

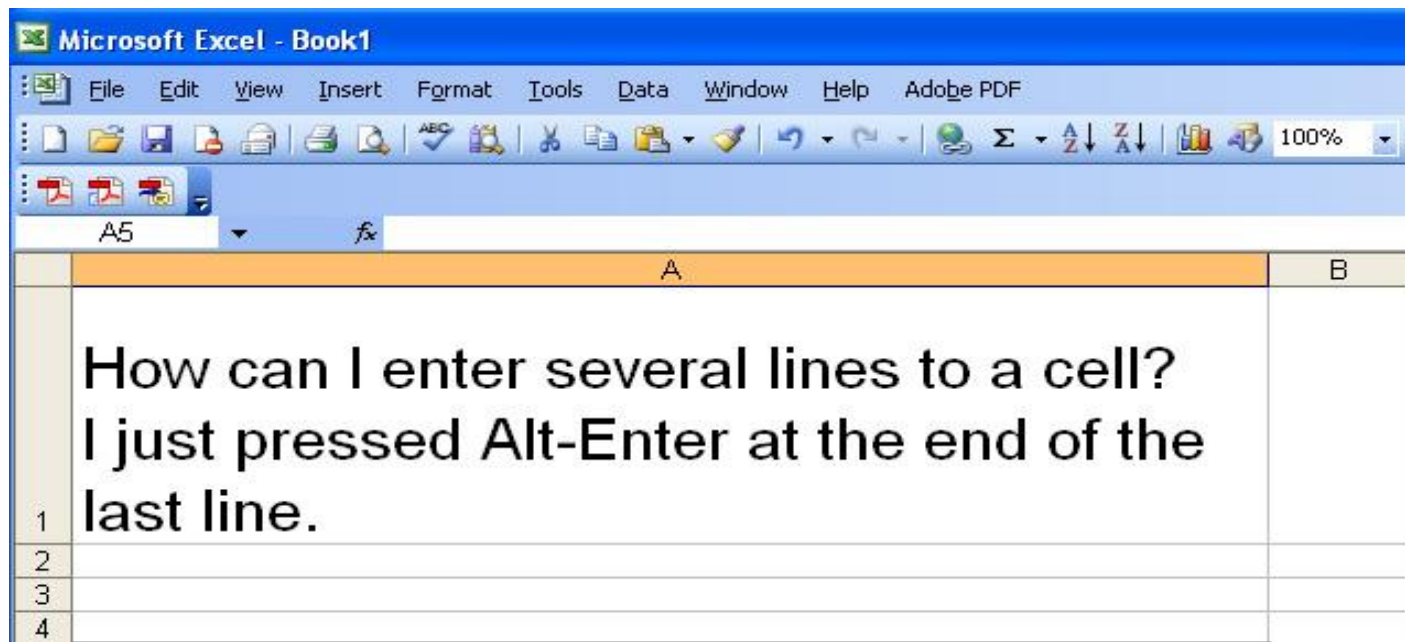
Excel

Angelo Gargantini
Basic CS Medical school

Tips and Tricks with Excel

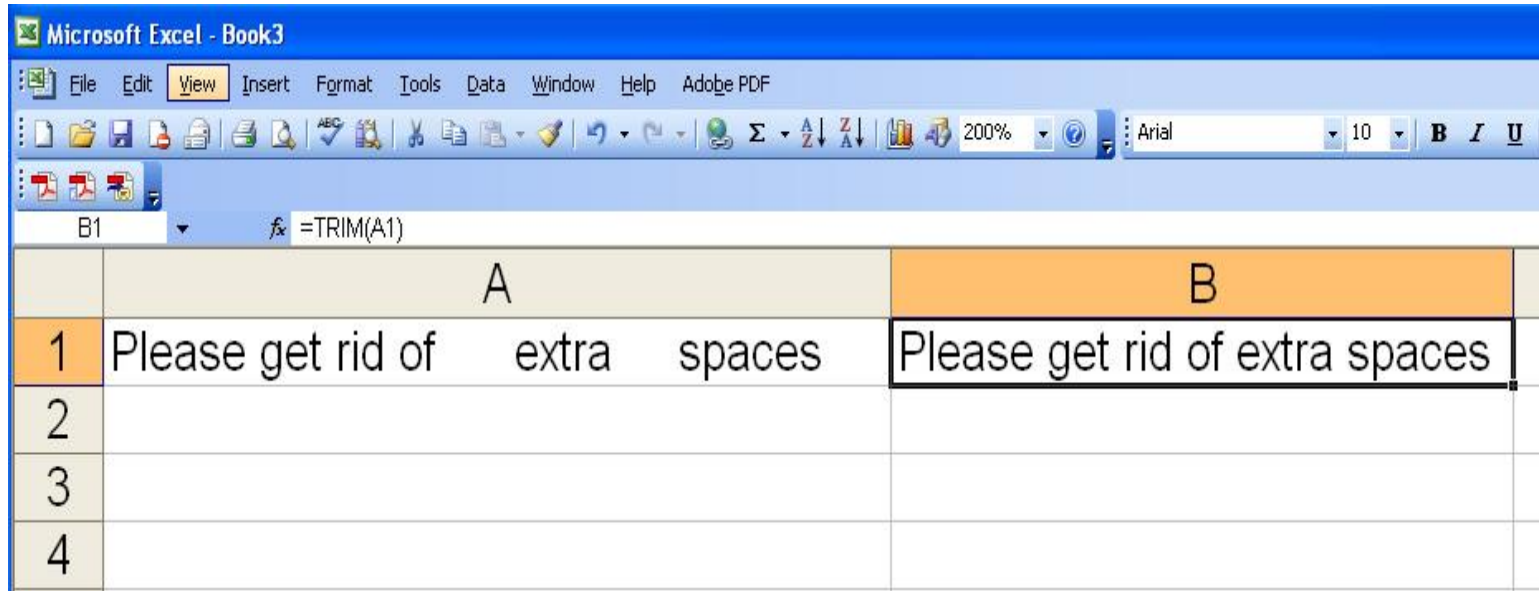
Add a carriage return to a cell's contents

