data is an integral part of data analysis. You might want to arrange a list of names in alphabetical order, compile a list of product inventory levels from highest to lowest, or order rows by colors or icons. Sorting data helps you quickly visualize and understand your data better, organize and find the data that you want, and ultimately make more effective decisions.

You can sort data by text (A to Z or Z to A), numbers (smallest to largest or largest to smallest), and dates and times (oldest to newest and newest to oldest) in one or more columns. You can also sort by a custom list (such as Large, Medium, and Small) or by format, including cell color, font color, or icon set. Most sort operations are column sorts, but you can also sort by rows.

### Basic Sorts

To execute a basic descending or ascending sort based on one column from a table:

1. Click on the  button at the top of the column
2. Excel will recognize the type of data in the column.
  - For text, it will ask you to choose **Sort A to Z** or **Sort Z to A**
  - For number, it will ask you to choose **Sort Smallest to Largest** or **Sort Largest to Smallest**
  - For date, it will ask you to choose between **Sort Oldest to Newest** or **Sort Newest to Oldest**

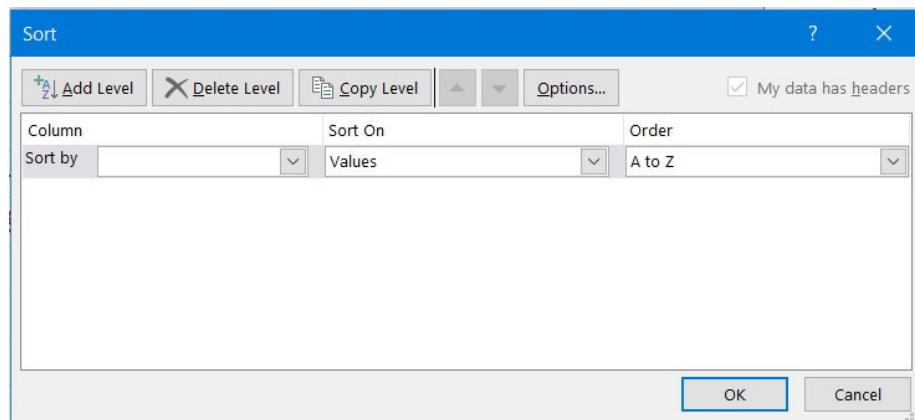
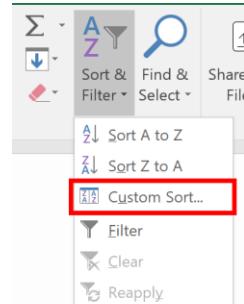
Try it: - Use file **Excel\_101\_Epis.xlsx/ worksheet EX3** and basic sort by **FAV\_COLOUR**.

### Custom Sorts

To sort on more than one column at a time

(e.g., sort by sex and then age):

1. Click on any cell of the table
2. In the **Editing** group of the **Home** tab, click on the **Sort & Filter** button and select **Custom Sort**.
3. Choose the **Column**, the basis on which to sort (**Sort on**), and the **Order** you want to sort by first.
4. Click the **Add Level** button and choose the second column you would like to sort the data on.
5. Click **OK** when you have added all the columns in the order you want to sort by.



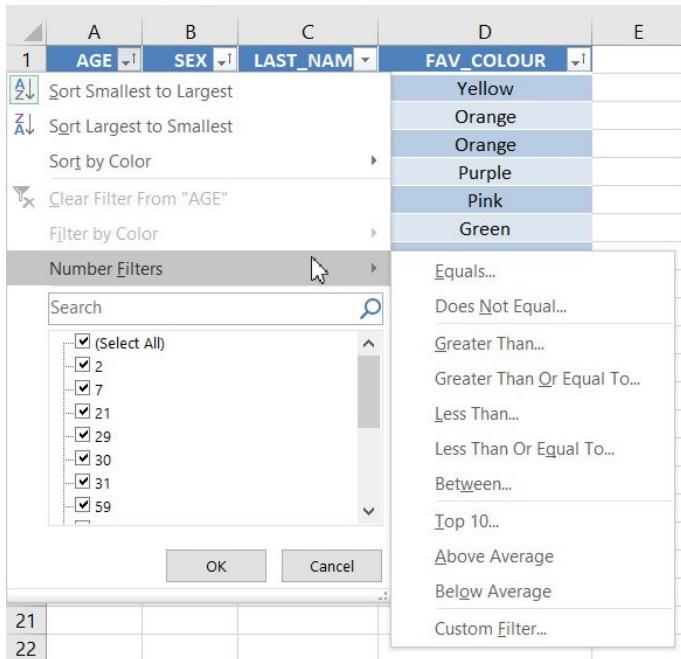
Try it: - Use file **Excel\_101\_Epis.xlsx/ worksheet EX3** and custom sort by **SEX** and **AGE**.

AGE	SEX	LAST_NAME	FAV_COLOUR
31	F	Leblanc	Green
59	F	Smith	Green
64	F	Lee	Blue
72	F	Johnson	Blue
79	F	Johnson	Red
90	F	Tremblay	Pink
94	F	Green	Black
2	M	Marshall	Yellow
7	M	Ashley	Orange
21	M	Mathews	Orange
29	M	Frosst	Purple
30	M	Leblanc	Pink
65	M	Lee	Brown
74	M	Kelly	Blue
97	M	Jones	Green

## Filtering

To display only data that meets certain criteria:

1. Click the  button at the top of the column that contain the data you wish to filter.
2. Click the items you wish to filter or Number/Text Filters to put a rule (e.g. Greater Than 19)
3. To clear the filter, choose **Clear Filter from** in the menu



The screenshot shows a portion of an Excel spreadsheet with columns labeled A through E. The 'FAV\_COLOUR' column has a filter applied, displaying a list of colors: Yellow, Orange, Orange, Purple, Pink, and Green. A context menu is open over the 'FAV\_COLOUR' header, showing options like 'Sort Smallest to Largest', 'Sort Largest to Smallest', 'Sort by Color', 'Clear Filter From "AGE"', 'Filter by Color', and 'Number Filters'. The 'Number Filters' option is selected, opening a dialog box. This dialog box contains a search bar, a list of numbers from 2 to 59 each with a checkbox, and buttons for 'OK' and 'Cancel'. A red arrow points to the 'OK' button in the bottom left corner of the dialog box.

Try it: - Use file [Excel\\_101\\_Epis.xlsx](#)/ worksheet EX3 and filter for females only within the SEX variable.

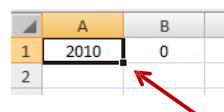
## Auto-Filling and Fill Series

### Auto-Filling

Excel has an auto-filling feature that lets you replicate a given entry into multiple cells in a column. This can be very handy.

Suppose you would like to replicate the A1 cell entry of “2010” into cells A2 through A12.

1. Click on 2010 in cell A1.
2. A small black square in the lower right corner of the cell should appear.
3. Click and drag down A3, A4 and so on to A12.
4. When you release the mouse, notice that all highlighted cells now contain 2010, and a small **Auto-Fill Options** button appears.



Try it: - Use file [Excel\\_101\\_Epis.xlsx](#)/ worksheet EX4 (column A).

### Fill Series

Excel can also save you time if you need to enter a regular series of numbers, days of the week, etc. Suppose you want to add 1 to each successive cell down a column, starting with 0:

1. Enter **0** in the top row (e.g., B1), **1** in the second row (e.g., B2).
2. Highlight the two cells, click on the lower right corner of the section, as above, and drag down the column. Excel will fill the series, adding 1 to each successive cell.

**Try it:** – Use file **Excel\_101\_Epis.xlsx/ worksheet EX4 (column B)**.

Suppose you want to add 5 to each successive cell down a column, starting with 0:

1. Enter **0** in the top row (e.g., C1), **5** in the second row (e.g., C2).
2. Highlight the two cells, click on the lower right corner of the section, as above, and drag down the column. Excel will fill the series, adding 5 to each successive cell.

