Unit 3

EXCEL MAC 2016 for Epidemiology

"Technical skills, like fire, can be an admirable servant and a dangerous master."

- A. Bradford Hill (1971)

Microsoft Excel, MS Excel, is the standard program for creating spreadsheets, maintaining them, and producing charts. Other programs are available, such as Quattro Pro or Lotus 1-2-3, but you are unlikely to encounter them.

This introduction to MS Excel focuses on its use for data set creation, manipulation (eg- sorting and selecting) and selected calculations.

While it can be done, we will not be using MS Excel to do statistical analyses and data visualizations. The focus of BIOSTATS 690C is, instead, on R and Stata.

Table of Contents

Learning Objectives

When you have finished this unit, you should be able to:

- Navigate in and out of Excel (launch, exit, enter data, format cells, arrange columns, freeze rows and columns for easy viewing, sort, and autofill)**;**
- Specify the format and layout of an excel spreadsheet for printing (portrait v landscape, headers, footers, etc);
- Create new fields (what we think of as variables) using functions and user-specified formulae; and
- Create a data set and document it.

1. Introduction to MS Excel

1.1 What is MS Excel?

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

What is a spreadsheet? *It is simply a grid of information that might include numbers or words or a mix. Its storage in a grid is handy way to do its organization.*

What might I like to do with a spreadsheet?
Lots of things, actually – lists, sorted lists, picture summaries. Also charts

1.2 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.
- 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 data display (e.g. – number of significant digits, display of dates as month-day-year) with no loss of information.

Disadvantages

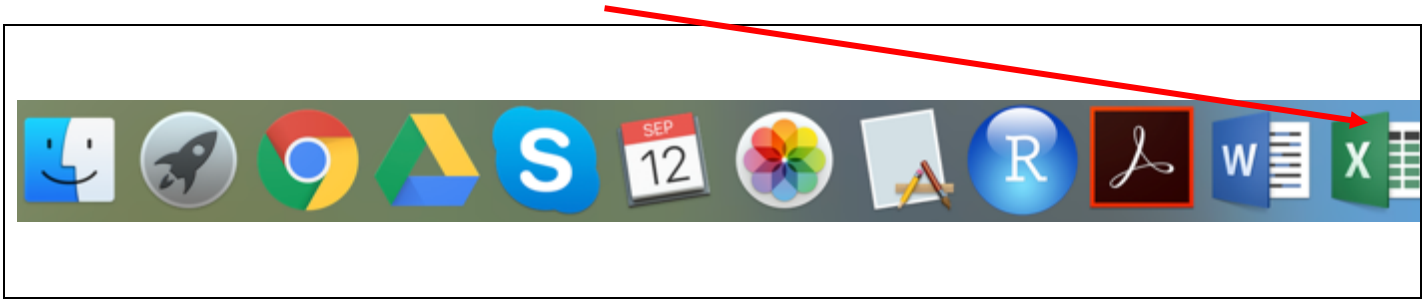
- The entry of a negative number is awkward. Excel might interpret the negative sign as the beginning of a mathematical formula. Solution: enter the negative number with a leading **apostrophe ‘** in the cell, e.g., ‘-0.28.

- Excel can make mistakes in mathematical formulae if you inadvertently mix character 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.

- Sometimes, the charts produced by Excel are not correct.

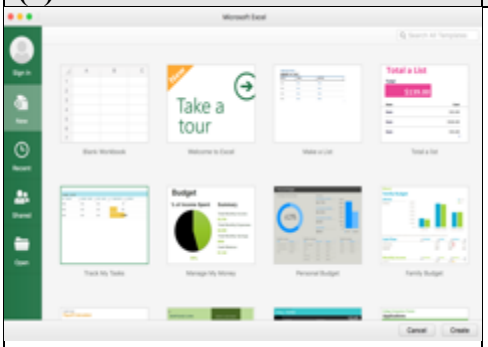
2. Getting Started - Spreadsheet Basics

Launch Excel by clicking on the excel icon on your dock.

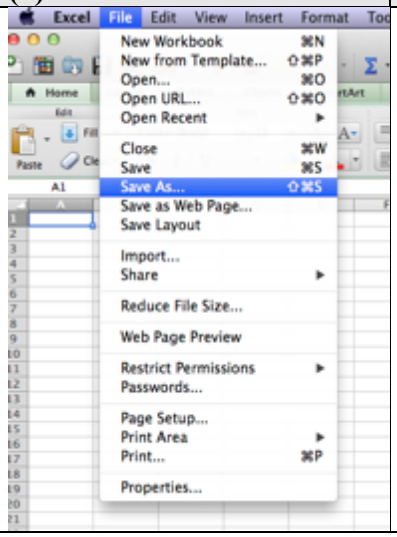


Tip! **Begin your session by saving your file:**
(1) Upon launching, Excel 2016 offers you a gallery of templates. Select "Blank Workbook" and click "Create" at the bottom left. (2) From the menu bar at the top of your screen**: FILE > SAVE AS.** (3) In the "Save As" box, enter a name of your choosing; no need to type the extension ".xlsx" Excel provides it. Click Save. I chose to name my saved file ***carol_practice.xlsx.***

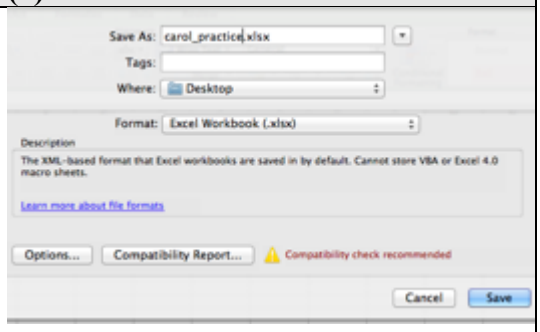
(1)



(2)

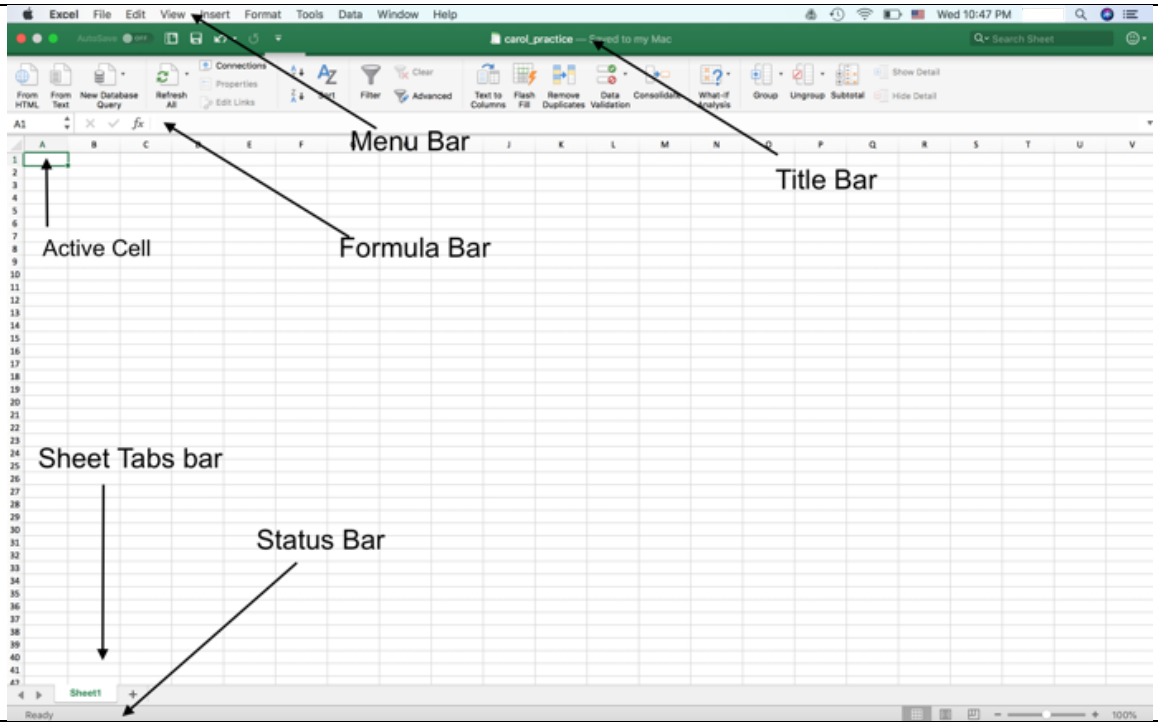


(3)



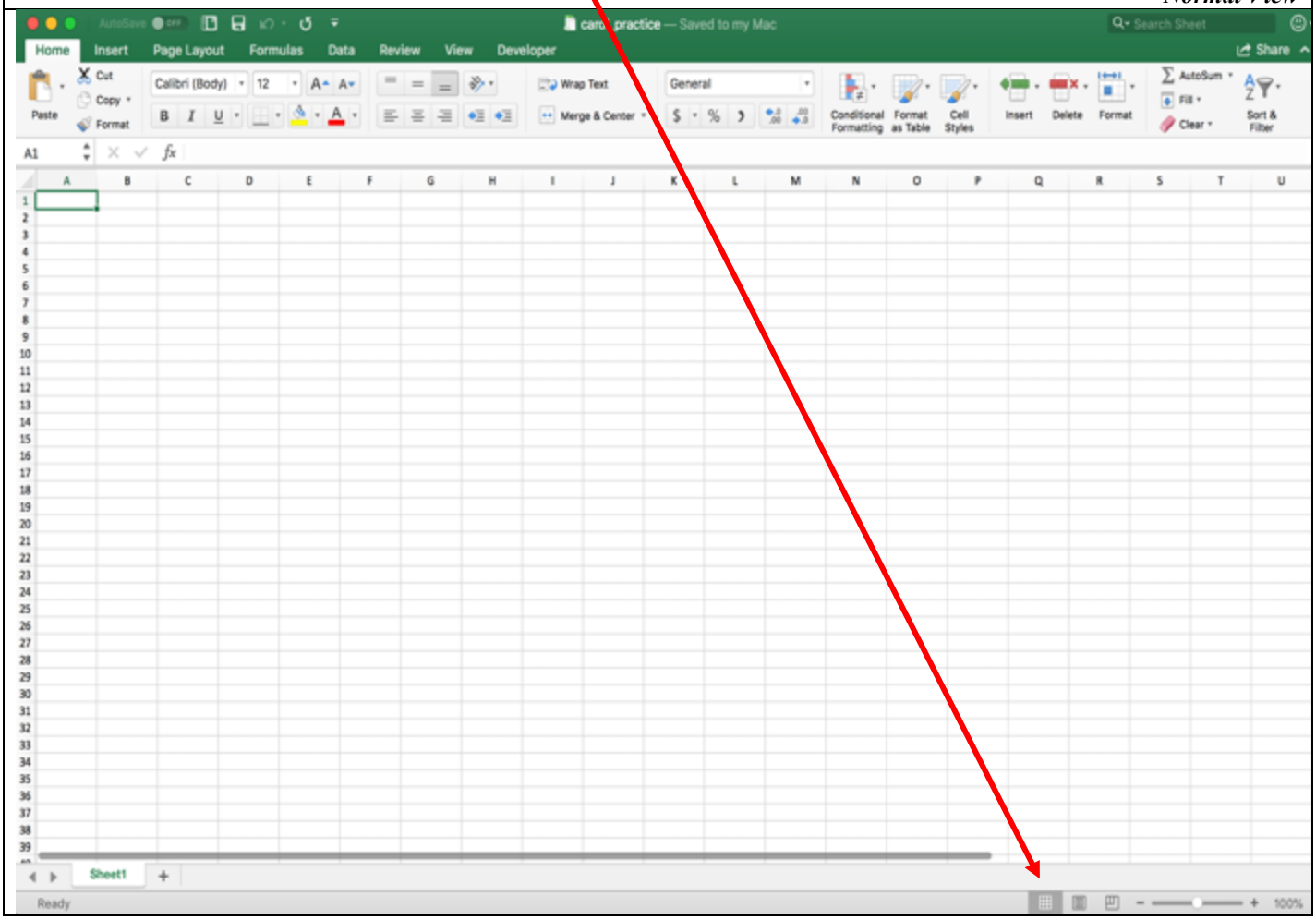
The Excel Window - Basics

Note – Your screen may look different depending on what toolbars, ribbons, and icons have been selected for display.

Excel 2016 for MAC


Excel offers three "views": page layout, normal, and page break review. Choose the view you like using the icons located at the bottom.

Normal View



The Excel Window – Detailed Key.

Again … Your screen may look different depending on what toolbars, ribbons, and icons have been selected for display.

Excel Mac Menu Bar

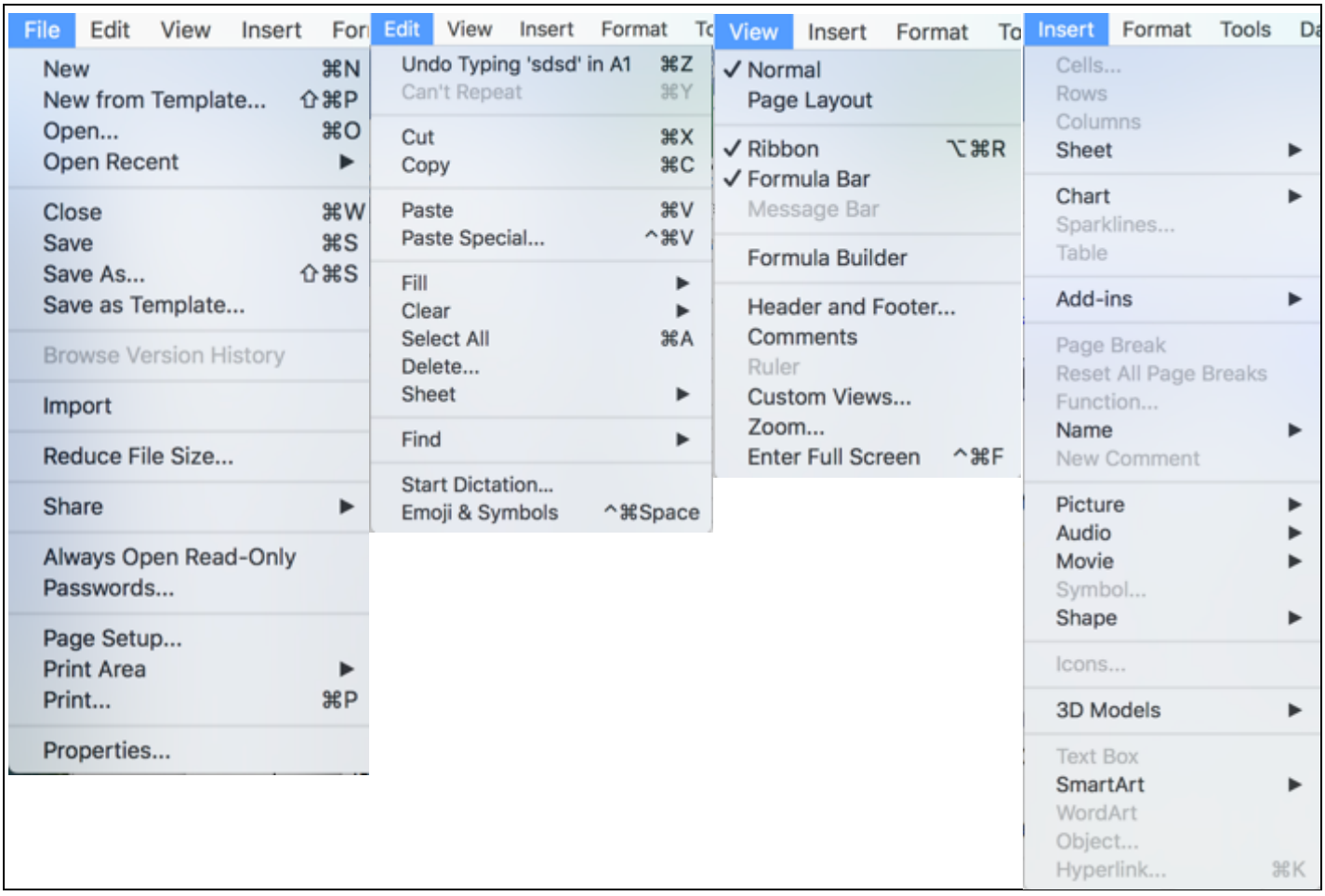
Quick Access Toolbar	Common commands such as “Save” and “Undo”
Ribbon	A series of tabs that provide access to tool groups
Formula Bar	Shows the contents of cells and is used to compose formulas.