- **Problem:** You need to enter a long text to a cell, each time you press Enter to separate lines, Excel just takes you to a new cell.
- **Solution:**
 - PC: press **Alt-Enter** at the end of a line
 - MAC: press **Command-Option Enter** at the end of a line



Get rid of spaces

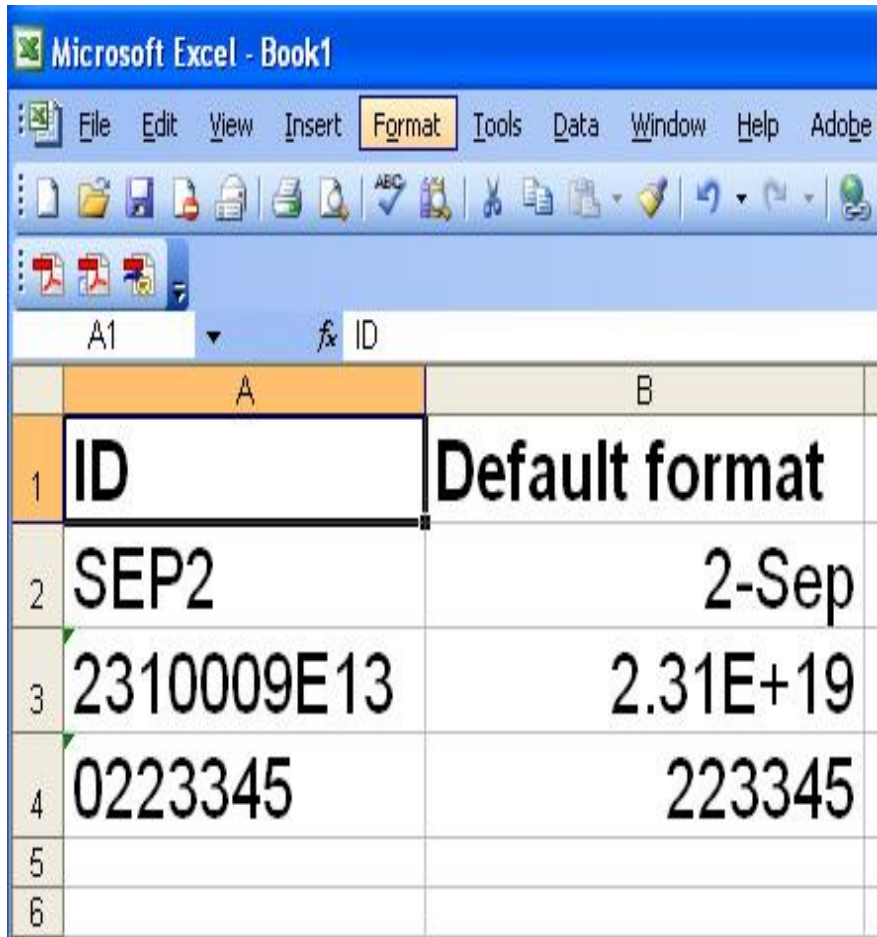
- **TRIM()**



The screenshot shows the Microsoft Excel interface with the following details:

- Window title: Microsoft Excel - Book3
- Menu bar: File, Edit, View, Insert, Format, Tools, Data, Window, Help, Adobe PDF
- Toolbar: Standard toolbar with icons for file operations, editing, and formatting. The font is set to Arial, size 10, with bold, italic, and underline options.
- Formula bar: B1, fx =TRIM(A1)
- Worksheet grid:
 - Column A: Please get rid of extra spaces
 - Column B: Please get rid of extra spaces
 - Row 1: Header row with columns A and B.
 - Row 2: Data row where the text from A1 is processed by the TRIM function in B1.
 - Row 3: Empty row.
 - Row 4: Empty row.