**Try it:** – Use file **Excel\_101\_Epis.xlsx/ worksheet EX4 (column C)**.

Excel can also recognize some words such as month and day of the week. Suppose you want to look at the various months in 2010.

1. Enter **Jan** in the top row (e.g., D1)
2. Highlight the cell, click on the lower right corner of the cell and drag down the column. Excel will fill the series with each month of the year.

**Try it:** – Use file **Excel\_101\_Epis.xlsx/ worksheet EX4 (column D)**.

## Page Setup and Printing

Before you do any printing, specify your page layout. Click the **Dialog Box arrow** in the **Page Setup** group of the **Page Layout** tab. Four tabs with a variety of menus will appear:

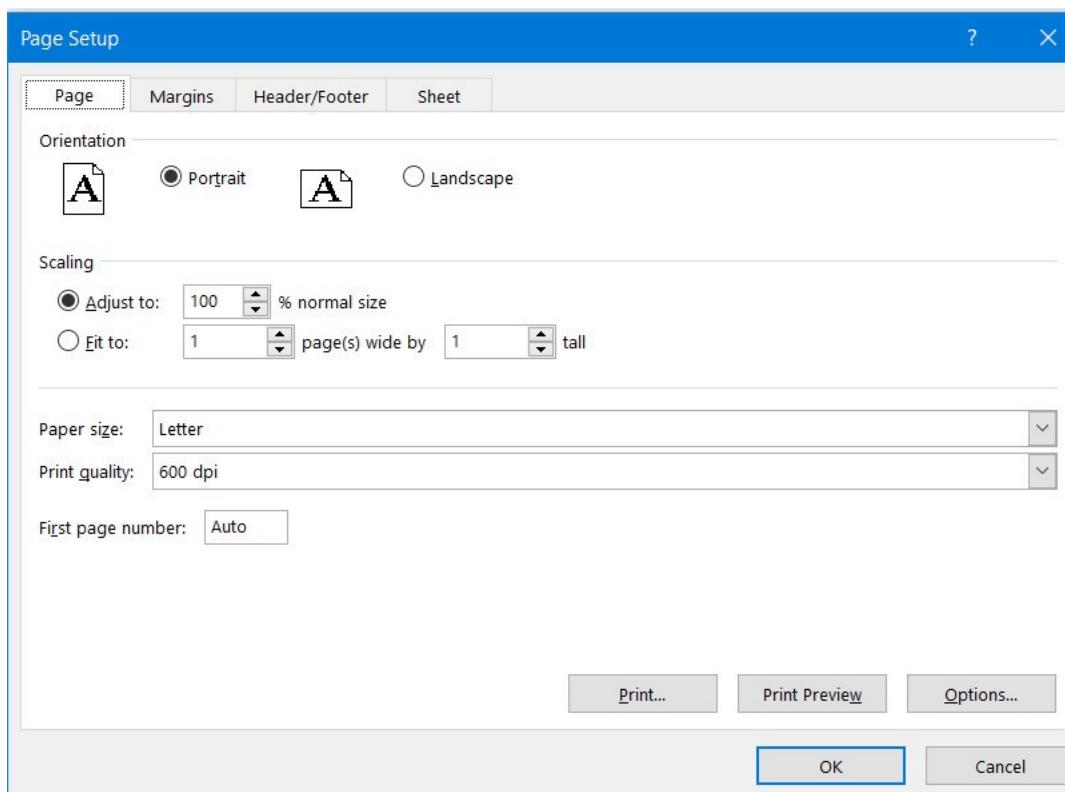
**Page:** Choose page orientation (Portrait or Landscape). If you do not have too many columns, choose the option "Fit to 1 page wide" so that all of your variables appear on one page.

**Margins:** Use this tab to change the margins of your page.

**Header/Footer Tab:** Use this tab to specify custom headers and footers. A good practice is to use headers and footers to document your name, date, file name, analysis code program names, etc.

**Sheet Tab:** Use this tab to choose rows to be repeated at the top of each printed page and columns to appear at the left of each printed page. The **Sheet** tab also allows you to choose whether or not to show gridlines in your printed table.

The **Print Preview** button will let you see just what your worksheet will look like when printed. When you are satisfied with the appearance of your worksheet, select **Print**.



## Creating a Data Set

Creating a data set is best explained in an example (adapted from Liz Procter-Gray, Carol Bigelow and Yiwei Jiang's "ICU Example"). Let's say we are studying 25 consecutive patients who entered the general medical/surgical intensive care unit at a large urban hospital. For each patient, the following data were collected:

Variable Name	Description	Code
ID	Personal Health Number	unique
AGE	Age (years)	numeric
TYPE_ADM	Type of Admission	1 = emergency 0 = elective
ICU_TYPE	ICU Type	1 = medical 2 = surgical 3 = cardiac 4 = other
SBP	Systolic Blood Pressure (mm Hg)	numeric
ICU_LOS	Number of days in ICU	integer
VIT_STAT	Vital Status at Discharge	1 = dead 0 = alive

We will use this example of 25 shown below:

ID	Age	Type_Ad	ICU_Type	SBP	ICU_LOS	Vit_Stat
1	15	1	1	100	4	0
2	31	1	2	120	1	0
3	75	0	1	140	13	1
4	52	0	1	110	1	0
5	84	0	4	80	6	0
6	19	1	1	130	2	0
7	79	0	1	90	7	0
8	74	1	4	60	1	1
9	78	0	1	90	28	0
10	76	1	1	130	7	0
11	29	1	2	90	13	0
12	39	0	2	130	1	0
13	53	1	3	250	11	0
14	76	1	3	80	3	1
15	56	1	3	105	5	1
16	85	1	1	145	4	0
17	65	1	1	70	10	0
18	53	0	2	130	2	0
19	75	0	3	80	34	1
20	77	0	1	130	20	0
21	52	0	2	210	3	0
22	19	0	1	80	1	1
23	34	0	3	90	3	0
24	56	0	1	185	3	1
25	71	0	2	140	1	1

## Designing Your Database

Before entering data into an excel spreadsheet, or any database application, the file's structure must first be defined. Exactly how this is done varies by software type, but the following components are available in good database software:

**Name of field (also your variable name):** a single word name used as a shorthand reference for a field.

- *Keep it short (8 characters or less is recommended, though not required).*
- *Avoid special characters such as #,-,\*,... While some software will allow these special characters in a name, others will not, creating problems when you transfer data between formats.*
- *Avoid spaces in a name for the same reason. Use an underscore ( \_ ) in place of a space.*

**Label for field (optional):** A longer description of data stored in the field.

**Type of field:** There are two basic types of fields that dictate the manner in which data are stored (text/character and numeric). Other field formats are often available too (e.g., date).

**Numeric:** Allows numbers only.

**Text or Character:** Allows letters, numbers, and other keyboard characters.

**Logical:** Yes/No or True/False.

**Date:** Contains dates in a specified format (in some programs dates are stored as character data, in others, numeric)

**Format for field:** Specifies the number of digits or spaces available for entering and displaying data, or other specialized formats.

- Numeric formats specify the number of digits before and after the decimal place.
- Character formats typically define the number of spaces or columns needed.
- Date and Date/Time formats specify the order (e.g., month/day vs. day/month) and presentation of data (e.g. 07JUN2001 vs. 06/07/2001).