2.1 Excel 2016 for Mac Toolbars

In Excel 2016 for MAC you have 3 choices of toolbars: 1) Excel Mac menu bar 2) Quick Access toolbar or 3) Ribbon.

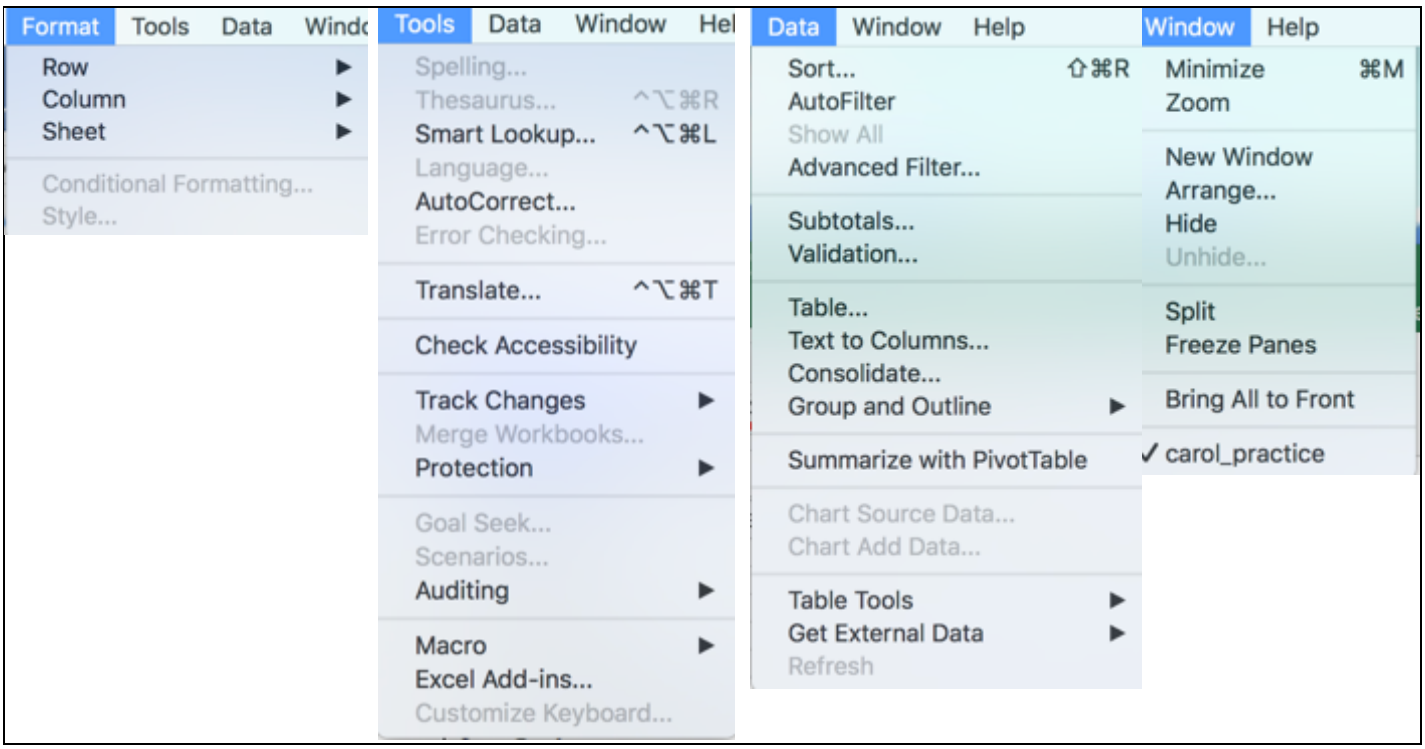
1) Excel Mac Menu Bar

FILE, EDIT, VIEW, and INSERT drop down menus:



Design ..... Data Collection ..... Data Management ..... Data Summarization ..... Statistical Analysis ..... Reporting

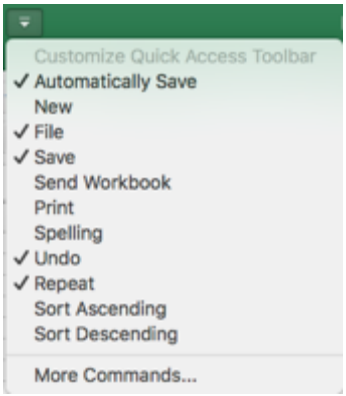
FORMAT, TOOLS, DATA, and Window drop down menus:



Design ..... Data Collection ..... Data Management ..... Data Summarization ..... Statistical Analysis ..... Reporting

2) Quick Access Toolbar

The quick access toolbar contains common commands. Hover your mouse (from left to right) to locate the icons for, in order: **start a new worksheet from a template, save, undo, refresh, and customize access toolbar.**



More commands can be added to the quick access toolbar by clicking the last icon and check any commands listed in the dropdown or select "More Commands" for more options.

Note: the Toolbox icon is dropped from Excel 2016

- **Formula builder** can be found under "View" in the Excel menu bar
- **Reference tools** can be found under "Tools" in the Excel menu bar

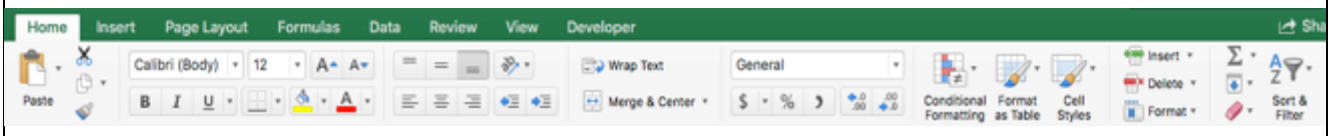
3) Ribbon

A green ribbon is located below the tool bar. Notice 8 headings or "tabs." "tab" contains groups of functions that have been organized for easy use.

At the far right, notice the down arrow next to "Share": the arrow is for showing or hiding functions

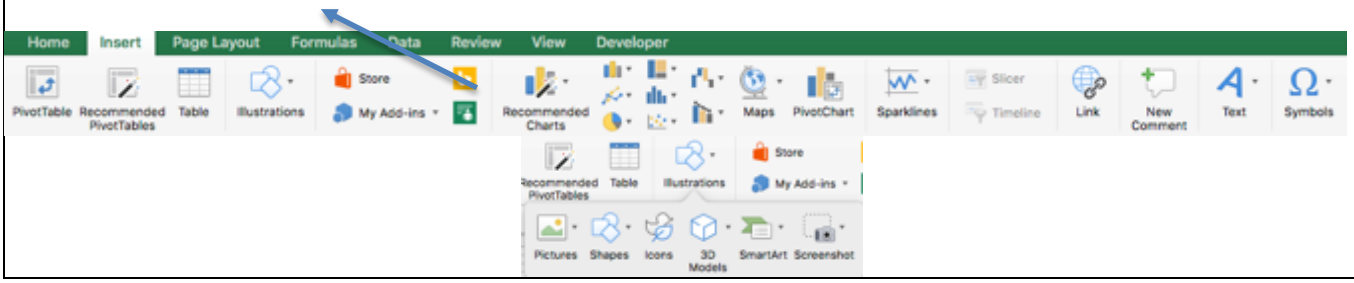
Home tab

Most frequently used. Contains many familiar commands e.g. – formatting cells, text and numbers, inserting and deleting columns and rows, and adding formula (near Sort & Filter icon on the right).



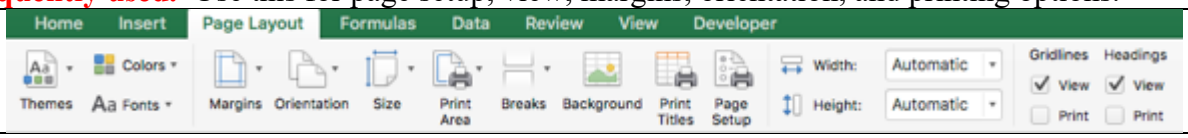
Insert tab

Frequently used. Use to format and edit tables created within Excel. Here for pivot tables, too. Also use to create graphs and charts. Pictures and SmartArt can be added under the Illustrations dropdown



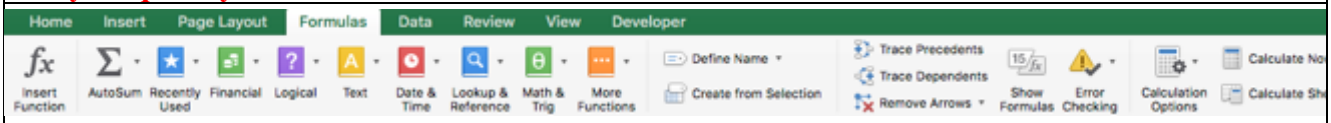
Page Layout tab

Frequently used. Use this for page setup, view, margins, orientation, and printing options.



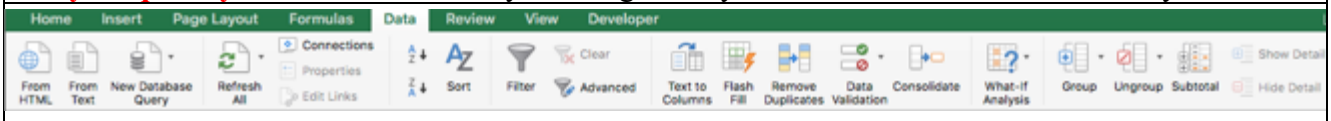
Formulas tab

Very frequently used. Use to create or build formulas and functions



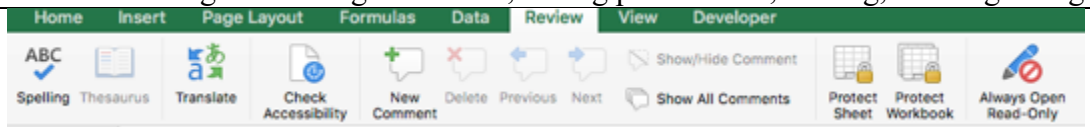
Data tab

Very frequently used. Use to actually do things with your data such as: sort, filter, analyze.



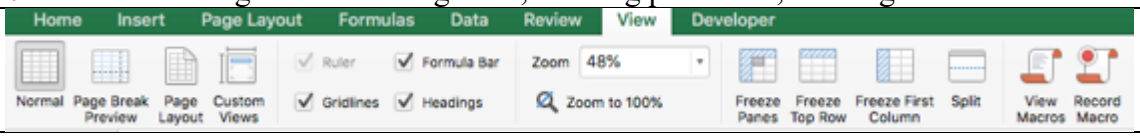
Review tab

Useful. Use for creating and editing comments, setting permissions, sharing, tracking changes, etc.



View tab

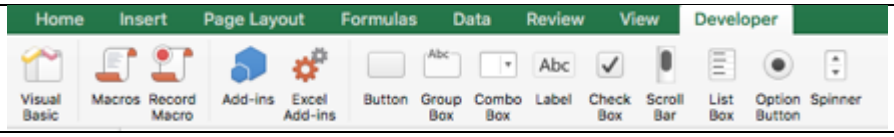
Useful. Use for selecting or customizing view, freezing panes and, viewing macros.



Design ..... **Data Collection** ..... **Data Management** ..... **Data Summarization** ..... **Statistical Analysis** ..... **Reporting**

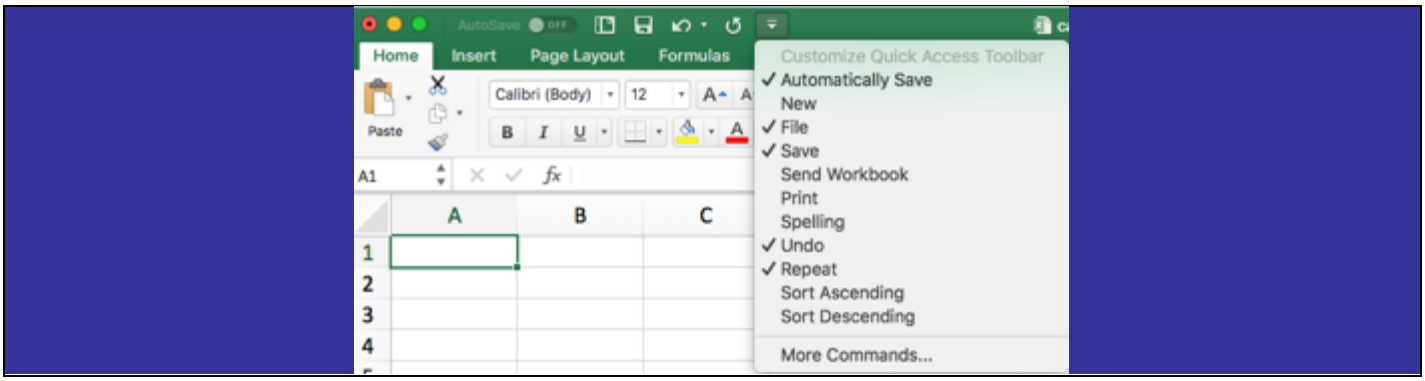
Developer tab

Useful. Use for adding Visual Basic for Applications (VBA) to the workbook and managing add-ins.