Symbol changed by Excel Unintended format



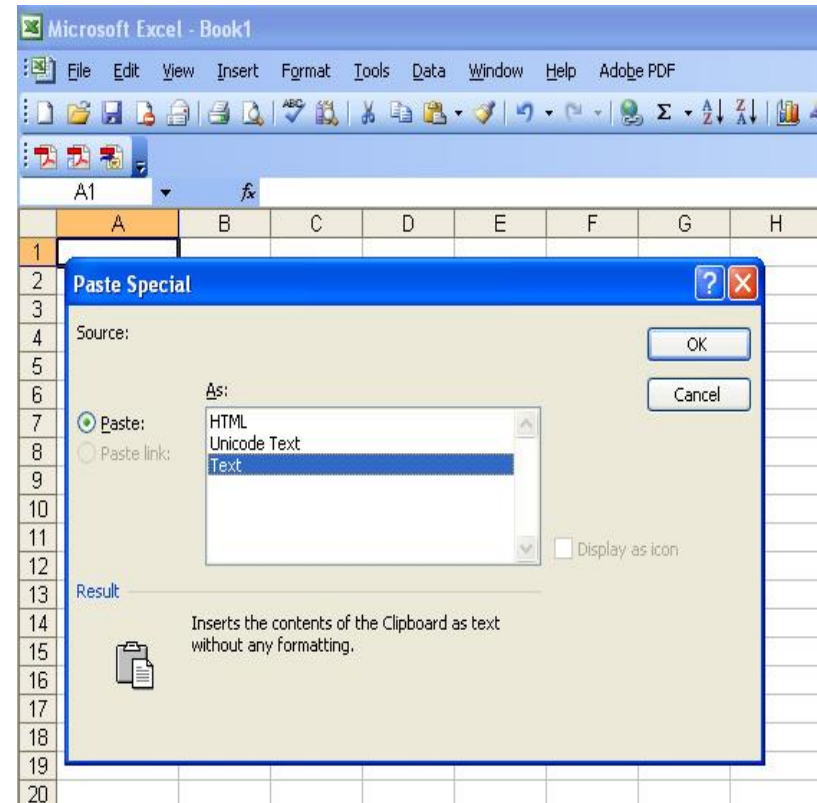
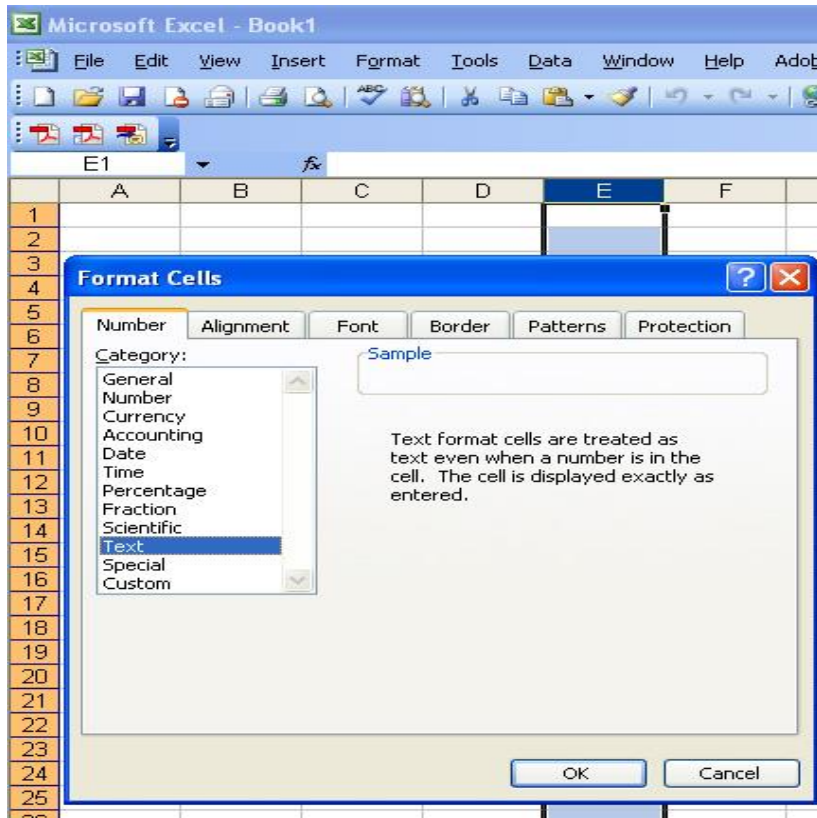
The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

	A	B
1	ID	Default format
2	SEP2	2-Sep
3	2310009E13	2.31E+19
4	0223345	223345
5		
6		

- Cause:
 - Excel automatically applies a built-in number format to a cell, based on the following criteria:
 - If a number contains month JAN to DEC, it may be converted to a date format.
 - If a number contains the letter E (in uppercase or lowercase letters; for example, 10e5), or the number contains more characters than can be displayed based on the column width and font, the number may be converted to scientific notation, or exponential, format.
 - If a number contains leading zeros, the leading zeros are dropped.