**Data type:** It is necessary to define the data type for each field. In Excel, relates to formatting cells (see p 10).

*Some notes on data type:*

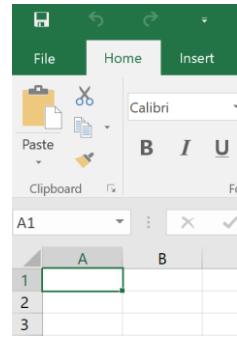
- *Numeric and character data are stored quite differently, and you should be clear ahead of time as to the data type required.*
- *Numbers can be stored in character fields. Don't do this!! It can cause great confusion later in the data management process if you think you have numeric data and attempt computations when the field was defined as character.*
- *It is not always obvious when data should be numeric and when it should be character. For example, while a phone number or social security number could be entered as numeric data, you will never want to compute with these numbers -- they serve as ID or identifier types of variables. By entering these as character data, you can include hyphens (e.g., 545-1000 or 999-99-9999), and when printed they will appear in a familiar format.*
- *There are also occasions when numbers are clearly codes and can be entered as character data since you will never compute with these numbers (such as 1=White, 2=Black, 3=Asian, 4=other). However in some situations it may be advantageous to enter these as numeric data. Some statistical applications (e.g., Minitab) will not allow character variables in analyses -- even if the variable is used solely to define groups. If you know this is true of the software you will be using for analysis, plan accordingly. If you have defined a variable as character and need numeric data or vice versa, it is always possible to convert the data or create a new variable in the required format from the values of the current one, but planning ahead saves work.*
- *Pay attention to dates – some applications store dates as character data and others as numeric. This affects how information is transferred between programs, and you will often need to do some special programming to handle dates. This is particularly important if you will be using dates to compute durations (e.g., length of stay in hospital, time between patient interviews).*

**Some programs have other special field types (e.g., currency, percent, phone numbers, SSN).** In some software these are considered formats rather than field types.

## Data Entry

You are now ready to create your dataset. It's easiest to explain by example...

Try it: – Use file **Excel\_101\_Epis.xlsx/ worksheet Dataset\_Example.**



### Step 1: Create your data set

1. You should see an empty spreadsheet and the cell A1 with a bold border. Cell A1 is the active cell; your cursor (you can't actually see it) is positioned here.
2. To enter your variable names, enter the column headings as shown in the data table on page 17 (these are your fields and correspond to your variable names). Proceeding horizontally across the first row, type the variable names in cells A1, B1,..., G1. Use the right arrow key after each entry so that your cursor moves right along the horizontal. You should now have the following:

	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu-los	vit_stat	
2								
3								

3. Convert the dataset in a table. To do this, select any of the variable names and select the **Table** in the **Insert** tab. Ensure to check “My table has headers”.
4. Now enter your data, column by column. To do this, begin by highlighting cell A2. Type a “1” in this cell (this is value of ID for the first record) and then press the down arrow key. Having pressed the down arrow key, you can now use the enter key after each data entry. When you are done, you should have the following (*only a partial picture is shown here*):

	A	B	C	D	E	F	G
1	ID	Age	Type_Adm	ICU_Typ	SBP	ICU_LO	Vit_Sta
2	1	15	1	1	100	4	0
3	2	31	1	2	120	1	0
4	3	75	0	1	140	13	1
5	4	52	0	1	110	1	0
6	5	84	0	4	80	6	0
7	6	19	1	1	130	2	0
8	7	79	0	1	90	7	0
9	8	74	1	4	60	1	1
10	9	78	0	1	90	28	0

## Step 2: Deal with missing data

It is good practice to create a standard entry value for missing data (i.e., when you do not have a value for one or more cells). Missing data are generally indicated by a period ( . ) or code such as "99".

*Note: do not use a numeric value for missing data in a numeric field. For example, do not use the value „99“ for missing values in the AGE variable; this value could accidentally be included in calculations of average age, range, standard deviation, etc. It is better to use a period for numeric variables.*

## Step 3: Format fields

Refer back to instructions on page 16 for formatting fields.

The following are reasonable choices:

Column	Variable	Format cells category	Notes
A	id	General	Two decimal places with or without 1000 separator
B	age	Number	
C	type_adm	Number	
D	icu_type	Number	
E	sbp	Number	
F	icu_los	Number	
G	vit_stat	Number	

*Tip! Except for the variable ID, each variable in this example is formatted as numeric. This makes programming convenient, as it spares having to remember special conventions in working with character fields.*

You should now have the following: (only a partial picture is shown.)

	A	B	C	D	E	F	G
1	ID	Age	Type_Adm	ICU_Typ	SBP	ICU_LO	Vit_Sta
2	1	15.00	1.00	1.00	100.00	4.00	0.00
3	2	31.00	1.00	2.00	120.00	1.00	0.00
4	3	75.00	0.00	1.00	140.00	13.00	1.00
5	4	52.00	0.00	1.00	110.00	1.00	0.00
6	5	84.00	0.00	4.00	80.00	6.00	0.00
7	6	19.00	1.00	1.00	130.00	2.00	0.00
8	7	79.00	0.00	1.00	90.00	7.00	0.00
9	8	74.00	1.00	4.00	60.00	1.00	1.00
10	9	78.00	0.00	1.00	90.00	28.00	0.00

Compare the dataset you just created to Excel\_101\_Epis.xlsx/ worksheet Dataset\_Example\_Complete.

## Step 4: Create a Data Dictionary

Always create a data dictionary (called a 'Coding Manual' in the example below) when you create a new database. A data dictionary is a coding manual for the data variable names, labels, type, value labels and a notes column.

*Tip: Be sure to include missing value codes in your data dictionary*

This step has already been done for you -see [Excel\\_101\\_Epis.xlsx/ worksheet DataDictionary](#).

Coding Manual ICU Study (n=25)				
<u>Variable</u>	<u>Label</u>	<u>Type</u>	<u>Coding</u>	<u>Remarks</u>
<b>id</b>	Patient ID	character	Unique (1, 2, etc.)	
<b>age</b>	Age at admission, years	numeric	.=missing	
<b>type_adm</b>	Admission type	numeric	1=emergency 0=elective .=missing	
<b>icu_type</b>	ICU admission type	numeric	1-medical 2=surgical 3=cardiac 4=other .=missing	
<b>sbp</b>	systolic blood pressure, mm Hg	numeric	.=missing	
<b>icu_los</b>	Length of stay ICU, days	numeric	.=missing	
<b>vit_stat</b>	Status at discharge	numeric	1=dead 0=alive .=missing	

# Creating Charts and Epidemic Curves

MS Excel can be used to produce some very nice graphs (called “charts” in Excel). They are not “canned” graphs. Rather they are selected graphs that are often of interest in epidemiology. This overview covers the basics of creating charts. You will learn how to create more epi-specific graphs in the Data Manipulation session.

## Basics of creating a chart

It's easiest to explain by an example. Let's say we want to make a chart of the number of cases of Disease X by year.

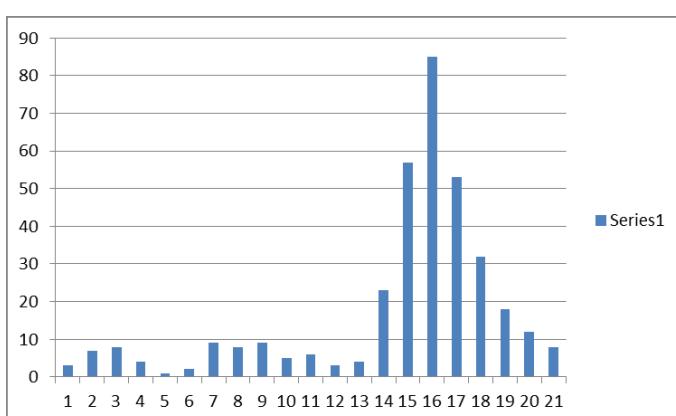
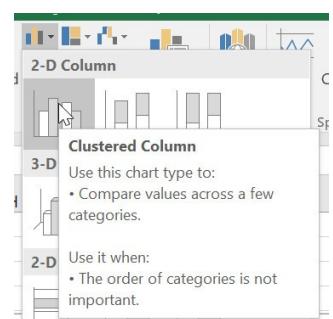
Try it: – Use file [Excel\\_101\\_Epis.xlsx/worksheet CHART\\_EX](#).

### Step 1: Creating your chart

1. Open your worksheet containing the data you wish to display in your chart (for this example, use [Excel\\_101\\_Epis.xlsx/ worksheet CHART\\_EX](#)).
2. Highlight only the data you wish to show in your chart (do not highlight variable names or axis labels).