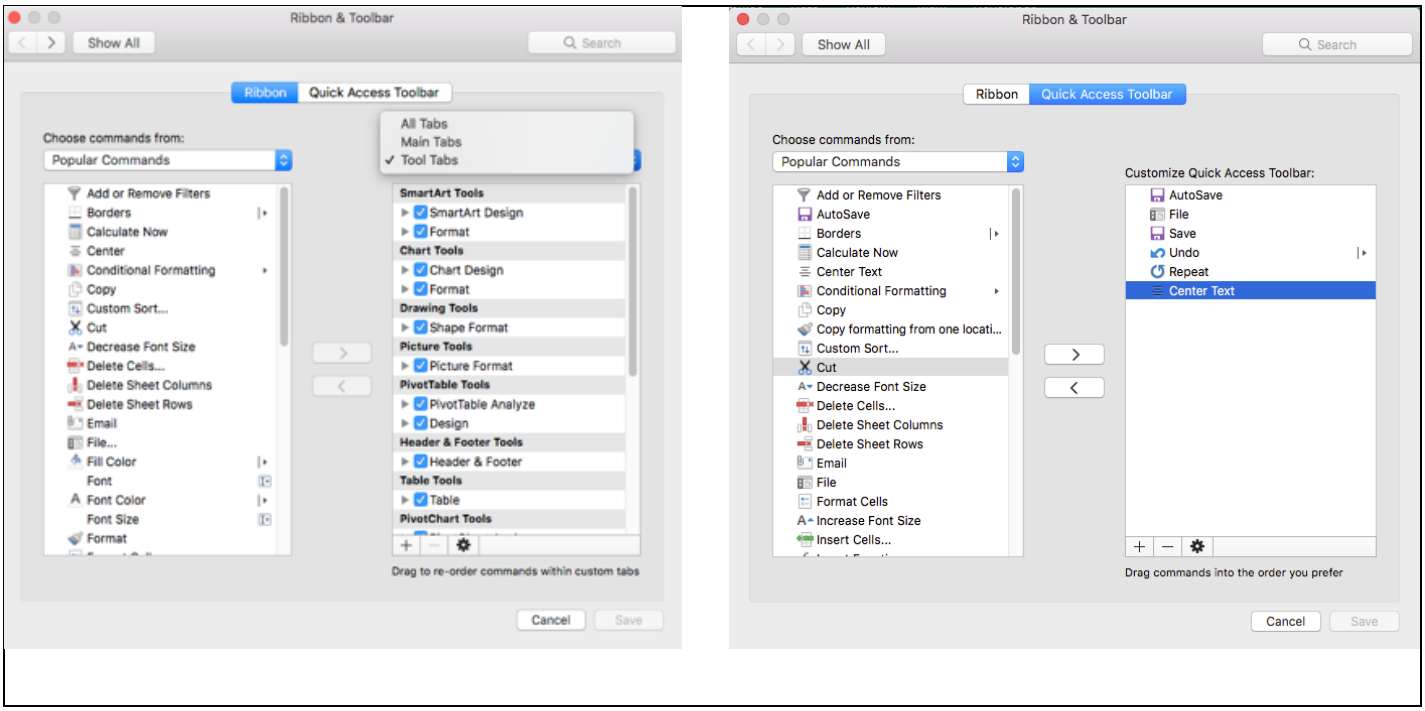


Tip! **Customize the display of toolbars.**

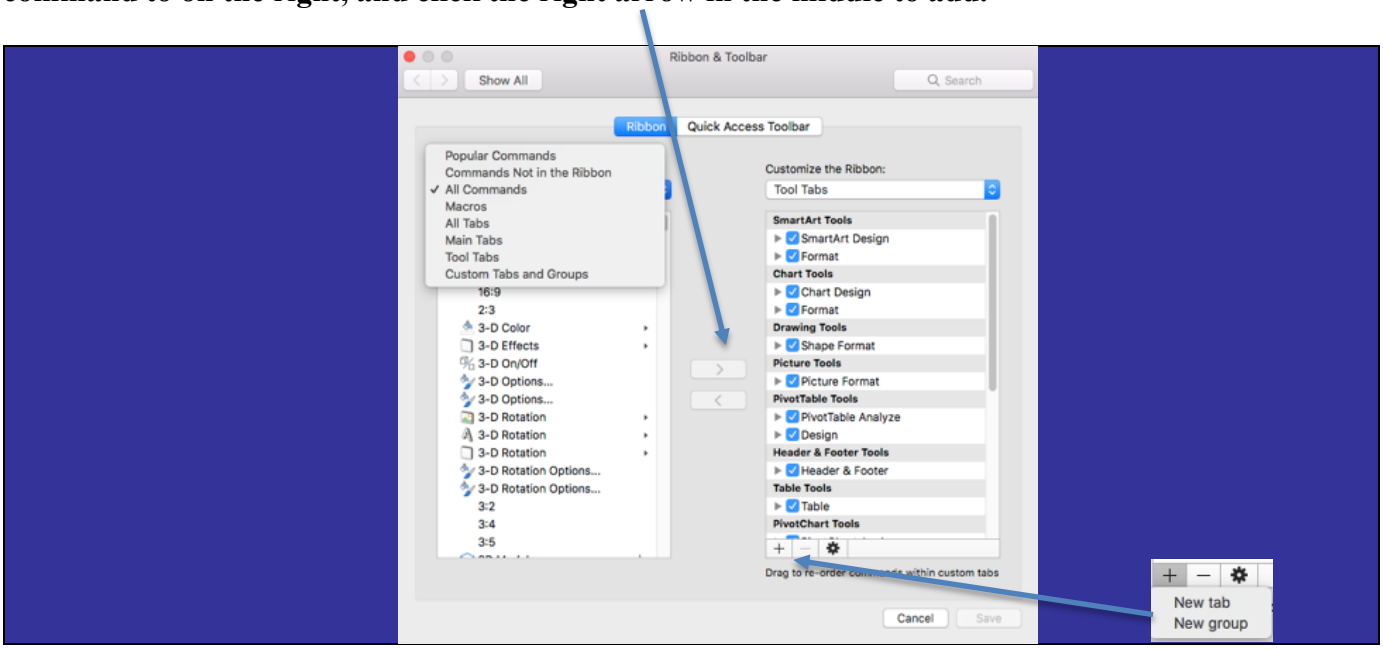
From the Quick Access Toolbar: Down arrow icon > MORE COMMANDS



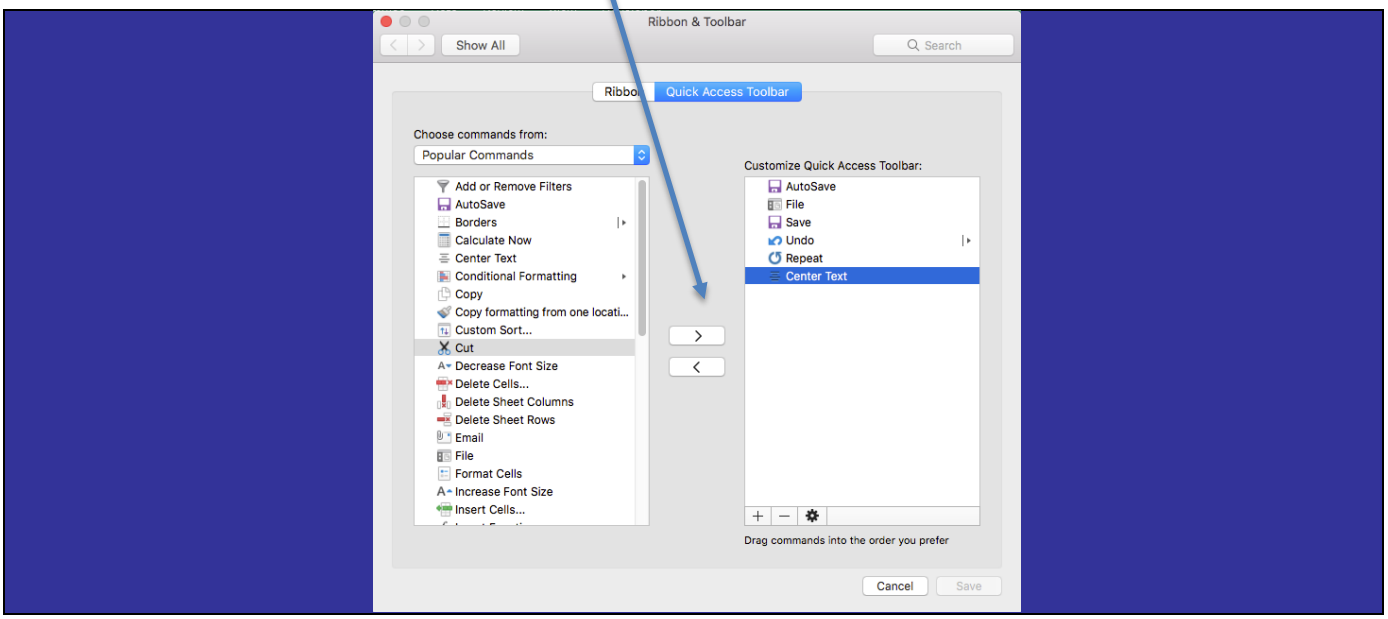
Click on RIBBON or QUICK ACCESS TOOLBAR.



Customize RIBBON by 1) Select a tool under a tool group (e.g. SmartArt Design under SmartArt Tools) on the left and click the add sign at the bottom to create "New Group" 2) Scroll down to find short cuts to the icons you want on the left, and select. 3) Select a tool group that you want to add a command to on the right, and click the right arrow in the middle to add.



Customize QUICK ACCESS TOOLBAR by 1) Scroll down to find short cuts to the icons you want on the left, and select. 2) Click the right arrow in the middle to add.

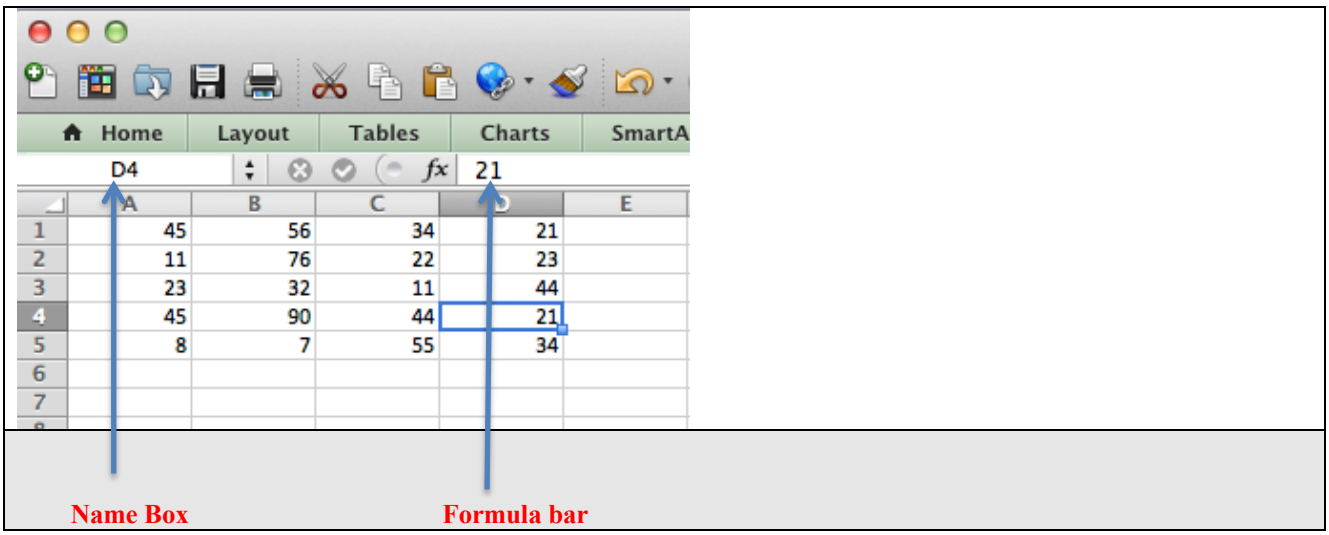


2.2 How to move through cells

Moving through cells in Excel 2016 for Mac is mostly (but not entirely) the same as in versions of Excel for the PC. One distinction is in moving to a particular cell. See the last row of the following table.

To Get to ….	Key strokes to use are …
One cell up	up arrow key
One cell down	down arrow key or **ENTER**
One cell left	left arrow key
One cell right	right arrow key or **TAB**
Top of the worksheet (cell A1)	**CTRL+HOME**
End of the worksheet (last cell containing data)	**CTRL+END**
End of the row	**CTRL**+right arrow key
End of the column	**CTRL**+down arrow key
Any cell **Example: Cell D4 (column "D", row "4")**	**Step 1:** Click in the name box located to the left of the formula bar (picture below) **Step 2:** Type the cell reference, eg; d4 **Step 3**: Press enter

Location of the "name box" – Left of the formula bar. In this picture, the "name box" as in it D4



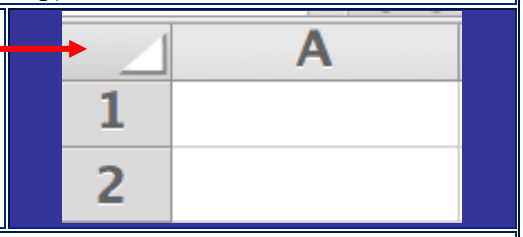
2.3 How to Modify the Layout of Your Worksheet

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: 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. Left-click and drag to obtain the row size you want and then release.
  - To resize a column: 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. Left click and drag to obtain t he column size you want. Release.

- **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** (triangle shape located in upper left, above row 1 and to the left of column "A") 
Cluster of cells	drag mouse over the cells or hold down the **SHIFT** key while using the arrow keys

- **Insert/Delete a row**

  Your new row will be ABOVE your current location
  To Insert: **INSERT > ROW**
  To Delete: **EDIT > DELETE.** Choose: **"entire row"**

- **Insert/Delete a column**

  Your new column will be LEFT of your current location
  To Insert: **INSERT > COLUMN**
  To Delete: **EDIT > DELETE** Choose: **"entire column"**

- **Move or copy cells**

  Highlight and select the cells you want to move or copy

  To move cells: **EDIT > CUT**
  Position cursor in upper left cell of destination
  EDIT > PASTE

  To copy cells: **EDIT > COPY**
  Position cursor in upper left cell of destination
  EDIT > PASTE

- **Freeze panes**

  TIP!!! This is a wonderful feature! It allows you to retain for viewing some "header" rows and columns. **Example –** Row 1 might contain variable names and column A might contain study id's. I want to be able to always see these, regardless of where I am in the worksheet.

  To freeze panes: Position cursor (1) ***below*** header row and (2) ***right*** of header column
  VIEW TAB > FREEZE PANES

  To UN freeze panes: Use **VIEW TAB > 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 File > Page Setup > Sheet. More on this later (see section 2.8).***

- **Adding, Deleting, Renaming, and Moving Worksheets**

  At the bottom of your screen you will see a tab for each worksheet: **sheet1, sheet2** and so on.