Symbol changed by Excel: Solution

1. Before you populate the data into excel, select the cells that will contain the numbers stored as text
2. Right-click and choose **Format Cells >Text > OK**
3. If a symbol is copied from an website, you may need to use **Paste Special > Text** or **Unicode Text**



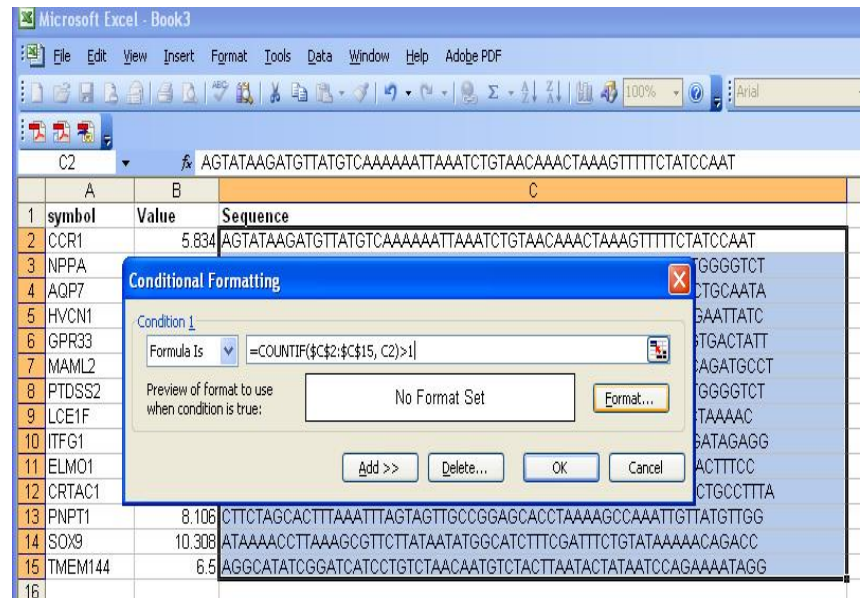
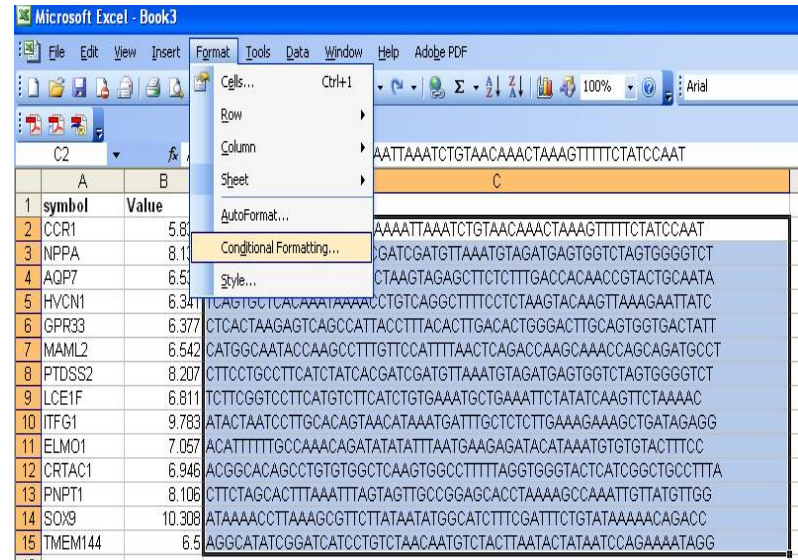
Highlight duplicated cells