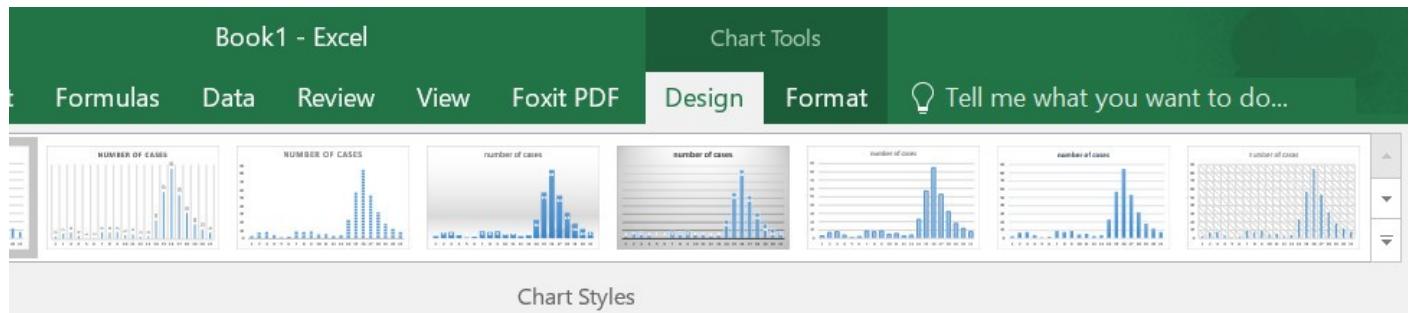
	A	B
1	year	number of cases
2	1947	3
3	1948	7
4	1949	8
5	1950	4
6	1951	1
7	1952	2
8	1953	9
9	1954	8
10	1955	9
11	1956	5
12	1957	6

3. Click the **Insert** tab on the ribbon; select the desired chart type (e.g., column, line, bar) in the **Charts** group.
4. In the drop down menu, choose the desired sub-type of chart you wish to create (for example, **2-D Clustered Column** as shown).
5. Right after you've clicked the picture, you should see an unsatisfying and plain-looking chart that may or may not display the data in the way you intended.

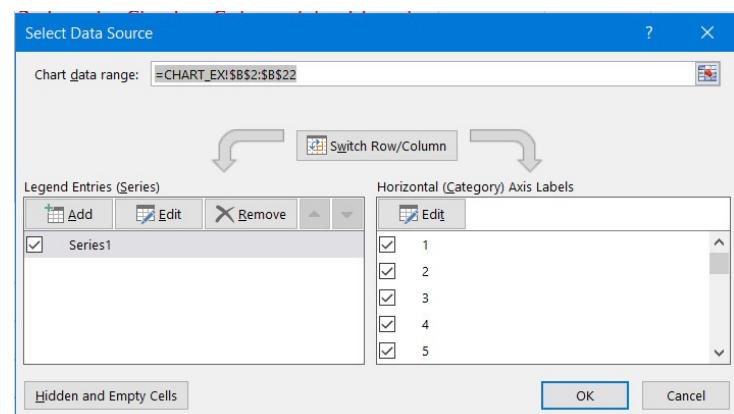


## Step 2: Basic formatting

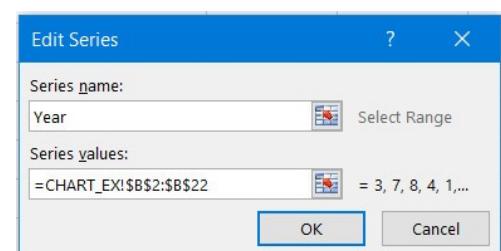
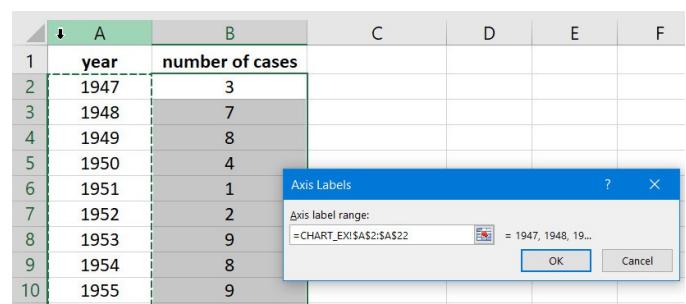
1. Next, use the **Design**, and **Format** tabs under the **Chart Tools** menu to format your chart.



2. To orient your data along the correct axis, choose the **Design** tab under **Chart Tools**. Click the **Select Data** button.
3. To change the chart data range, click on the button and select the data you wish include in the chart.
4. Use the **Add**, **Edit**, and **Remove** buttons to add/delete/modify variable names and data ranges.

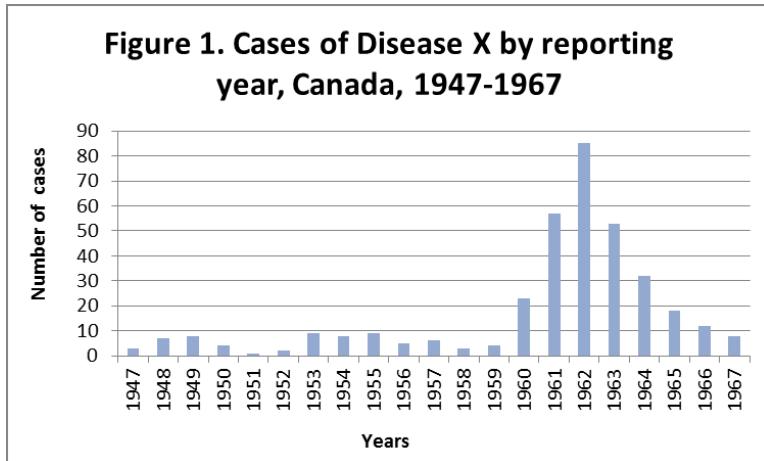


5. In this case, to add the year labels to the x-axis, click on **Edit** under the **Horizontal (Category) Axis Labels** menu.
6. Click on the button and select the range of data cells that contain the year labels in the worksheet. Click **OK**.
7. To add the variable name to the chart, highlight the generic variable name (e.g., Series 1) and click on **Edit** under the **Legend Entry (series)** menu.
8. Type the variable name into the **Series name** box or click on the button to select the name directly from the worksheet.



9. To add a chart title and axis labels, choose the **Design** tab under **Chart Tools**. Click the **Add Chart Element** button and select the **Axis Titles** buttons. Pick the Primary Horizontal, the Primary Vertical or both options. Type your title into the text boxes. You can also choose the position you would like the titles to be in each menu of the buttons.

10. Your chart should now look something like this....

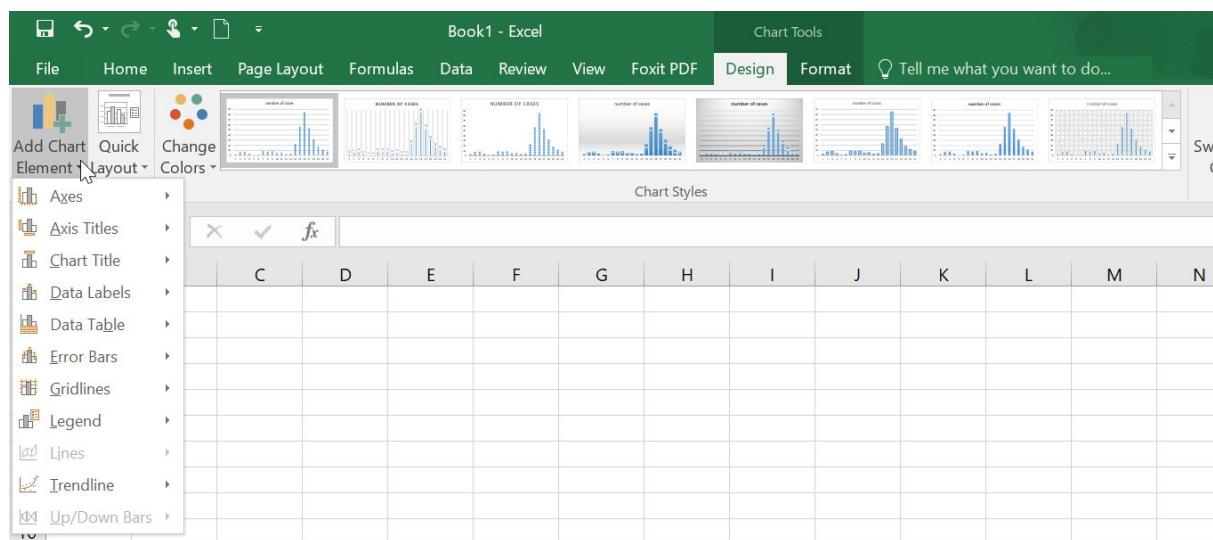


### Step 3: Aesthetics

There are lots of things you can do at this point. Play with the different functions and formatting options under **Add Chart Elements** on the **Design** tab in the **Chart Tools** menu bar.