  To **add** a worksheet: Click on the "+" button located to the right of your last worksheet

  To **delete** a worksheet: Locate yourself anywhere in the worksheet to be deleted.
  EDIT > DELETE SHEET

  To **rename** a worksheet: **Right click** on the tab of the worksheet you want to rename.
  The tab is at the bottom of your screen.
  Choose from menu: **RENAME**

  To **move** a worksheet: **Right click** on the tab of the worksheet you want to move.
  Choose from menu: **MOVE OR COPY**

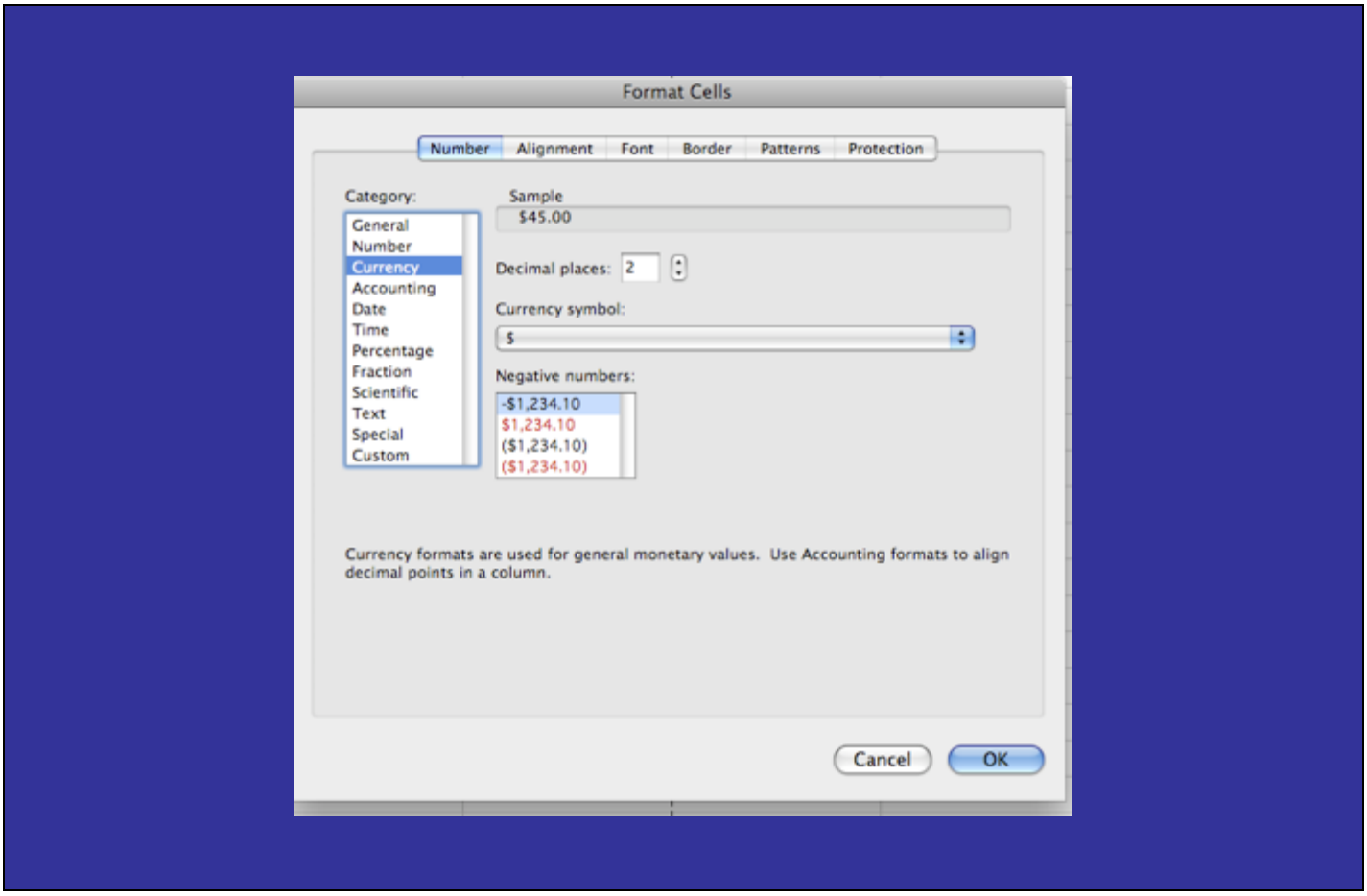
2.4 Formatting Cells

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

(1) Highlight the cells. Typically, you will select an entire column by clicking on the column heading

(2) From the menu bar at top: **FORMAT > CELLS**
This will open the dialog box below. It has several tabs. You will be positioned in the NUMBER tab

(3) Choose the tab and category that you want to format. Then select from the drop down menus that are provided. **Example –** Suppose I want to format this column as US currency with 2 units places for cents. I also want negative dollar amounts to appear in black with a minus sign in front. See below.



(4) The other tabs (eg – "Alignment", "Font" "Border") can be accessed to change the font of text entries, to align entries on the right, left, or center of cells, etc. Try it!

2.5 Formulas and Functions

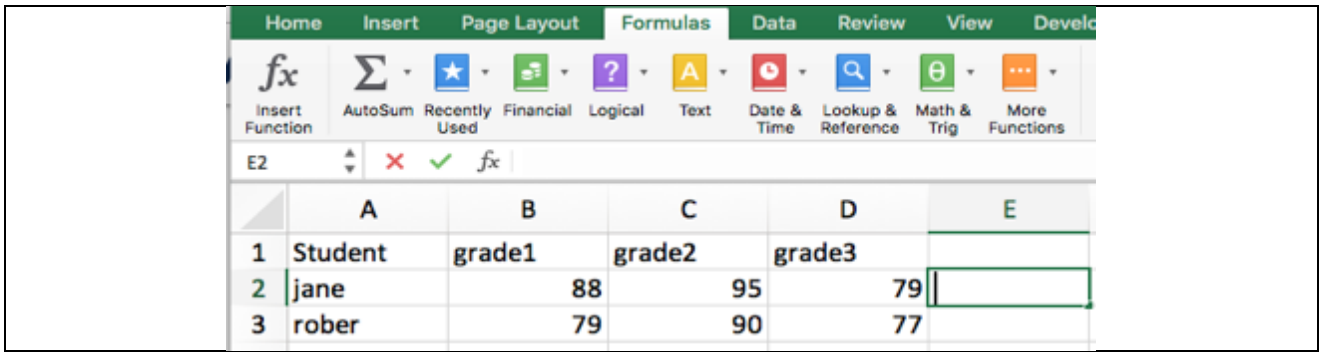
The creation of new variables (or fields) that are the result of calculations is easy. Here is a little data set that I just made up:

	A	B	C	D	E
1	Student	grade1	grade2	grade3	
2	jane	88	95	79	
3	rober	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					

Suppose we want to obtain, separately for each student, the average of "grade1", "grade2", and "grade3."

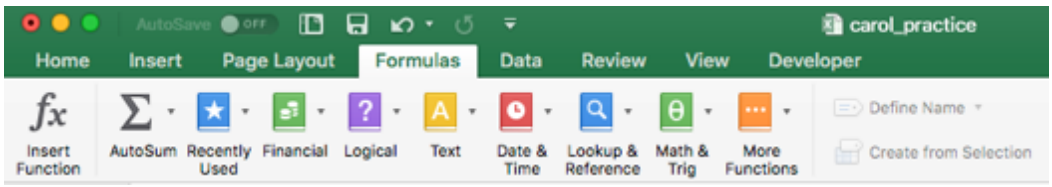
First, we obtain the average for the student whose scores are in the first row. This is "Jane" and her scores are in row "2".

Step 1: Highlight (this makes the cell "active") the cell where the result is to be stored. This is cell E2. Note the reassuring blue bold border. To the left of the formula box, you will see E2. Again, reassuring. The result of your calculation will be put into cell E2.

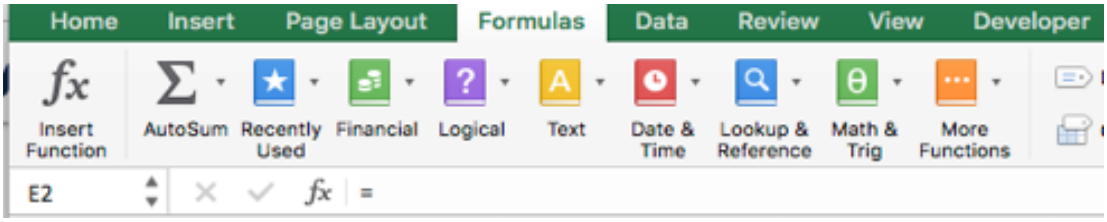


	A	B	C	D	E
1	Student	grade1	grade2	grade3	
2	jane	88	95	79	
3	rober	79	90	77	

Step 2: Click on the **FORMULAS TAB** icon located in the (green) ribbon.
The formula tab will be opened, showing you several icons related to formulae.

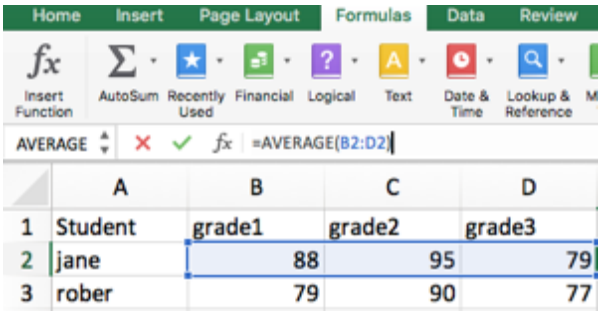


Step 3: **Always begin your calculation with an equal sign, "="**
Position your cursor in the formula box. Type the equal sign.



Step 4: **Make use of the function average and the use of "select" and "drag" to complete calculation**

Type the word **average** followed by a left parenthesis. Take care that there is NOT a space.
Next - highlight cell B2 (Jane's score for grade 1)
Next - Using shift key, select and drag to select all three scores: B2, C2, and D2
Now you can type your right parenthesis.
Press enter.



	A	B	C	D
1	Student	grade1	grade2	grade3
2	jane	88	95	79
3	rober	79	90	77

The result now appears in cell E2:

	A	B	C	D	E	F
1	Student	grade1	grade2	grade3		
2	jane	88	95	79	87.333333	
3	rober	79	90	77		
4	kamil	92	88	83		

Replicate this calculation for all the other students.
There are at least two ways to do this.

Approach 1

(1): Highlight (make active) the cell that has the first result; this is cell E2 in this example.

(2): From the top menu bar, choose **Copy**

(3): Highlight all the destination cells; these will be cells E3, E4, and so on down to the last row in your data set.

(4): From the top menu bar, choose **Paste.**

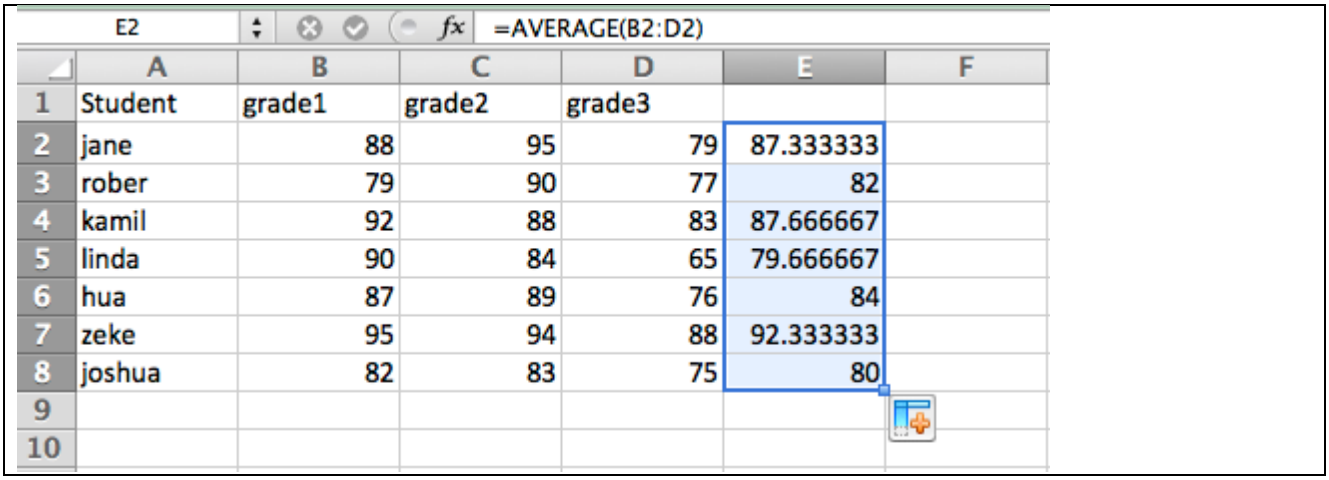
Approach 2

(1): Highlight (make active) the cell that has the first result; this is cell E2 in this example.

(2): Click the bottom right corner of this cell. A little "cross hair" should appear.

(3): Click on the "cross hair" and drag down through E3, E4, etc to the last row in your data set. Release.

Voila!

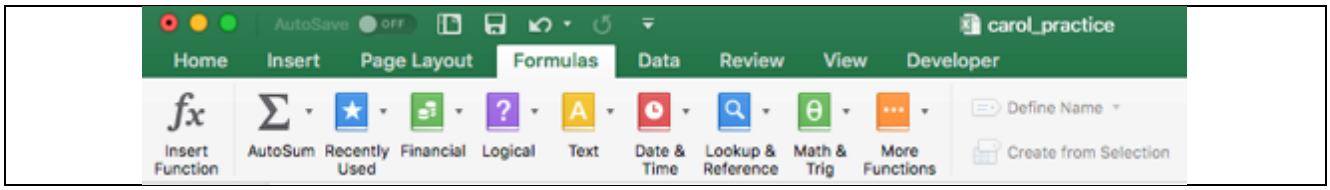


	A	B	C	D	E	F
1	Student	grade1	grade2	grade3		
2	jane	88	95	79	87.333333	
3	rober	79	90	77	82	
4	kamil	92	88	83	87.666667	
5	linda	90	84	65	79.666667	
6	hua	87	89	76	84	
7	zeke	95	94	88	92.333333	
8	joshua	82	83	75	80	
9						
10						

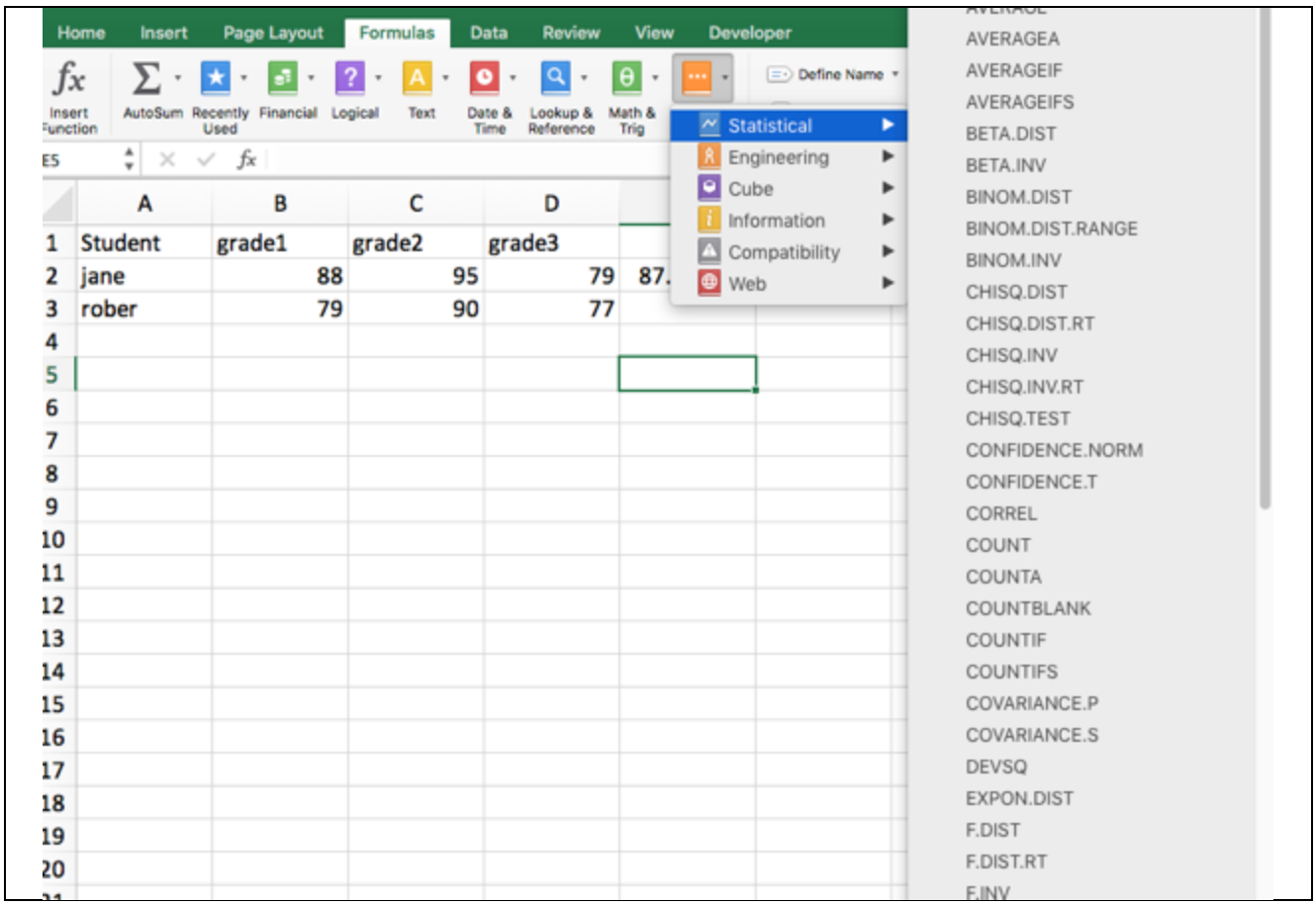
MS Excel has hundreds of "Built in" Functions

Excel 2016 for MAC offers easy access, help, and examples.