- Highlight the region that may have duplicated values

- **Format > conditional formatting**

- **Formula Is:**

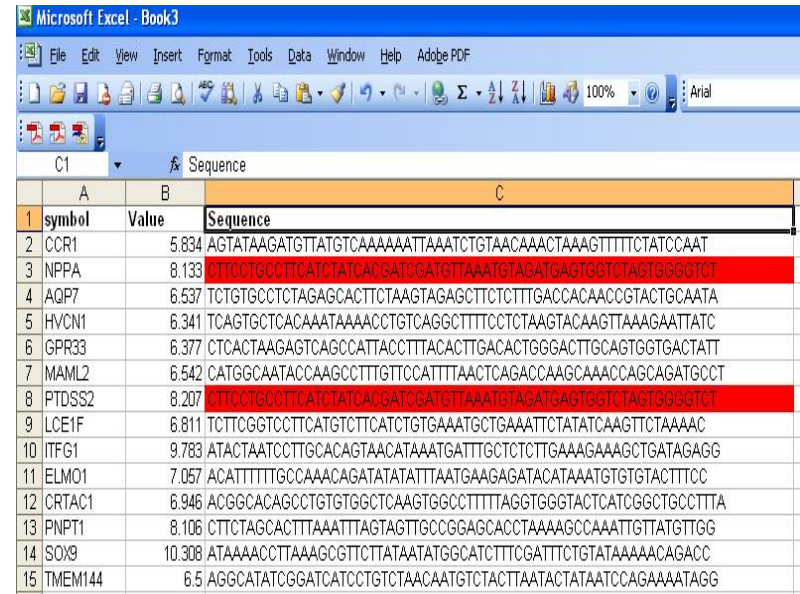
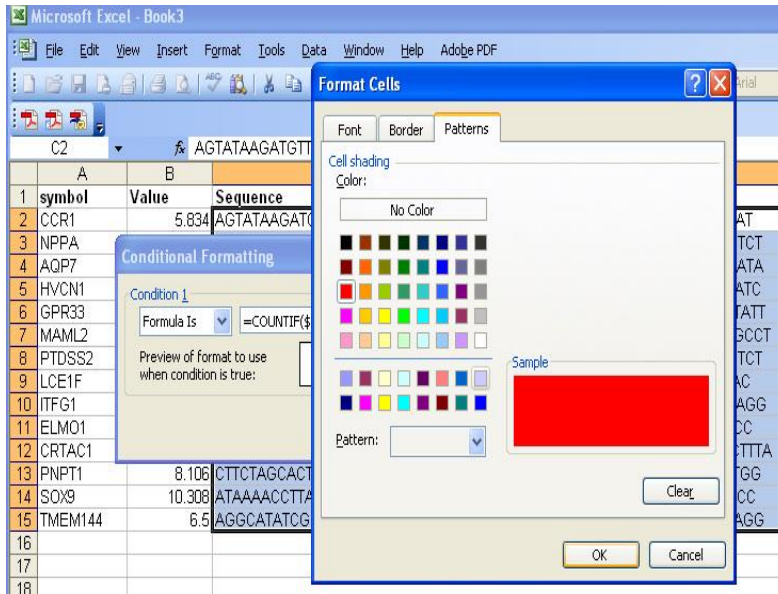
- **=COUNTIF (region, cell)>1**
- Region: absolute rows and columns
 - \$C\$2:\$C\$15
- Active cell: relative reference
 - C2



Maintain the position(s) of cell(s) in a formula:
Prefix "\$" to create an absolute link