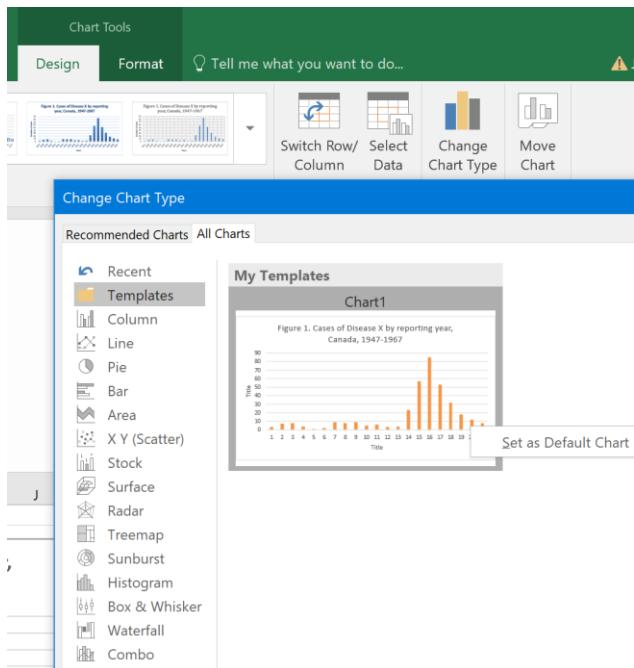
For example, you can:

- Change the colours of the bars, plot area, and outline.
- Change font type and size.
- Add/remove/re-position the legend.
- Add/remove gridlines.
- Change the scales of the x and y axes.
- Add data tables and trend lines.
- And much more.....



*Tip: You can also access chart formatting options by clicking on the chart and selecting the green cross that appears on the right side.*

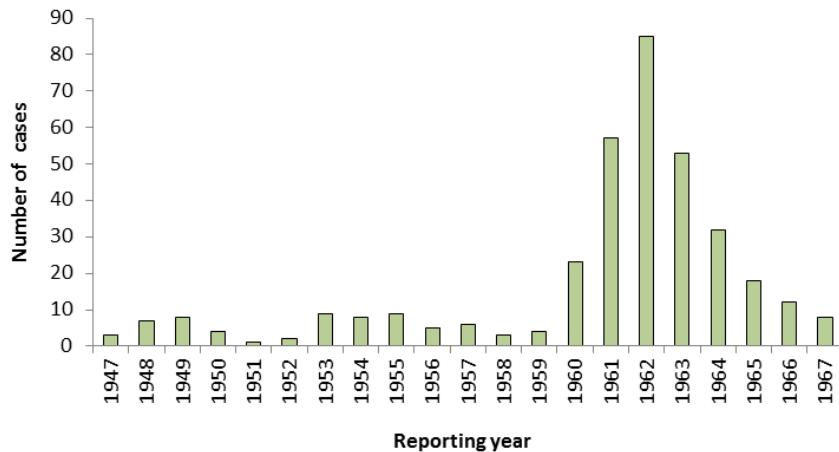




**Another tip!! You can save your custom formatting as a graph template for future use by right clicking your chart and selecting **Save as Template**. Apply your template by selecting your new chart and navigating to **Templates** under **Change Chart Type** on the **Design** tab in the **Chart Tools** menu. Set your template as the default chart type by right clicking on your chart template in the **Change Chart Types** window and clicking on **Set as Default**.**

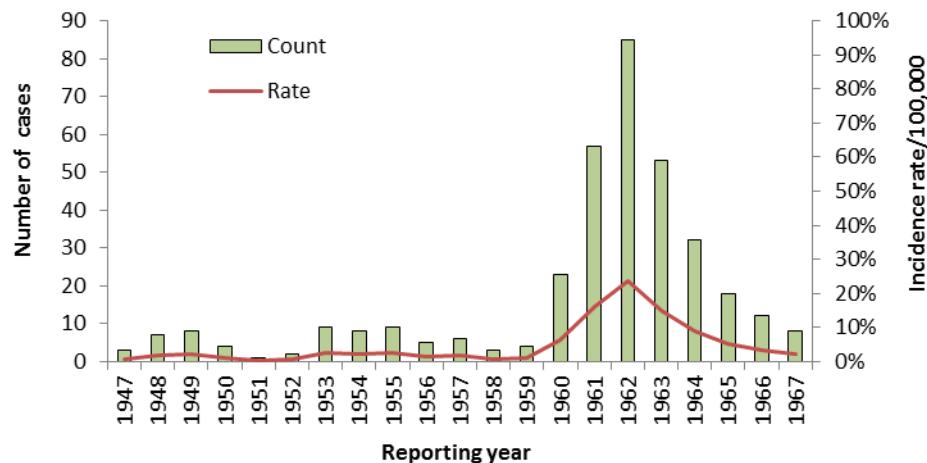
Here is the same chart from above after a bit of re-formatting:

**Figure 1. Cases of Disease X by reporting year, Canada, 1947-1967**



You can also plot data on a second axis. For example, here is the same chart from above with the corresponding incidence rate plotted on a secondary y axis (steps not shown):

**Figure 2. Cases and incidence rate of Disease X by reporting year,  
Canada, 1947-1967**



## Charting Do's and Don'ts

### DO....

- Keep it simple!
- Limit the number of variables shown in one graph
- Use an appropriate scale that displays data along the full spectrum of X and Y axes
- Use colours or textures that will print well either in colour or in black and white
- Include a title and label the axes and legend properly, clearly, concisely
- Select the appropriate type of graph for your data
- Include footnotes when necessary

### DON'T....

- Include several ideas on one graph
- Include redundant data (data duplicated in another table or graph)
- Use fancy 3-dimensional proportions, colours or textured patterns or other non-standard formats
- Use a high ink to data ratio
- Use non-standard or unexplained acronyms

# Creating PivotTables

## Basics of creating a PivotTable

PivotTables are used to quickly and easily summarize, analyze, explore and present data. The following example will take you through the basics of PivotTable creation.

Try it: – Use file **Excel\_101\_Epis.xlsx/ worksheet PIVOT\_Data**.

### Step 1: Format the data to be used in the PivotTable

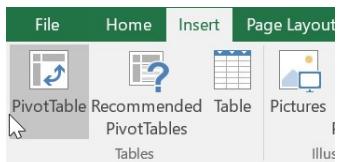
1. It's good practice to use a copy of your master data just in case something goes wrong. *For this exercise, make a copy of the PIVOT\_Data worksheet and work from the copy.*
2. It is important to check the consistent formatting of your data (e.g., date fields, etc.).
3. All the columns must be titled. Ensure that variable names are listed in Row 1.
4. If you have a large worksheet with many variables, then delete any unnecessary columns/variables.