Click on the icons in the FORMULAS tab on the ribbon bar:

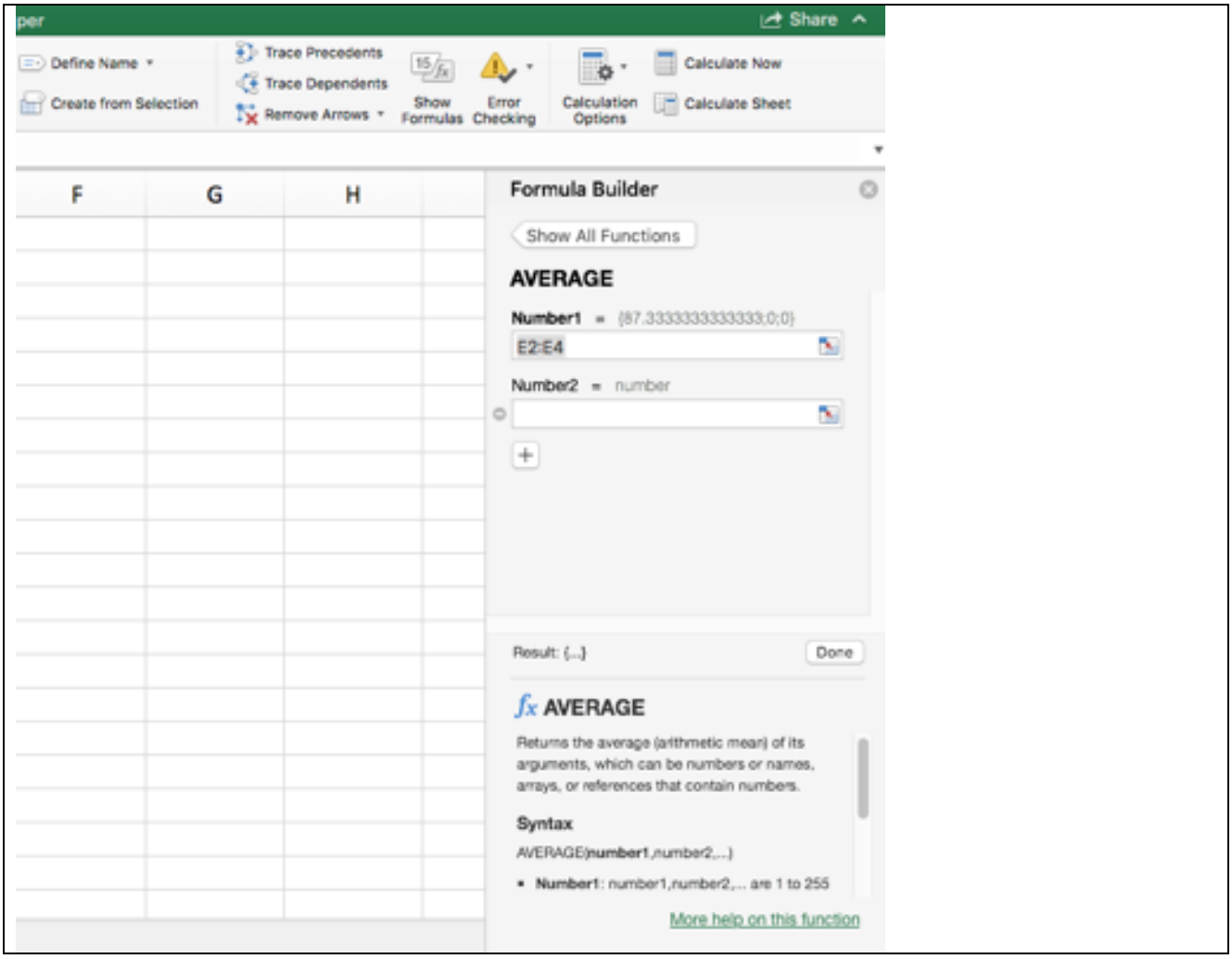


Excel returns a dropdown of functions in that category:

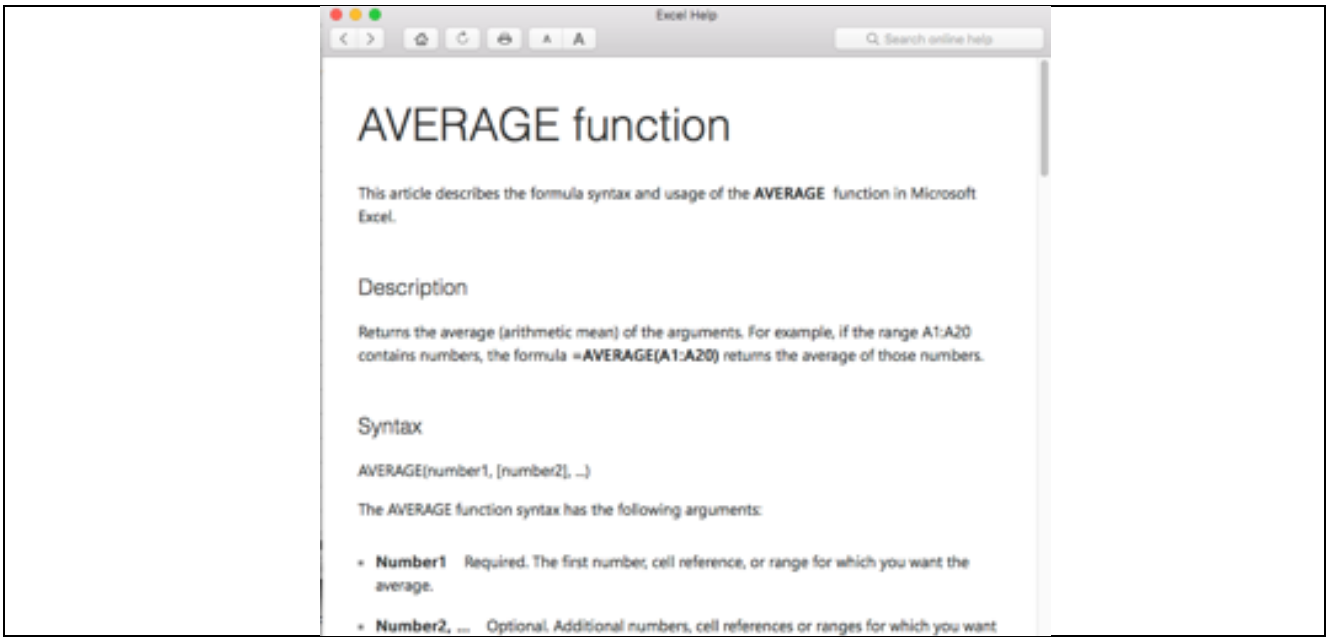


Design	.....	**Data Collection**	.....	**Data Management**	.....	**Data Summarization**	.....	**Statistical Analysis**	.....	**Reporting**

Click on function of interest and the Formula Builder will appear on the right.



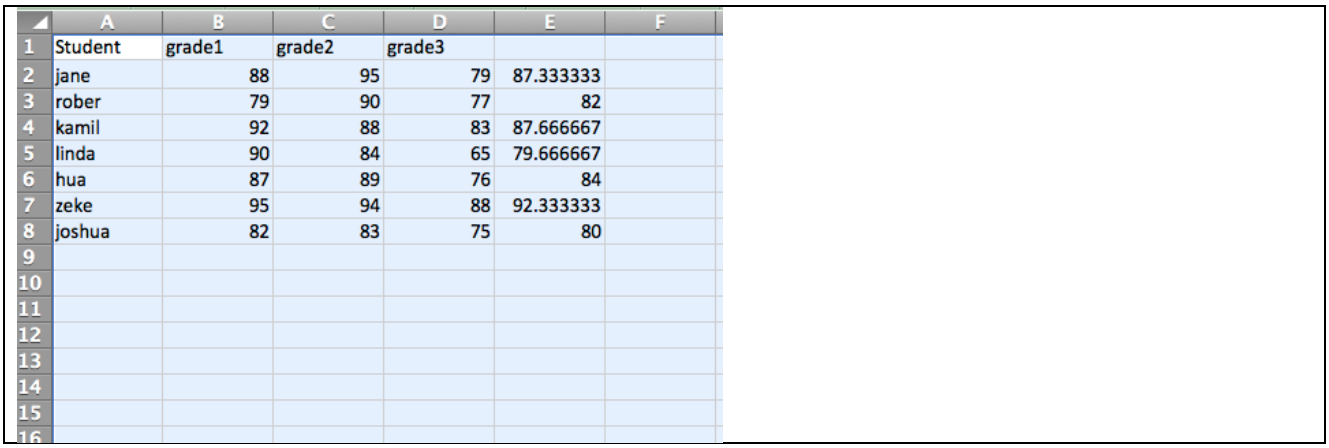
For more information about that function, click “More help on this function” (in green on the bottom right of previous picture) and a window with more detailed information will appear:



2.6 Sorting

Excel lets you sort the data in your spreadsheet by the entry in one column while retaining the integrity of the entire profile for each record.

Preliminary - Before sorting, highlight **ALL** of the cells that are to be sorted; this is usually the entire worksheet (recall – to select the entire worksheet, click on the triangle at upper left). The whole screen will turn light blue.

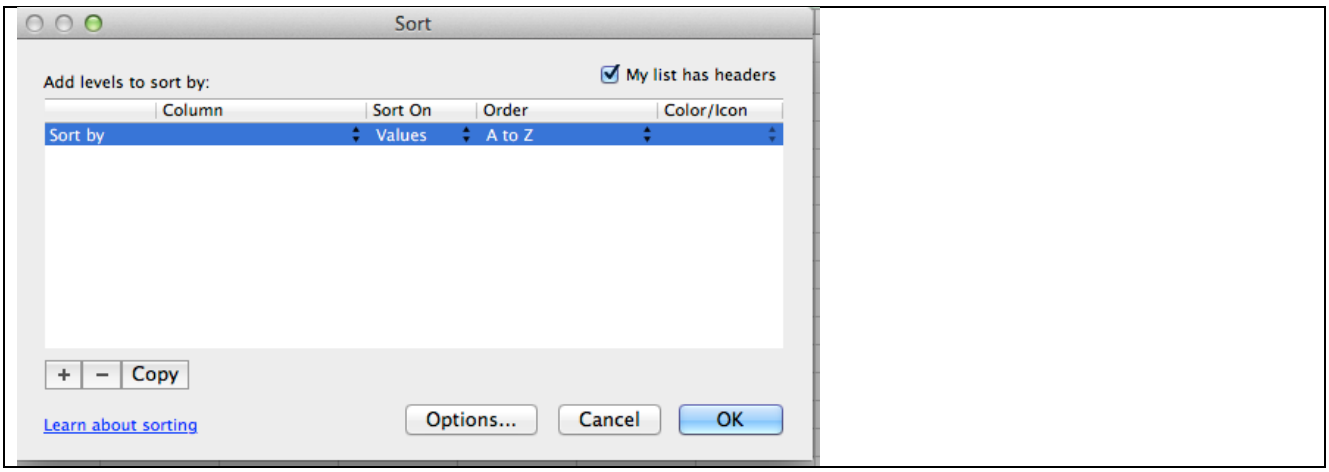


	A	B	C	D	E	F
1	Student	grade1	grade2	grade3		
2	jane	88	95	79	87.333333	
3	rober	79	90	77	82	
4	kamil	92	88	83	87.666667	
5	linda	90	84	65	79.666667	
6	hua	87	89	76	84	
7	zeke	95	94	88	92.333333	
8	joshua	82	83	75	80	
9						
10						
11						
12						
13						
14						
15						
16						

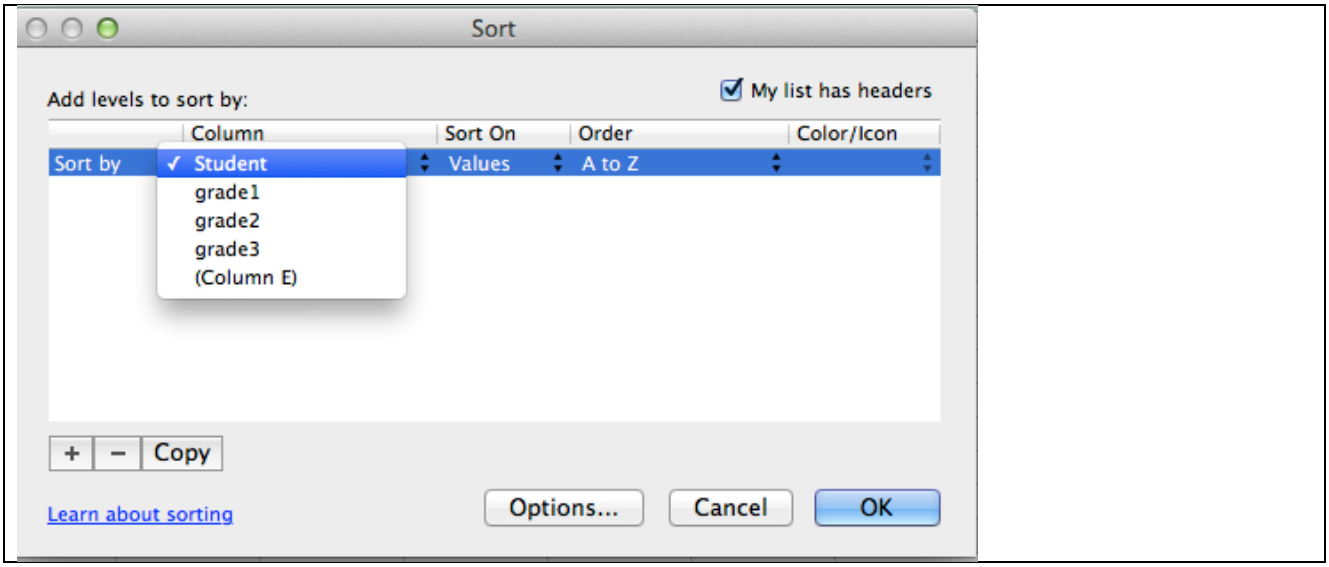
Again, there are at least two ways to do a sort, depending on the toolbar you choose to use.

Approach 1

From the **Excel MAC Menu Bar**: **DATA > SORT**. The following "Sort" window will appear. Check that the box next to "My list has headers" is checked.



Choose your sort. Here, I chose to sort the data by the entries in the column "Student", alphabetically, A to Z. Click OK.