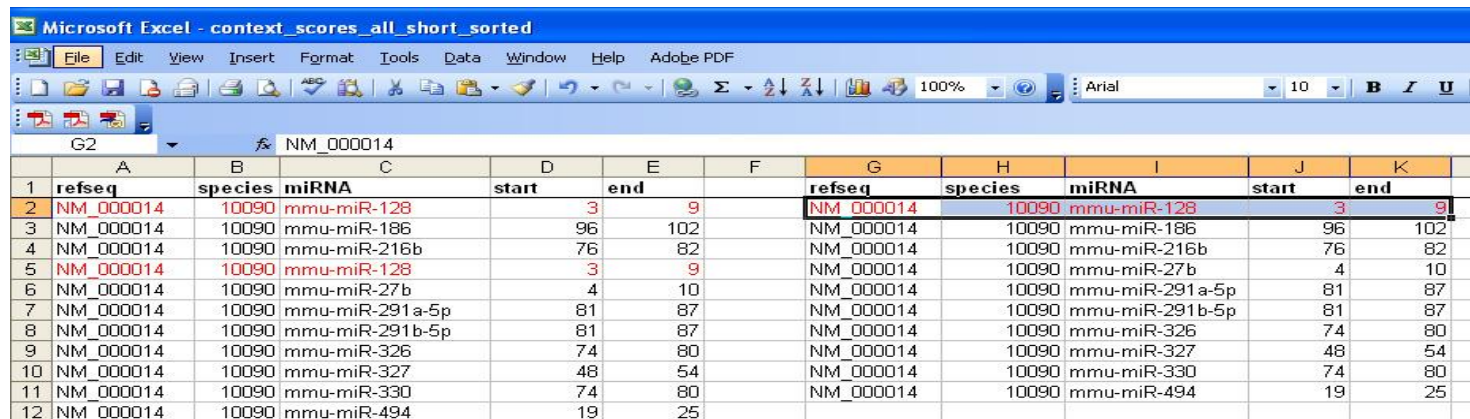
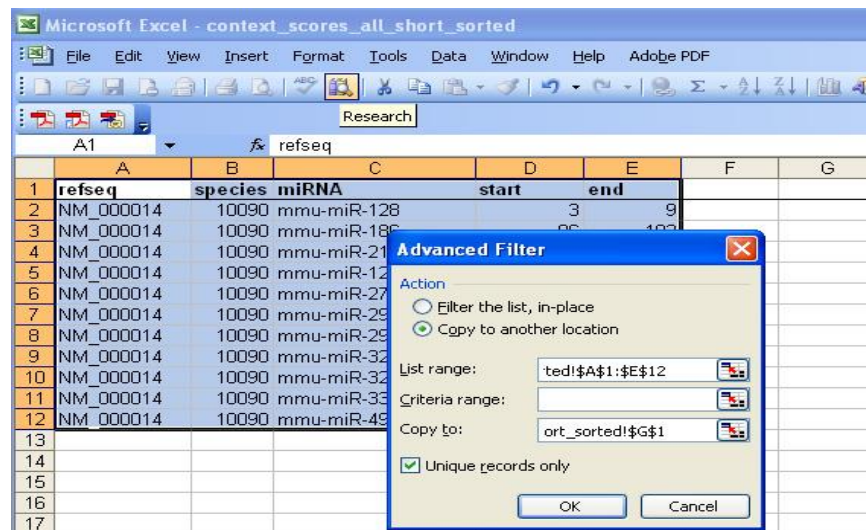
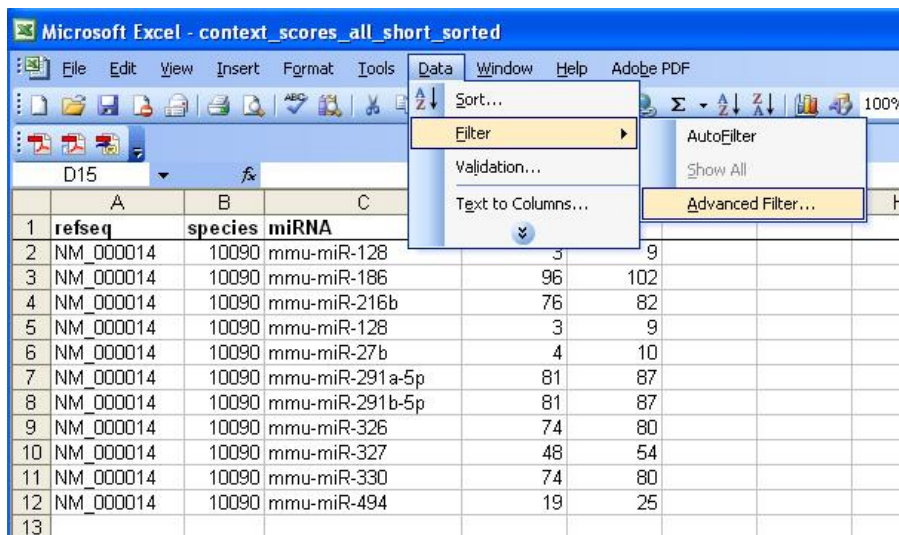
Highlight duplicated cells

- **Format > Patterns > choose color > OK**



Remove duplicated records

- **Data > Filter > Advanced Filter > Unique records only**



Move to the last row in a list

- PC:

- First cell in the active column: **Ctrl + up arrow**
- Last cell in the active column: **Ctrl + down arrow**
- First cell in the active row: **Ctrl + left arrow**
- Last cell in the active row: **Ctrl + right arrow**
- The very first cell (A1): **Ctrl+ Home**
- Bottom right corner: **Ctrl+ End**

- MAC:

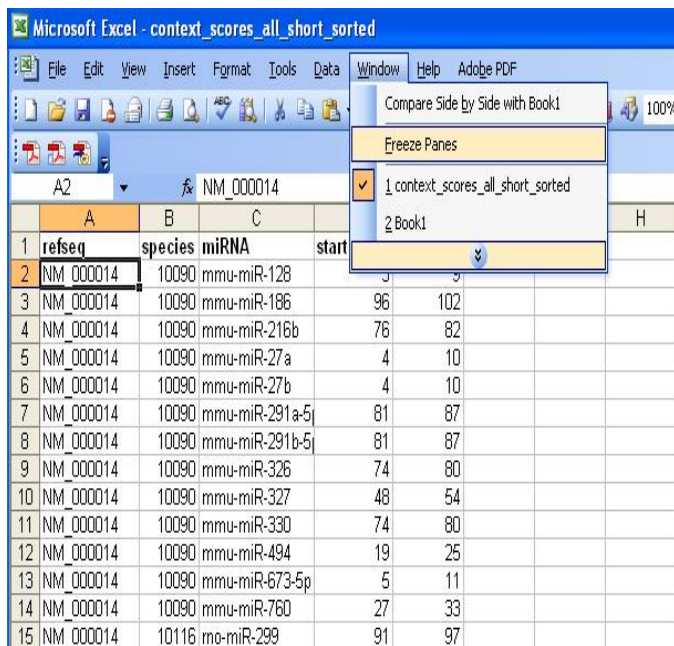
- use either **Command** or **Control**

Selecting a Range of Cells

- Selecting a large range of cells can be a pain. Here's one trick that could help.
 - To start the selection of a large matrix, **start by selecting the bottom right cell** of the desired matrix (**Ctrl + End**)
 - Use "**Ctrl + Shift + up arrow**" to select everything above the original cell.
 - Use "**Ctrl + Shift + left arrow**" to select everything to the left of the original cell.
 - Use "**Shift + down arrow**" to move down one row (to avoid the header line, if present).
 - Use "**Shift + right arrow**" to move to the right one column (to avoid any ID column(s), if present)

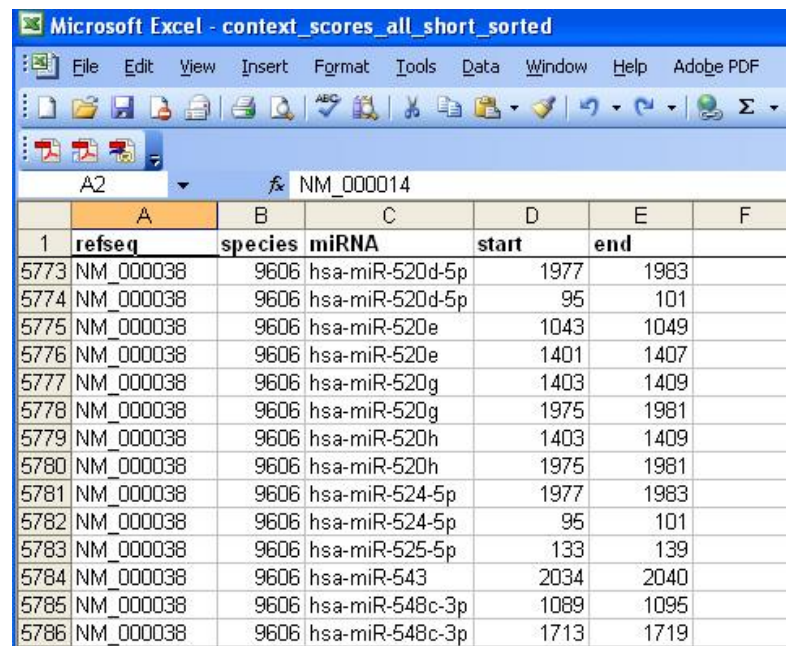
Keep headers constant as you scroll

- Click the first cell in the row below the last row you want to freeze
- Choose **Window > Freeze Panes**
- To unfreeze frozen rows, choose **Window > Unfreeze Panes**



The screenshot shows the Microsoft Excel interface with the 'Freeze Panes' menu open. The menu options are: 'Compare Side by Side with Book1', 'Freeze Panes', '1 context_scores_all_short_sorted', and '2 Book1'. The 'Freeze Panes' option is highlighted. The spreadsheet data is visible in the background, with the first row as headers and the second row selected.

	A	B	C		H
1	refseq	species	miRNA	start	
2	NM_000014	10090	mmu-miR-128		
3	NM_000014	10090	mmu-miR-186	96	102
4	NM_000014	10090	mmu-miR-216b	76	82
5	NM_000014	10090	mmu-miR-27a	4	10
6	NM_000014	10090	mmu-miR-27b	4	10
7	NM_000014	10090	mmu-miR-291a-5	81	87
8	NM_000014	10090	mmu-miR-291b-5	81	87
9	NM_000014	10090	mmu-miR-326	74	80
10	NM_000014	10090	mmu-miR-327	48	54
11	NM_000014	10090	mmu-miR-330	74	80
12	NM_000014	10090	mmu-miR-494	19	25
13	NM_000014	10090	mmu-miR-673-5p	5	11
14	NM_000014	10090	mmu-miR-760	27	33
15	NM_000014	10116	mo-miR-299	91	97



The screenshot shows the Microsoft Excel interface with the spreadsheet data. The first row is frozen, and the rest of the data is visible. The columns are labeled 'refseq', 'species', 'miRNA', 'start', and 'end'.