*Tip! Keep it simple!!*

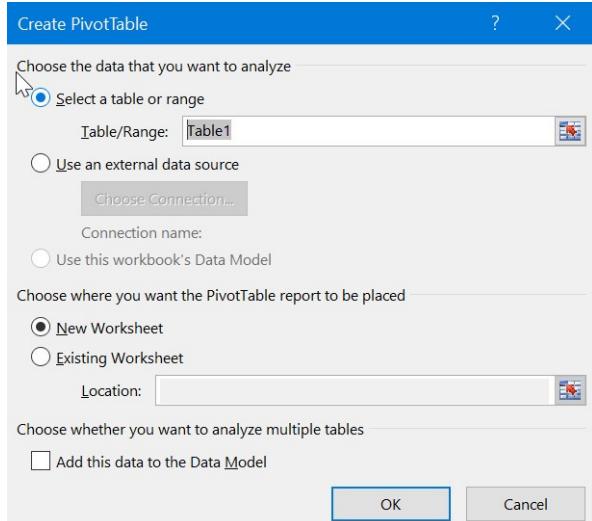
### Step 2: Create a PivotTable

Ideally, you should start by creating a table with your data so the pivot table refers to the table as a whole rather than just a range of data. By doing this, it will ensure that any new entry in the table is included in the pivot table without having to readjust the range of data.

1. Highlight the data you want to analyze (or if you are starting from a table, click on any of the cells within the table)
2. Select the **PivotTable** button in the **Tables** menu under the **Insert** tab



3. Double check that your **Table/Range** of data is correct (i.e., that all data that you wish to summarize in your table has been highlighted).
4. Select **New Worksheet** (or **Existing Worksheet** if you are already working in a PivotTable). Click **OK**.



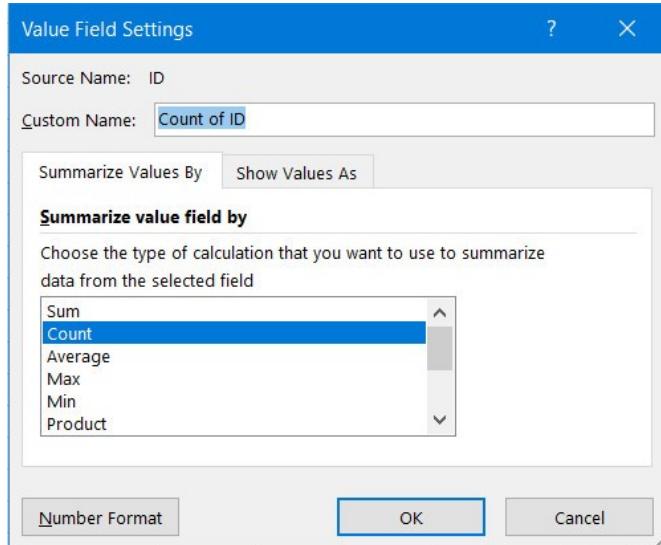
5. You will be taken to a new sheet with the PivotTable shell. There will also be a **PivotTable Field List** box. This contains all the fields for your selected dataset and an 'areas' section. You can pick any one of the items and drag it into one of the four areas in the PivotTable (**Report Filter**, **Row Labels**, **Column Labels**, **Values**) or into the 'areas' under the field list.

- For your **Values** variable (i.e. data items) use a variable that is complete in your master list (i.e., no blanks) and is unique for each case in your dataset. Case ID or Personal Health Number often works well. For this example, we will use ID.
- Pick variables from the **Field List** and drag into **Rows** and **Columns** to summarize your data into columns and rows.
- For this exercise, drag AGE\_GRP into **Rows** and drag ID into the **Values** field. Now drag SEX into the **Columns** area. Your PivotTable should now look like this:

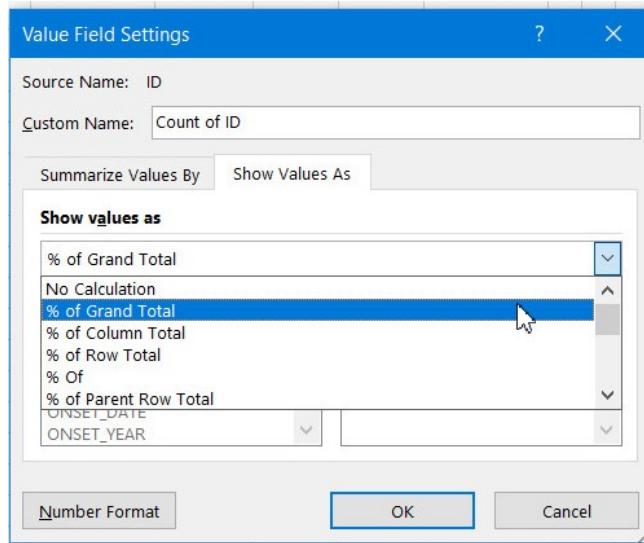
Count of ID		Column Labels	
Row Labels		F	M
15-19		1	1
20-24		5	5
25-29		1	3
30-34		3	3
35-39		3	4
40-44		2	5
45-49			7
50-54		1	3
5-9		1	
60-64		1	
65-69		1	
<b>Grand Total</b>		<b>19</b>	<b>31</b>
			<b>50</b>

*Note: When selecting your **Values** ensure that the data are complete and that there are no blanks because this is the data that is used to summarize the 'count'*

*Note: Make sure that you are ‘counting’ and ‘not adding’ the values (e.g., the case ID#’s). If Excel is adding the values, right-click on the **Values** area (the part of the table that contains the data), select the **value field settings** tab and ensure that ‘count’ is selected. You can also verify this by looking in the Σ value section in the PivotTable Field List.*

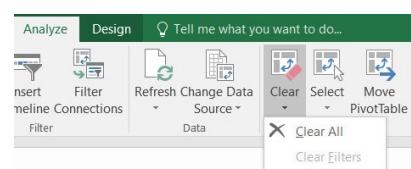


6. You can also use PivotTables to calculate proportions. Follow the same steps as above, then:
  - a. Click in the **Values** area
  - b. Click on **Value Field Settings** then **Show values as** then select “% of Grand Total”



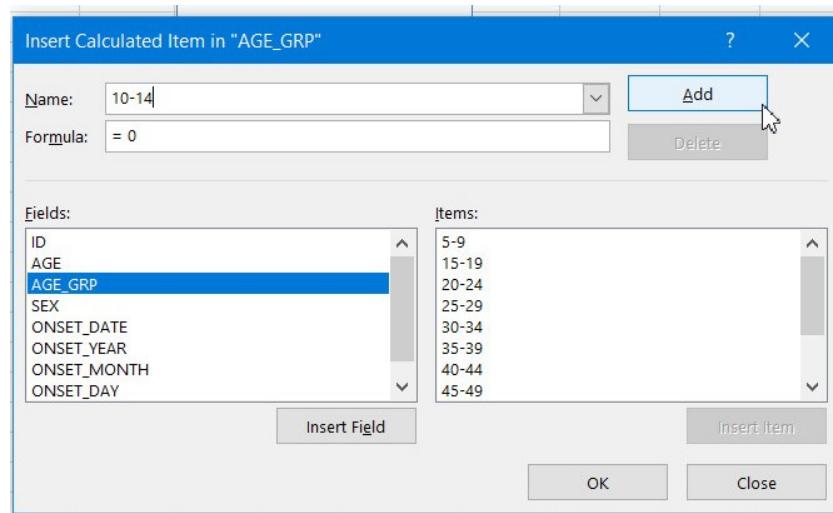
7. Continue to experiment by dragging other items into the PivotTable. To remove an item from the PivotTable, simply select it and drag it out of the table.