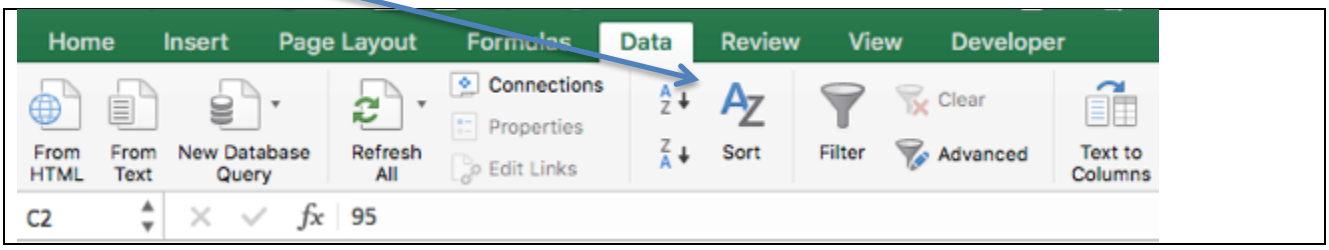


Voila!

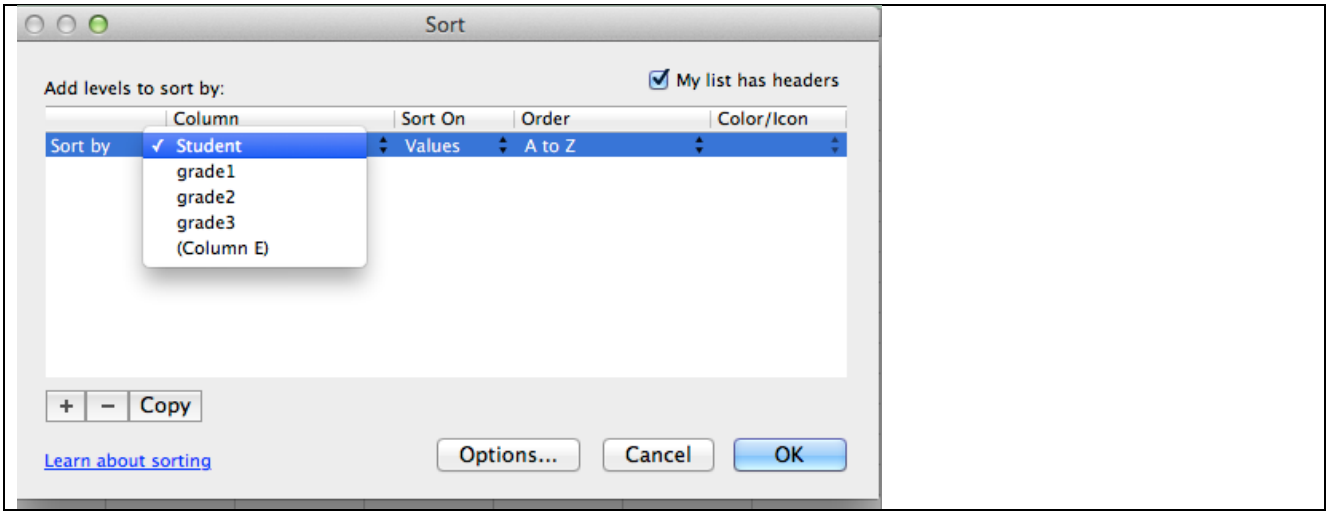
A	B	C	D	E	F
Student	grade1	grade2	grade3		
hua	87	89	76	84	
jane	88	95	79	87.333333	
joshua	82	83	75	80	
kamil	92	88	83	87.666667	
linda	90	84	65	79.666667	
rober	79	90	77	82	
zeke	95	94	88	92.333333	

Approach 2

From the **DATA tab on the ribbon**. The tab is expanded.
Click on the **sort** icon at the middle:



The window below will appear. Choose your sort. Again, I chose to sort the data by the entries in the column "Student", alphabetically, A to Z. Click OK.



Voila!

A	B	C	D	E	F
Student	grade1	grade2	grade3		
hua	87	89	76	84	
jane	88	95	79	87.333333	
joshua	82	83	75	80	
kamil	92	88	83	87.666667	
linda	90	84	65	79.666667	
rober	79	90	77	82	
zeke	95	94	88	92.333333	

2.7 Auto-filling (eg, 1 1 1, etc) and Fill Series (eg 1 ,2 3, etc)

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.

Example Suppose you would like to replicate the A2 cell entry of "2009" into cells A3 through A150.

Step 1: Enter 2009 in cell A2. Press enter. Select the cell A2 again.

Step 2: In the now active cell A2, position your cursor at the lower right corner of this cell so that the cursor arrow changes to a **small black cross**.

Step 3: Click one time on the small black cross. **Without releasing, drag down** A3, A4 and so on to A150.

Step 4: When you release the mouse, notice that all highlighted cells now contain 2009, and a small auto-fill options button appears. This brings us to Fill Series.

Fill Series (eg 0, 5, 10, 15 and so on….)

Excel can also save you time if you need to enter a regular series of numbers, days of the week, etc. Choosing the **Fill Series** option in the example above will result in the series 2009, 2010, 2011, etc., adding 1 to each successive cell.

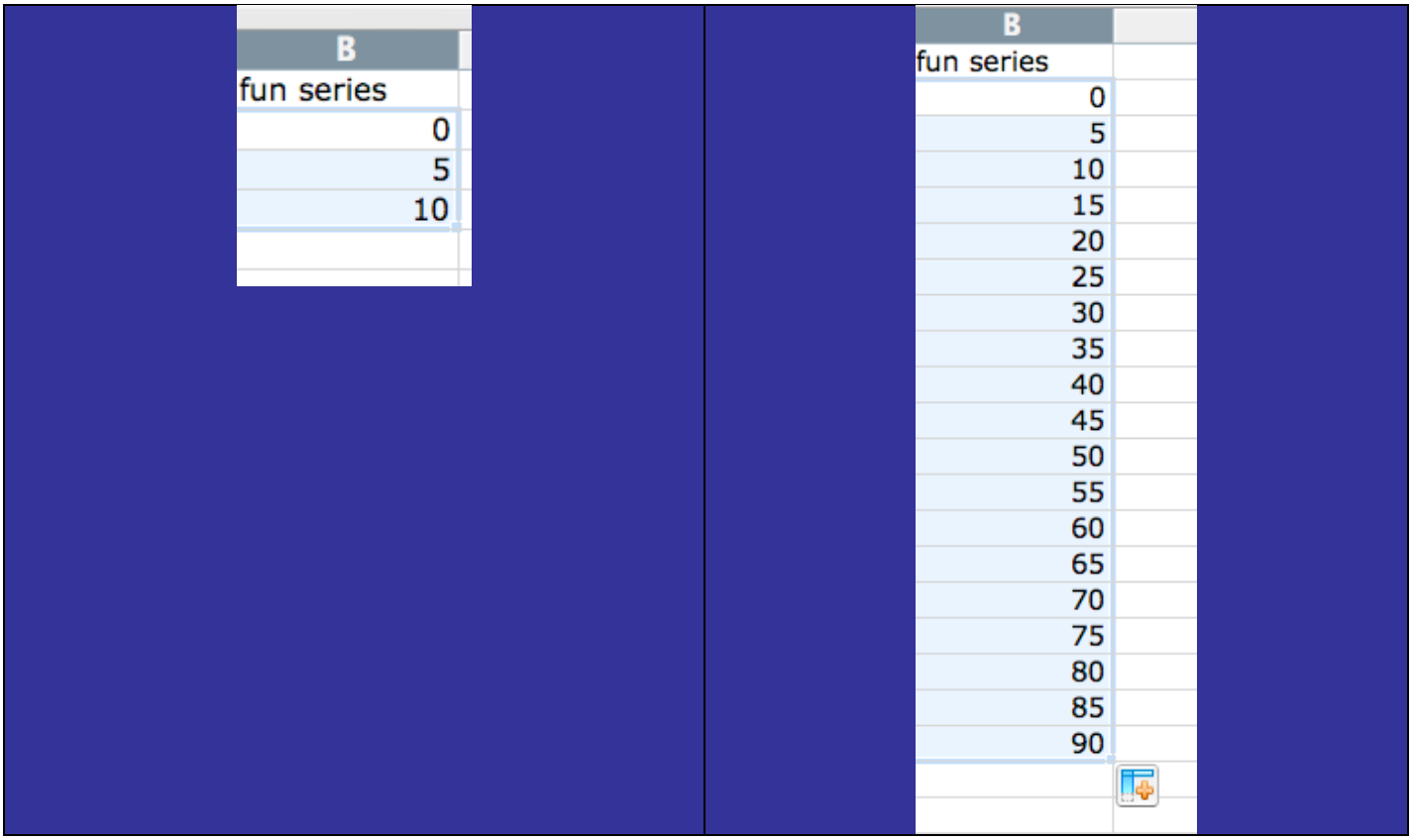
Example – Suppose you want to add 5 to each successive cell down a column, starting with 0.

Step 1: Enter 0 in cell B2, 5 in cell B3 and 10 in cell B4. Enter.

Step 2: Now highlight all three cells: B2, B3, and B4. Again, position your cursor at the lower right corner of this cell so that the cursor arrow changes to a **small black cross**.

Step 3: Click one time on the small black cross. **Without releasing, drag down** B5, B6 and so on

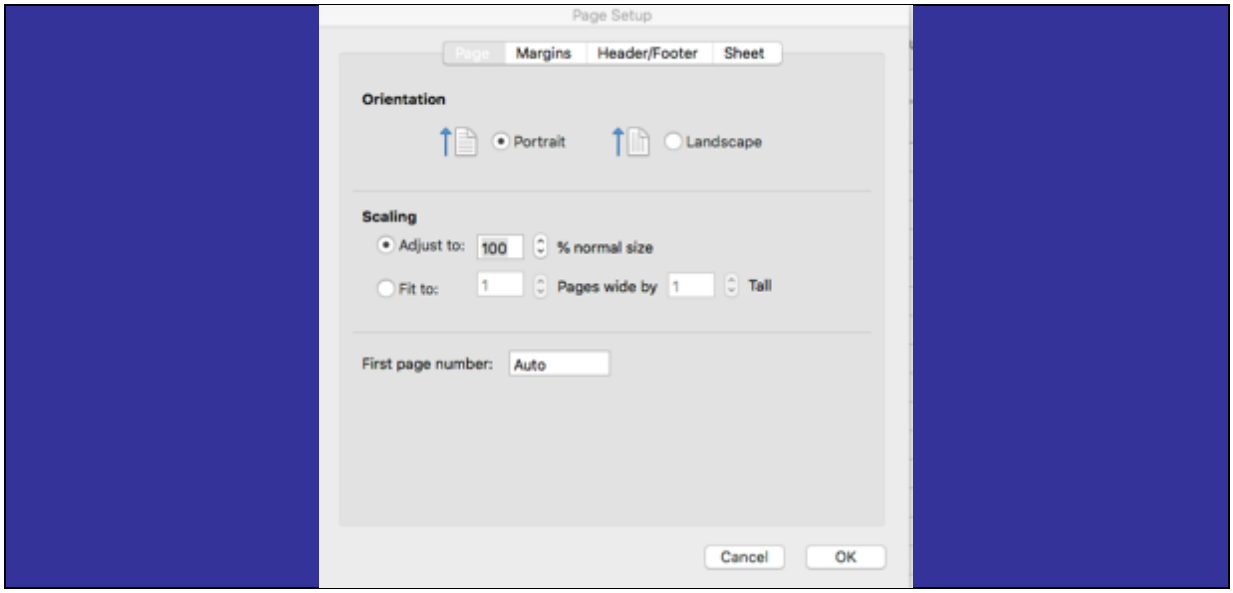
Step 4: When you release the mouse, Excel will fill in the series for you!.



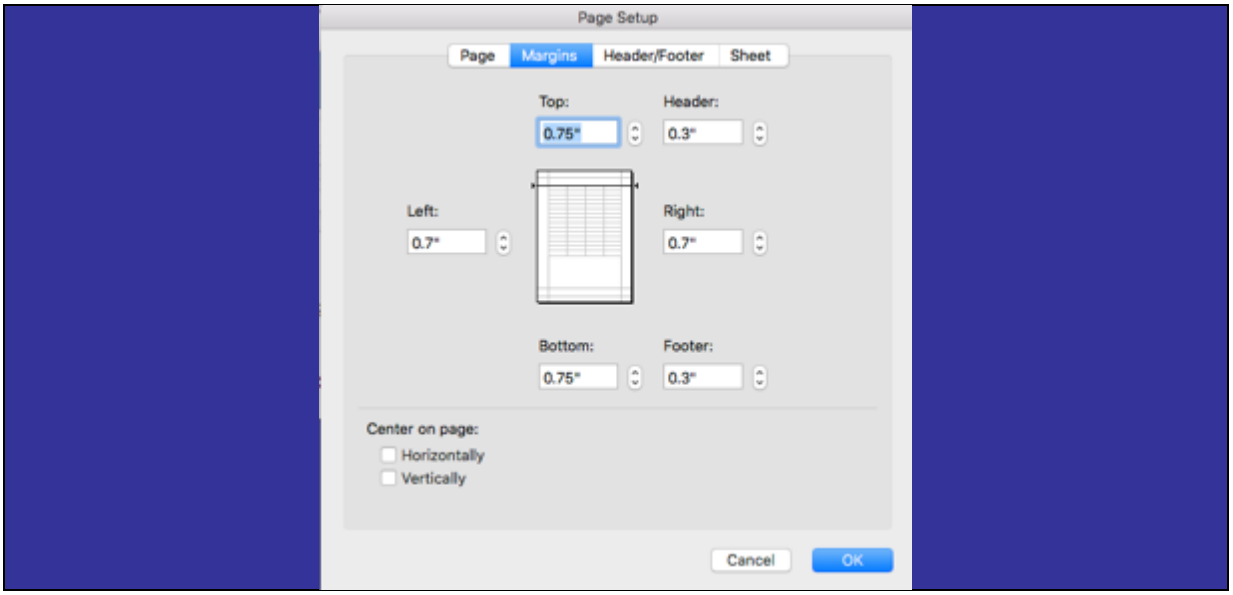
2.8 Page Setup and Printing

Before you do any printing, specify your page layout. From the top menu, choose **File>Page Setup** Four tabs with a variety of menus will appear: (1) **Page**, (2) **Margins**, (3) **Header/Footer**, and (4) **Sheet**

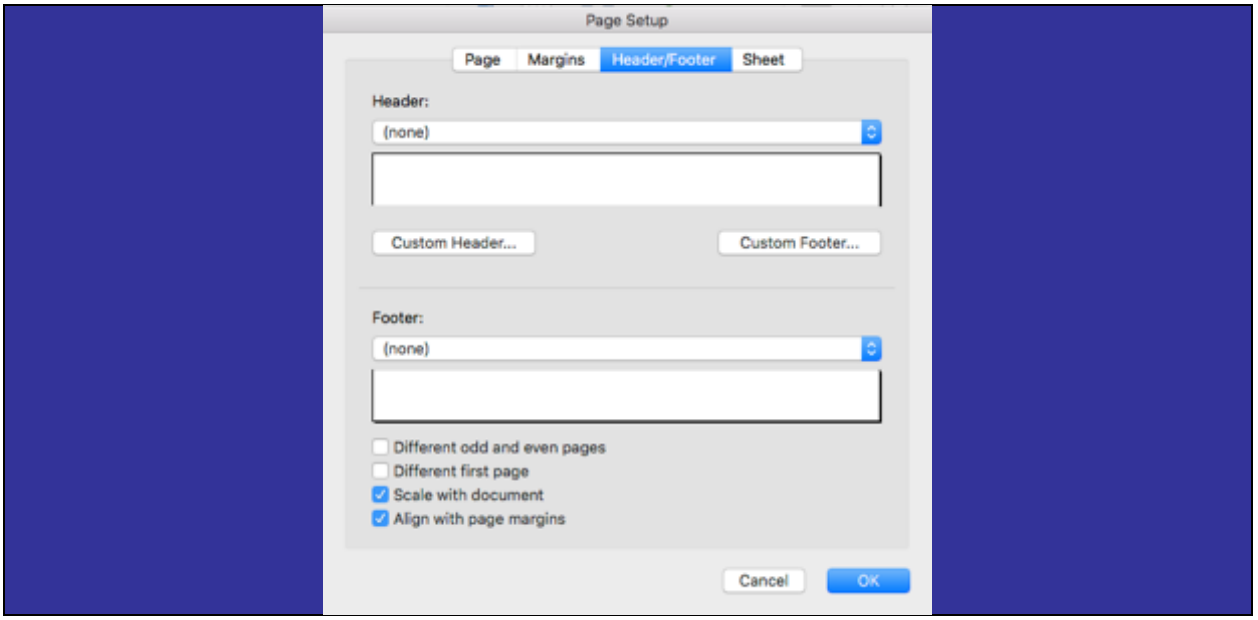
Page: Choose page orientation (**Portrait** or **Landscape**). If you do not have too many columns, choose the option **Fit to 1 page wide by 1 tall** so that all of your variables appear on one page.



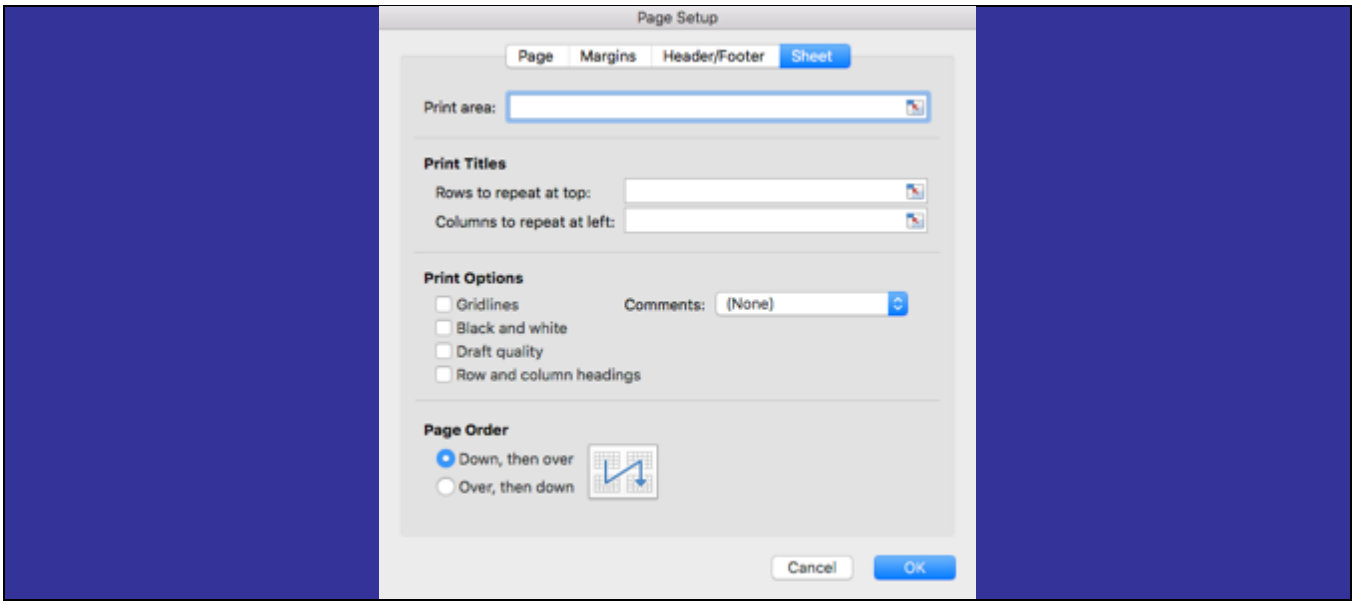
Margins: In the margins tab you can choose margins and centering.



Header/Footer: 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: **Tip!!** 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.



To preview your print out:

From the main menu bar: FILE > PRINT > PREVIEW

To print:

From the main menu bar: FILE > PRINT

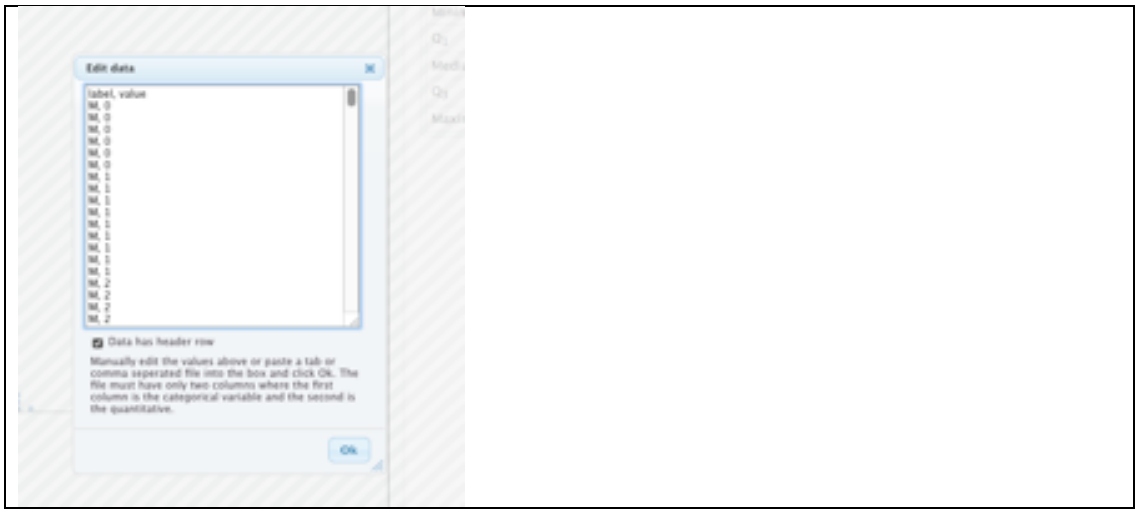
2.9 Handy for BIOSTATS 540 – How to Concatenate

Introduction – Why do I want to Concatenate?

In BIOSTATS 540, a number of online statistical software applications are introduced. These are terrific in that you can enter your data directly! The problem is that, sometimes, the required format of data entry is awkward. Two such online statistical software applications are:

1. Shodor Interactivate Box Plot
   (http://www.shodor.org/interactivate/activities/BoxPlot/)

2. StatKey
   http://www.lock5stat.com/StatKey/



What a nuisance! **In both instances, when the data consists of two variables, one qualitative and one quantitative, each row must contain the two values and they must be separated by a comma.**

Solution – Enter your data into Excel for pasting into Shodor or StatKey or whatever else

Example –
Suppose we want to enter the following data into our online statistical software application

Male, 44
Male, 16
Female, 37

Step 1
Launch Excel. Into Column A of your worksheet put your values of your first variable. For example, this might look like:
COLUMN A
Male
Male
Female

Step 2
Into Column B of your worksheet put your values of your second variable. For example, this might look like:
COLUMN B
44
16
37

Step 3
Now position your cursor in COLUMN C, row 1. Into the formula box, type the following, taking care not to forget the equal sign:

= concatenate(A1,",",B1)

You should now see in Column C, row 1:
COLUMN C
Male, 44

Step 4
Using copy>paste, repeat for the remaining rows of data.

All set!
Paste column C where you need it (StatKey or Shodor, etc)

3. Data Set Creation Basics

Example ("ICU Example")–

In BIOSTATS 540, a study was introduced of 25 consecutive patients entering the general medical/surgical intensive care unit at a large urban hospital. For each patient, the following data were collected.

Variable	Description	Code
ID	Confidential Patient Identifier	
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'll use this example of 25 observations to illustrate the steps and recommendations for data set creation using MS Excel. The data are on the next page (page 41).

Example ("ICU Example") – Data.

id	age	type_adm	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

3.1 Design Your Database First

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

- **Name of field (note – this is also your variable name)** -- a single word name used as a shorthand reference for a field

  Keep it short (8 characters or under 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 underbar (_) in place of a space.

- **Label for field** -- (optional) a longer description of data stored in the field.

- **Type of field** -- there are 2 basic types of fields that dictate the manner is which data is stored, character and numeric. Other field formats are often available, too. (e.g. date).

  Numeric -- containing only numbers

  Text or **Character** -- allowing letters, numbers, other keyboard characters

  Other field types:

  Logical -- Yes/No or True/False

  Date -- containing dates in a specified format
  (in some programs dates are stored as character data, in others, numeric)

  Some programs have other special field types – currency, percent, phone numbers, SSN, ...

  Note - In some software these are considered formats rather than field types.

- **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 (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 this is **formatting cells** (see page 21)

- Numeric and character data are stored quite differently, and you should be clear ahead of time, as to data the 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 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, etc…)

3.2 Data Entry

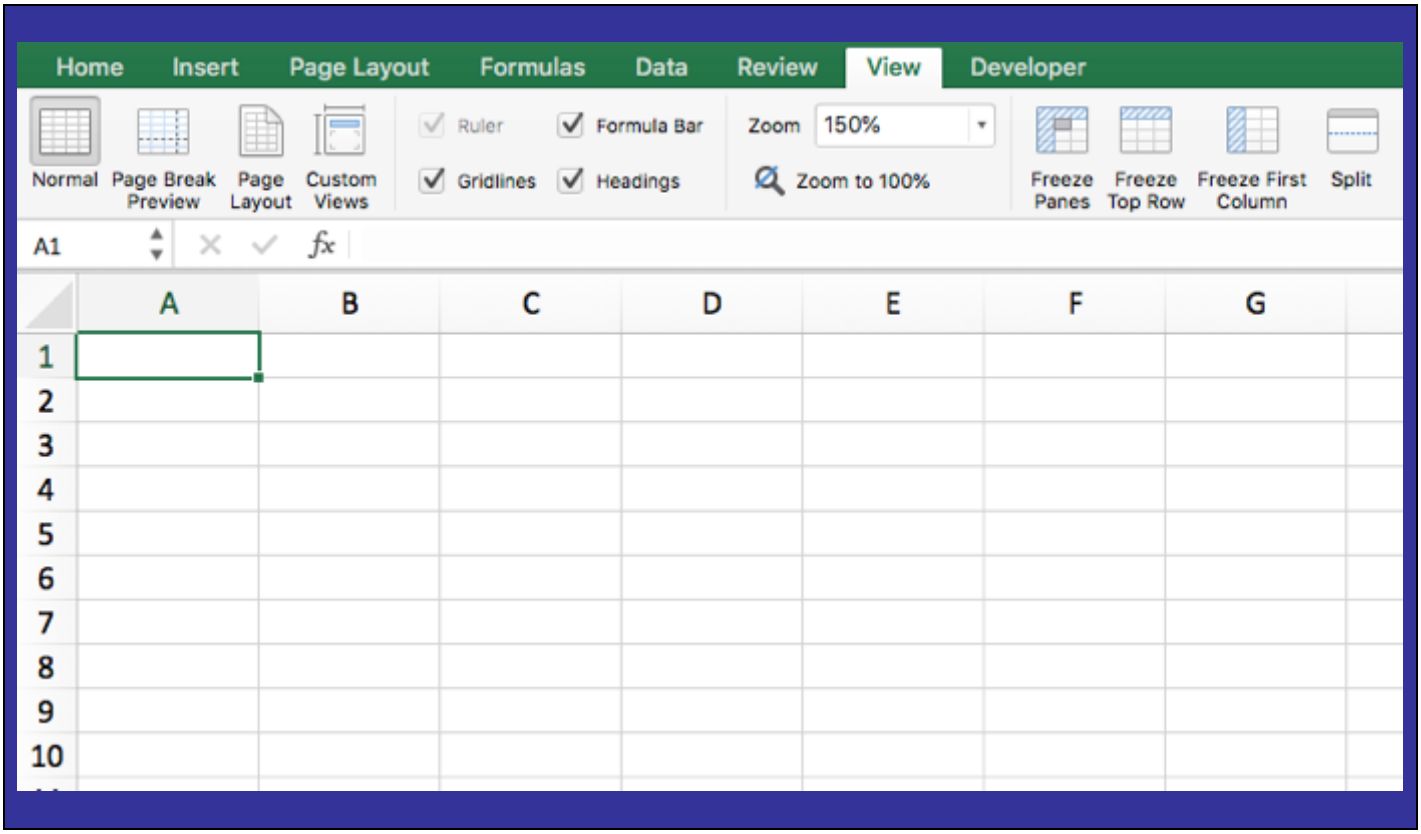
The steps are best explained in an example.

ICU Example -

Step 1: Launch MS Excel

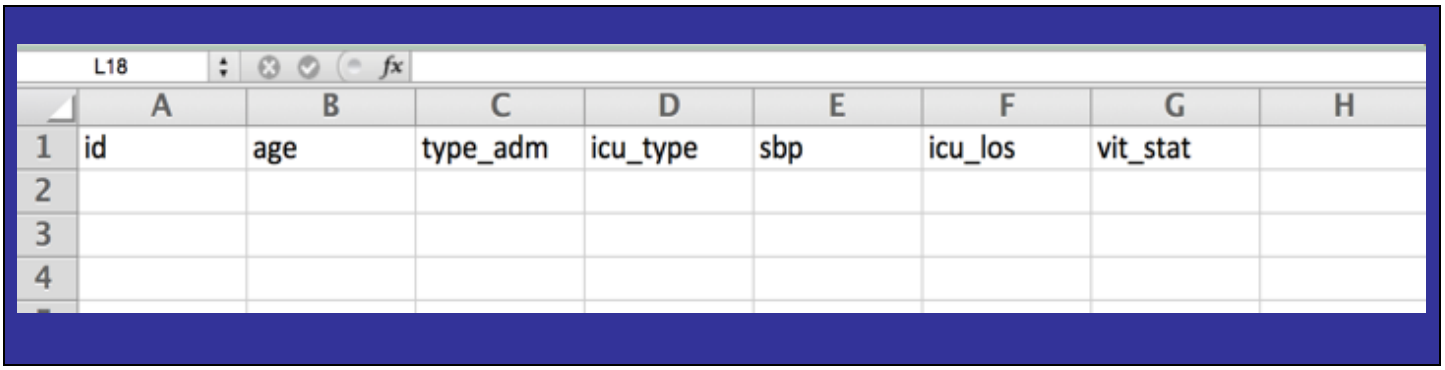
Suggestions:
Click **VIEW TAB**
Click on the **NORMAL** icon on the left
In the middle: **ZOOM > 150%**

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 see it actually) is positioned here.