*Note: If needed, you can clear your experimentation before continuing with the following steps by first clicking anywhere on the PivotTable. Then, click on the **Clear** button of the **Actions** option in the **Analyze** tab. Select the **Clear All** option and repeat step 5. This reset step might be helpful if you are struggling with some of the following steps namely inserting a slicer or creating groups.*



8. If you need to reorder data in the PivotTable, you can do this very simply. To reorder the AGE\_GRP data:
  - a. Ensure that AGE\_GRP is in the **Row Labels** area PivotTable
  - b. Right click on age group 5 to 9

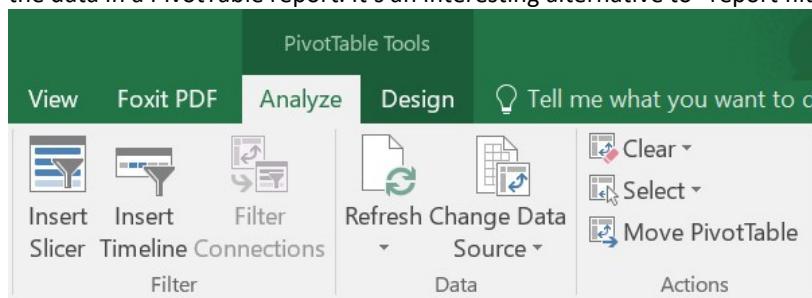
- c. Select **Move > Move “5 to 9” to Beginning** (should now be in proper order and will remain in this order even if you remove it from PivotTable and re-enter)
  
- 9. You need to add a field to a given variable in a pivot table. In the above table there are individuals between the ages of 10-14 and therefore the 10-14 age category is not visible in the table. To add the category:
  - a. Select on the of existing category in the pivot table
  - b. In the Analyze tab click on **Fields, Items, & Sets** in the **Calculations** menu and select **Calculated Item** in the drop box.
  - c. In the pop up box select the AGE\_GRP variable from the fields list on the left hand side. In box titled name type 10-14 and then select **Add** and **OK**.



- d. You can reorder the categories using **Move** option describe above.

*Tip! You can also filter the data by clicking on the arrow beside the variable name within the PivotTable. By clicking on the arrows beside your variables you can add/remove options (e.g. choose to see results only for confirmed cases, only males etc.).*

**Slicer** was a new feature added to MS Excel 2010. Slicers are easy-to-use filtering components that enable you to quickly filter the data in a PivotTable report. It's an interesting alternative to "report filter" within the PivotTable.



1. Select **Inset Slicer** in the **Filter** menu of the **Analyze** tab.
2. In the pop-up box, select the variable you want to filter. The example below show the use of the Slicer feature to quickly look at the age and sex distribution based on case status.

The image shows a PivotTable on the left and a list of case statuses on the right. The PivotTable has 'Count of ID' as the value field, 'Column Labels' set to 'F', and 'Row Labels' set to 'AGE'. The data shows age groups from 5-9 to 65-69, with counts for males (M) and females (F). The total count is 29. The list of case statuses includes Confirmed, Probable, and Suspect.

	Count of ID	Column Labels	M	Grand Total
Row Labels	F			
5-9		1	1	
10-14		0	0	0
15-19		1	1	2
20-24		2	2	4
25-29		2		2
30-34		2	2	4
35-39		2	3	5
40-44		1	3	4
45-49		4		4
50-54		2		2
65-69		1		1
<b>Grand Total</b>		<b>10</b>	<b>19</b>	<b>29</b>

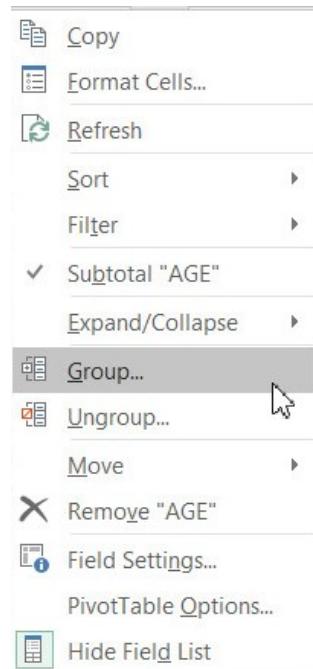
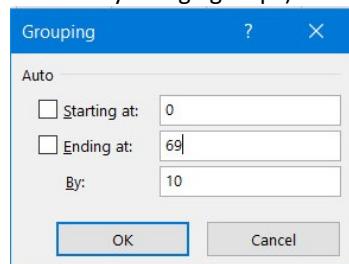
CASE_STATUS
Confirmed
Probable
Suspect

One of the very useful functions of PivotTables is the ability to group data. This is great for grouping individual days into weeks or ages into age categories. When grouping, you need to ensure that there is no missing data in the column being grouped.

The following steps take you through the steps required to group the AGE variable into 10-year age groups.

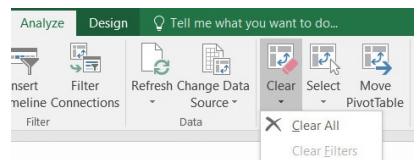
#### Try it: – Use file Excel\_101\_Epis.xlsx/ worksheet PIVOT\_Data.

1. Ensure that the PivotTable has the AGE variable in the **Row Labels**.
2. Right click anywhere in the AGE field.
3. Click **Group**
4. In the dialog box, deselect the two check boxes for **Starting at** and **Ending at**.
5. You may need to adjust the starting date so that it starts at age 0 and ends at age 69.
6. For **By**, enter **10** (to create 10 year age groups). Click **OK**.



7. To ungroup, click on the **Ungroup** button under **Options** or right click on any cell and click **Ungroup**.

*Note: Remember that you can clear the table before continuing with the grouping steps by first clicking anywhere on the PivotTable. Then, click on the **Clear** button of the **Actions** option in the **Analyze** tab. Select the **Clear All** option. This reset step might be helpful once again if you are struggling to group.*



#### Try it: – Use your PivotTable to answer the following questions:

1. How many males and females are included in your linelist?
2. How many confirmed cases?
3. How many males were suspect cases?

- 4. How many cases were reported in December of 2004?**
- 5. How many cases were between the ages of 30 and 39 years? How many were males? Females?**

## Some other tips and tricks for working with PivotTables

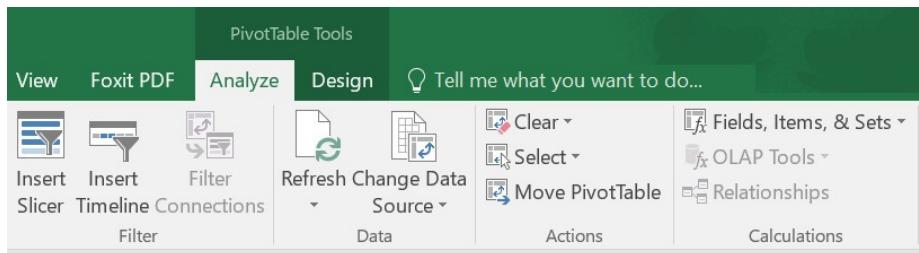
### Creating multiple PivotTables

You can have more than one PivotTable on each Excel worksheet, as well as multiple PivotTables on multiple sheets. To create a second PivotTable on your worksheet, simply select the existing PivotTable, copy and paste it into the worksheet.

### Refreshing your data

When you make changes to your master linelist, you will need to refresh each of the PivotTables. Be careful, if you don't refresh, your PivotTable will still be summarizing old data! To practice updating and refreshing, follow these steps:

1. Update your master linelist with the following changes (*make changes to a copy of your original linelist*):
  - a. ID 22 was actually a female.
  - b. ID 47 became a confirmed case.
  - c. ID 18 was actually a 14-year old male.
2. Refresh your PivotTable
  - a. Make sure that you have saved your updated linelist.
  - b. Go to PivotTable toolbar under **Analyze** and click on **Refresh** in the **Data** menu or right-click data and select **Refresh**.



**Try it: – Use your refreshed PivotTable to answer the following questions:**

1. How many males and females are now included in your linelist?
2. How many confirmed cases?
3. How many cases were between the ages of 10 and 14 years? How many were males? Females?

### Adding rows or columns of data to your linelist

If you are working from a table, any row added to your linelist will be integrated in the Pivot Table. If you are not working from a table, you will need to expand the reference area for the PivotTable. This is why it is strongly suggested to always start from a table.

*Note: the refresh button will only work for any changes made to data within your reference area.*

To expand the reference area for your PivotTable:

1. Go to the **Analyze** tab and select **Change data source**.
2. You will return to your linelist.
3. Highlight the area including the new row and click **OK**.

### Generate a linelist of specific cases

Let's say you would like to see a linelist of only Confirmed cases:

1. Go to the PivotTable and drag the CASE\_STATUS variable in the **Column** area of the **Fields List**
2. Double click on the **Grand Total** data cell containing the count of Confirmed cases.
3. A linelist of the specific cases will be generated in a new worksheet.

## Creating Charts with PivotTables

1. Quick method
  - a. Click on any cell in the PivotTable on which you wish to base the chart.
  - b. Under the **Analyze** tab, select **PivotChart** in the **Tools** menu.
  - c. Select the chart type you want and click **OK**.
  - d. Continue formatting chart as previous described (pages 24).
2. Alternate method
  - a. Copy the data you want to graph from your PivotTable.
  - b. Paste special (values and numbers) into a new excel worksheet.
  - c. Decide what to do with unknowns (typically exclude and add a note at the bottom of table).
  - d. Select your data and click on graph function - do not include totals in your data selected.

**Try it:** – Use file [Excel\\_101\\_Epis.xlsx/ worksheet PIVOT\\_Data](#) to create a simple bar chart of cases by AGE\_GRP.

## Creating Epidemic Curves with Pivot Tables

For this exercise you will need to use the following file:

[Excel\\_101\\_Epis.xlsx/ worksheet EpiCurve\\_data](#)

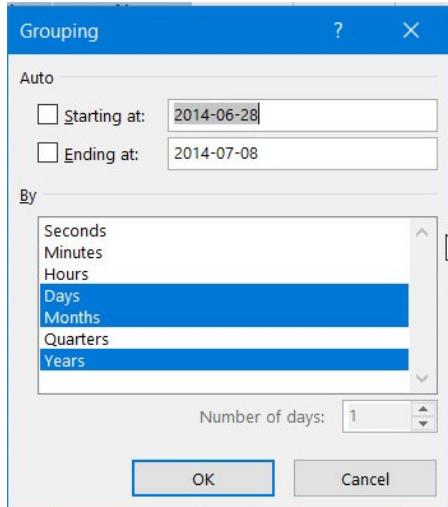
### Step 1: Create a Pivot Table containing the data you want in your epi-curve

1. Highlight all the data in the EpiCurve\_data worksheet and select **Tables** in the **Tables** menu of the **Insert** tab.
2. Select any cell in your table and click on the **PivotTable** button in the **Tables** menu under the **Insert** tab
3. Double check that your **Table/Range** of data is correct
4. Select **New Worksheet** (or **Existing Worksheet** if you are already working in a PivotTable). Click **OK**.

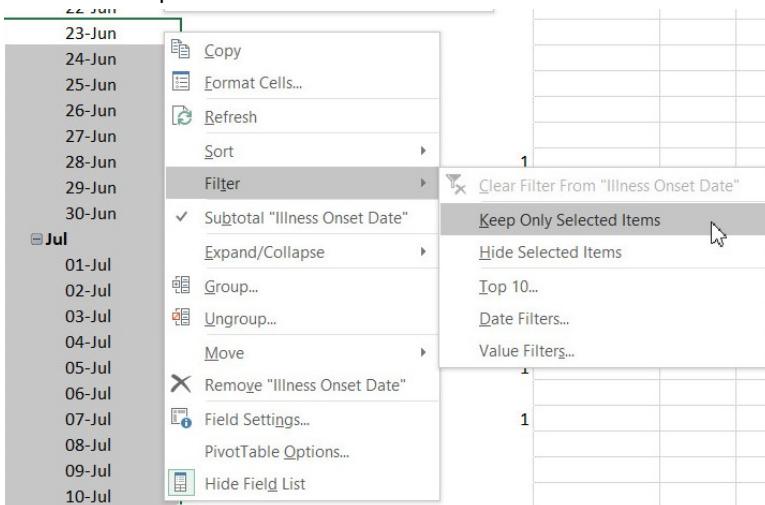
### Step 2: Regroup the date

1. For this exercise, drag Illness Onset Date into **Rows area** and drag National ID into the **Values** field. Now drag P/T into the **Columns** area
2. Ensure that the checkbox **Show items with not data** is checked. To do this, right click on any of the data in the PivotTable and select **Field Settings**. In the Layout & Print tab, check **Show items with no data**.

3. In order to show the dates between cases right-click on any of the dates in the PivotTable, and select **Group**
  
4. In the grouping window, unclick all selected options (highlighted in navy blue) and choose **Days** and set the **Number of days: to 1**. Next, click the **Months** and then **Years**. Click **OK**.

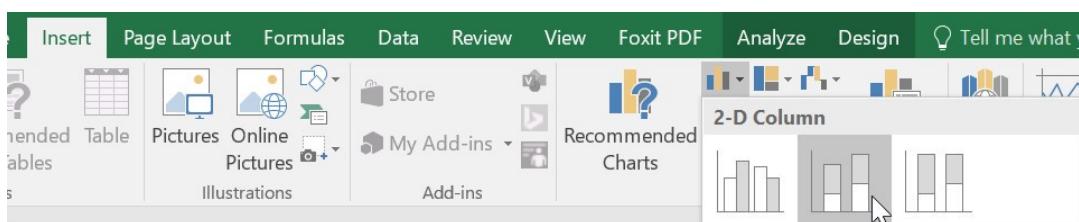


5. You will notice that all dates from January onwards are showing up. These can be hidden so that they do not show up in the Epi Curve. Scroll down to the dates where there are cases and highlight the area with data (in addition, a few days before and after the first and last illness onset dates can be highlighted). Since cases fall between June 28 and July 5, highlight the rows in the pivot table from June 23 and July 10. Right click the highlighted area and select **Filter > Keep Only Selected Items**. Scroll up to see the data.



### Step 3: Create the chart

1. Highlight the PivotTable. Select the **Insert** tab on the ribbon. In the **Charts** section, select **Column** and then **Stacked Column** (second option in the 2-D Column section). A graph should appear. Resize as needed.



2. To remove the spaces between the bars, right click on any bar on the graph. Select **Format Data Series**. Under the **Series Options** and the **Gap Width** section, drag the slider to the very left to the **No Gap** side and click Close (alternatively, in the box below the Gap Width section, delete 150% and enter 0%).
3. In the **Analyze** tab of the Pivot Chart Tools ribbon click **Field Buttons > Hide All**.
4. To remove the **2014** or **month** titles from the bottom of the epi-curve drag the **Years** and **Month** variable out of the **Axis Fields** box in the **Pivot Table Field List**.

## Formatting Guidelines

Page 24 of this document explains how to format charts in excel. The same steps can be used to format epi-curves.

Suggested formatting styles and guidelines for epi-curves are listed below.

- Title
  - Use consistent capitalization from figure to figure
  - Italicize Latin names
  - If a title is more than a single line of text use line breaks
- Axes
  - Use horizontal alignment for x- and y-axis title and labels
  - Select appropriate scales
  - Choose a sensible periodicity for x-axis labels
  - Use actual dates on the x-axis not just numbered weeks
  - Centre x-axis labels with histogram bars
  - Use a consistent font type for all labels, no bolded labels.
- Graph Area
  - Do not include a top border line on around the graph area
  - Do not include gridlines or coloured backgrounds
  - Do not include space between bars
  - Use stacked bars when using more than one category of cases. Ensure colour scheme/ patterns are distinct when printed in black and white.
  - Outline bars with a white or black line so they look more defined
- Legend
  - Place within the graph area
  - The ordering of categories within the legend should be identical to the order used in the curve

A formatted version of the epidemic curve can be seen below.

Figure 1. Number of cases of *E. coli* O157 in Canada, June 23 to July 10, 2014