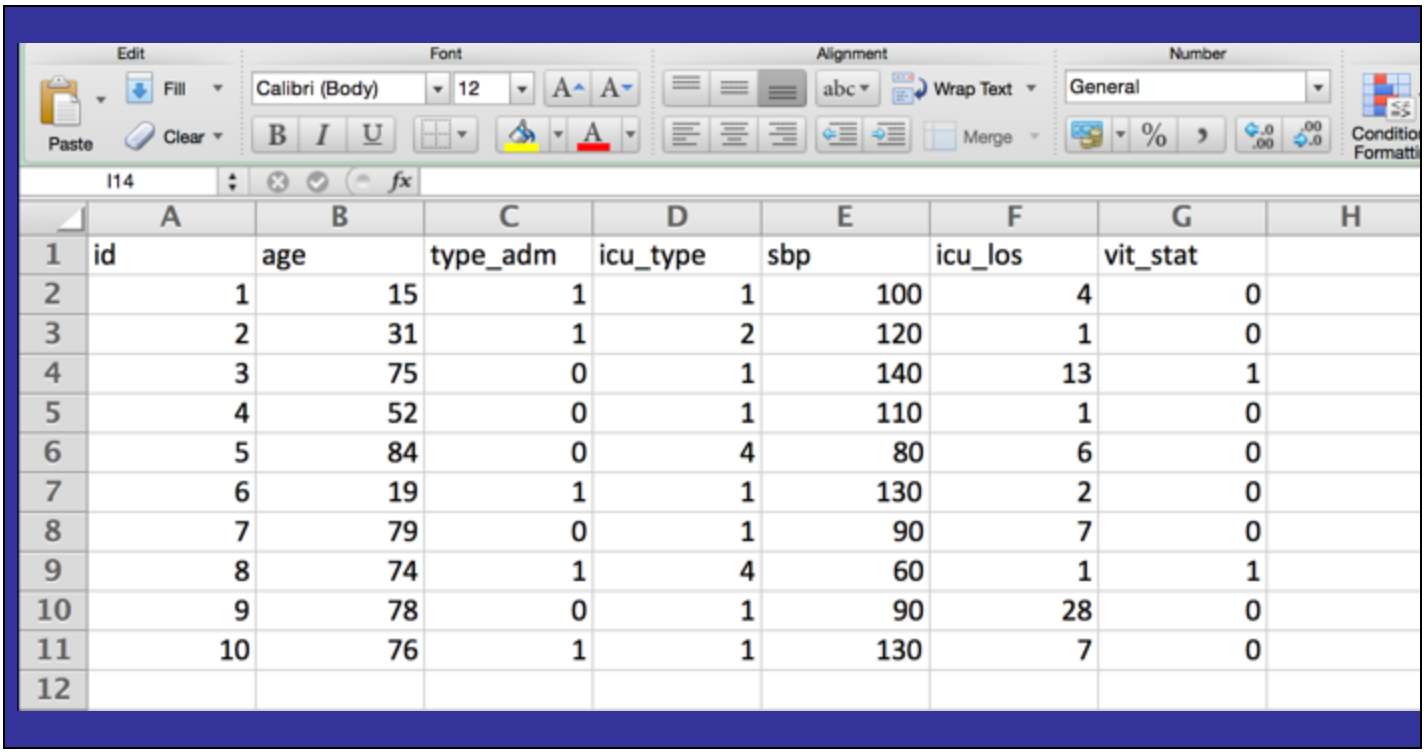
Step 2: Enter your variable names, horizontally, across row 1 as the column headings. Excel calls these fields.

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.



Step 3: 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). Press ENTER. Enter your data column by column. When you are done, you should now have the following; **note – Only a partial picture is shown here**.



	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	
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	
11	10	76	1	1	130	7	0	
12								

3.3 Formatting Fields, Field Names and Format Type

Step 4: Assign format types using instructions on page 21

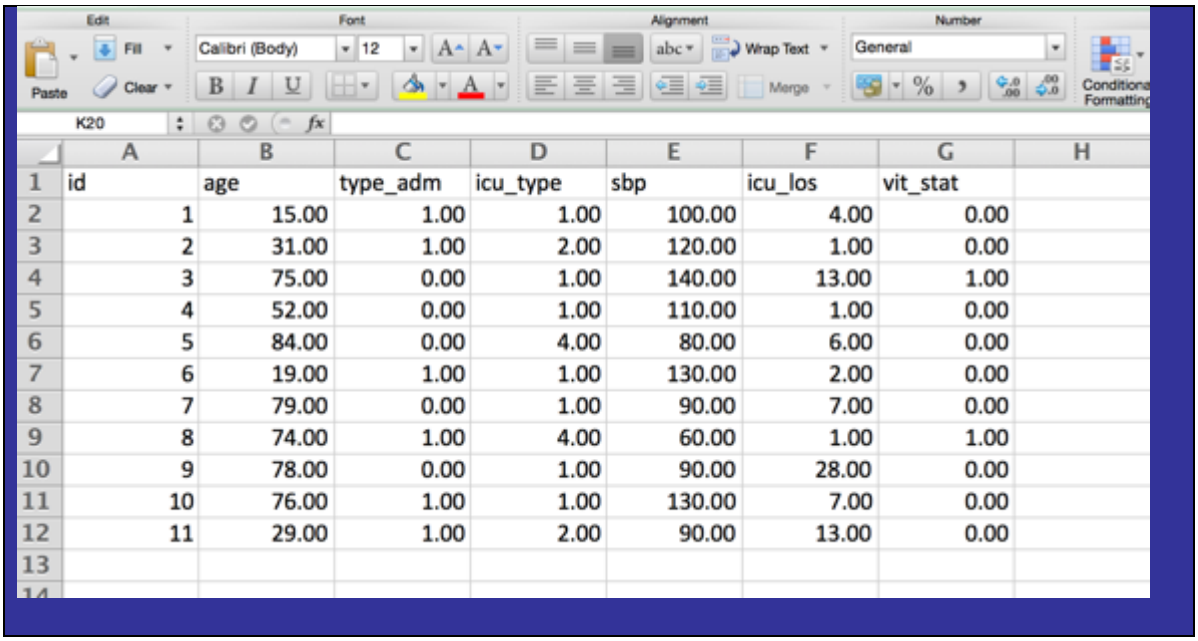
(1) For each column, select the entire column
(2) From the toolbar at top, click on **FORMAT.** From the drop-down menu, click on **CELLS.**

Example, continued –

The following are reasonable choices. **Tip!** Note that, except for the variable ID, I chose to format each variable as numeric. This makes programming convenient, as it spares having to remember special conventions in working with character fields.

Column	Variable	Format cells category:	Notes
A	**ID**	General	
B	**AGE**	Number	At right in the decimal places box, choose "2"
C	**TYPE_ADM**	Number	
D	**ICU_TYPE**	Number	
E	**SBP**	Number	At right in the decimal places box, choose "2"
F	**ICU_LOS**	Number	
G	**VIT_STAT**	Number	

You should now have the following; **note – A partial picture is shown here**.



	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	
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	
11	10	76.00	1.00	1.00	130.00	7.00	0.00	
12	11	29.00	1.00	2.00	90.00	13.00	0.00	
13								

3.4. Creating New Variables Using Formulae and Functions

Suggestion: As a general rule, it is recommended that you do *not* do variable creations in Excel. The advice is to reserve Excel file work for solely the raw data and that all new variable creation occurs in the software you are using for data analysis (e.g. R or Stata)

See again section 2.5, beginning on page 22.

Example ("ICU Example")–
We don't actually need to create a new variable, but let's do one for illustration. Suppose we want to create a new variable called AGEDAYS with the following definition:

AGEDAYS = AGE * 365.25

Step 5: Create AGEDAYS in Column H.

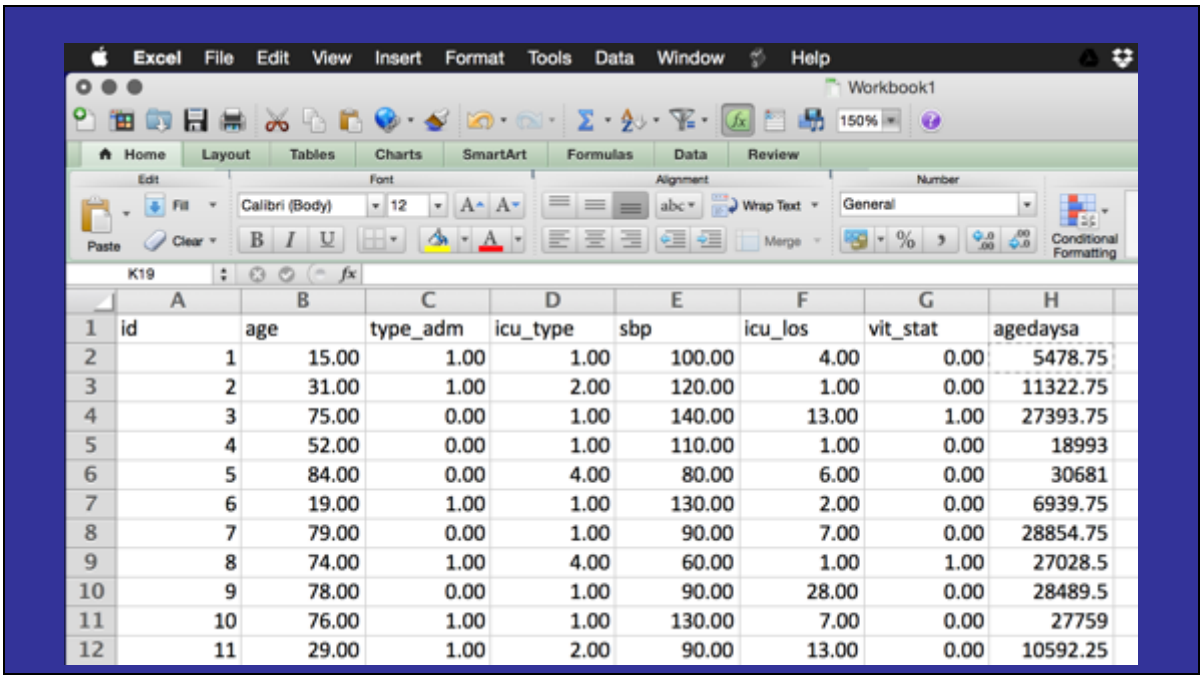
(1) In cell H1, enter the variable name **agedaysa**

(2) In cell H2, enter the calculation of agedays for the first record by typing
= B2*365.25 Press enter. You should see the result **5478.75** in cell H2

(3) Highlight cell H2. From the **menu bar**: **EDIT > COPY**. The border of cell H2 should now be dashed and vibrating!!

(4) Highlight cells H3 through H26. From the **menu bar EDIT > PASTE**

Your worksheet should now look like the following (this is a partial screen capture)



	A	B	C	D	E	F	G	H
1	id	age	type_adm	icu_type	sbp	icu_los	vit_stat	agedaysa
2	1	15.00	1.00	1.00	100.00	4.00	0.00	5478.75
3	2	31.00	1.00	2.00	120.00	1.00	0.00	11322.75
4	3	75.00	0.00	1.00	140.00	13.00	1.00	27393.75
5	4	52.00	0.00	1.00	110.00	1.00	0.00	18993
6	5	84.00	0.00	4.00	80.00	6.00	0.00	30681
7	6	19.00	1.00	1.00	130.00	2.00	0.00	6939.75
8	7	79.00	0.00	1.00	90.00	7.00	0.00	28854.75
9	8	74.00	1.00	4.00	60.00	1.00	1.00	27028.5
10	9	78.00	0.00	1.00	90.00	28.00	0.00	28489.5
11	10	76.00	1.00	1.00	130.00	7.00	0.00	27759
12	11	29.00	1.00	2.00	90.00	13.00	0.00	10592.25

3.5. Documentation with a Data Dictionary/Coding Manual

Include in your Excel file a worksheet that is a coding manual for the data. Document in the coding manual variable names, labels, type, value labels and a notes/remarks column.

Tip! Be sure to include missing value codes.

Tip!! How to get the carriage returns within a cell in Excel for MAC-

Notice that the entry in cell D7 has carriage returns. This was done as follows.

(1) Position cursor in cell D7

(2) After typing 1=emergency, do **NOT** press the enter key. Instead press **CONTROL – COMMAND -ENTER**

Example –

Sheets

	A	B	C	D	E
1	**Coding Manual ICU Study (n=25)**				
2	*version 9-29-2011*				
3					
4	**Variable**	**Label**	**Type**	**Coding**	**Remarks**
5	**id**	Patient id	character	1, 2, etc	
6	**age**	Age at admission, years	numeric	.=missing	
7	**type_adm**	Admission type	numeric	1=emergency 0=elective .=missing	
8	**icu_type**	ICU type	numeric	1=medical 2=surgical 3=cardiac 4=other .=missing	
9	**sbp**	systolic blood pressure, mm Hg	numeric	.=missing	
10	**icu_los**	Length of Stay ICU, days	numeric	.=missing	
11	**vit_stat**	Status at discharge	numeric	1=dead 0=alive .=missing	
12	**agedays**	New variable for laughs	numeric	age*265.25	
13					

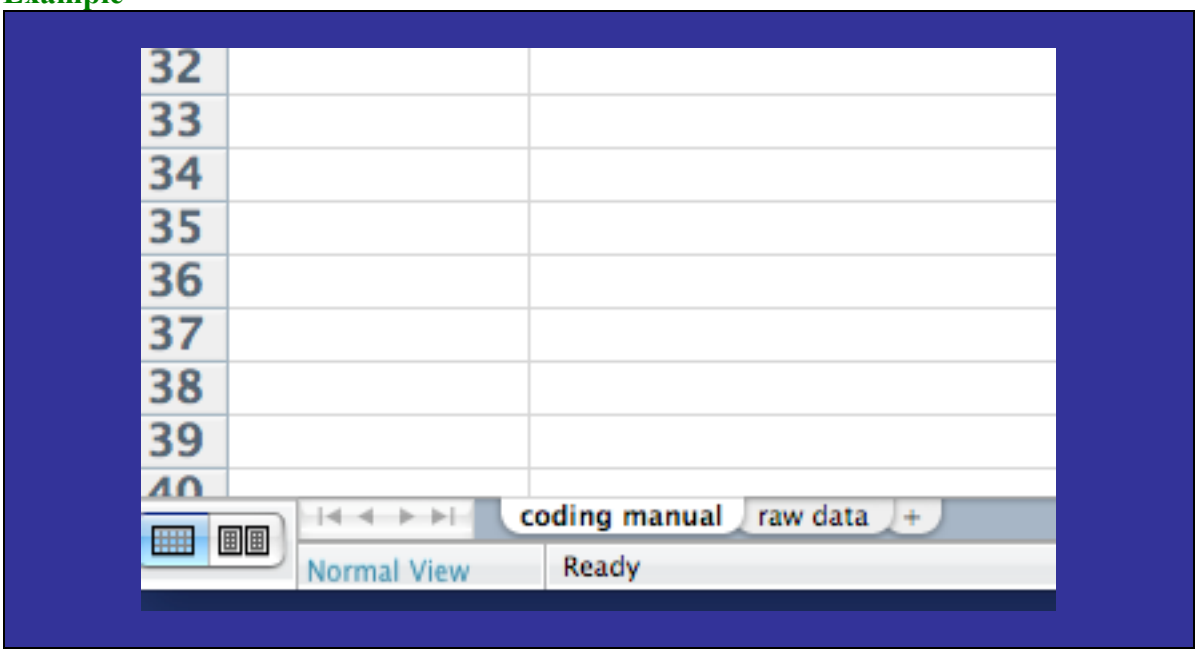
3.6. Saving and Exiting

Before exiting, lets give names to the worksheets and reorder them.

Rename worksheet
(1) Position cursor on the tab located at the bottom of your screen
(2) **Right click > RENAME**
(3) Type in the new name. Press **ENTER**

Rearrange the worksheets
(1) Activate the worksheet you want to move
(2) Position cursor on the tab at the bottom of your screen
(3) **Right click > RENAME**

Example –



Design ..... Data Collection ..... Data Management ..... Data Summarization ..... Statistical Analysis ..... Reporting