	A	B	C	D	E	F
1	refseq	species	miRNA	start	end	
5773	NM_000038	9606	hsa-miR-520d-5p	1977	1983	
5774	NM_000038	9606	hsa-miR-520d-5p	95	101	
5775	NM_000038	9606	hsa-miR-520e	1043	1049	
5776	NM_000038	9606	hsa-miR-520e	1401	1407	
5777	NM_000038	9606	hsa-miR-520g	1403	1409	
5778	NM_000038	9606	hsa-miR-520g	1975	1981	
5779	NM_000038	9606	hsa-miR-520h	1403	1409	
5780	NM_000038	9606	hsa-miR-520h	1975	1981	
5781	NM_000038	9606	hsa-miR-524-5p	1977	1983	
5782	NM_000038	9606	hsa-miR-524-5p	95	101	
5783	NM_000038	9606	hsa-miR-525-5p	133	139	
5784	NM_000038	9606	hsa-miR-543	2034	2040	
5785	NM_000038	9606	hsa-miR-548c-3p	1089	1095	
5786	NM_000038	9606	hsa-miR-548c-3p	1713	1719	

To freeze rows and columns along the top and left edges as you scroll

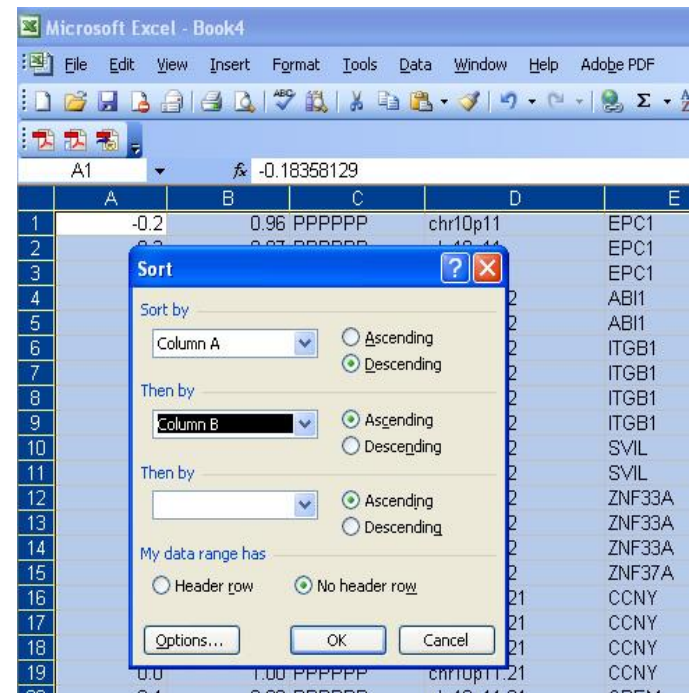
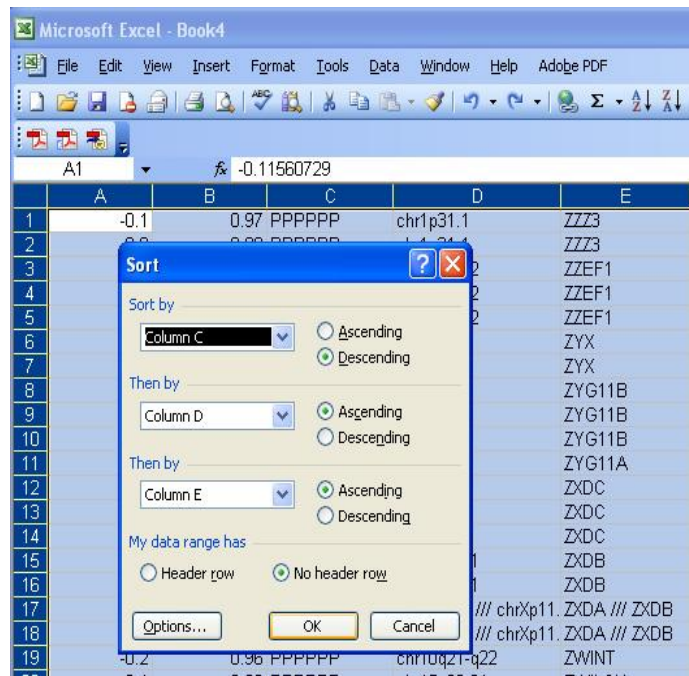
- Click the cell below and to the right of the rows and columns you want to freeze
- **Window > Freeze Panes**

	A	B	F	G
1	Name	Description	T	SJ ALLT
2	EAM193	hmr_miR-125a:bead_101-A	5	5.77132
3	EAM190	hr_miR-10b:bead_102-A	5	5
4	EAM187	hmr_miR-107:bead_103-A	5.0407	7.132
5	EAM185	hmr_miR-103:bead_104-A	6.3828	7.3855
6	EAM181	hmr_let-7f:bead_105-A	9.1065	10.2621
7	EAM179	hmr_let-7d:bead_106-A	5.84517	9.37032
8	EAM177	mr_miR-101b:bead_107-A	5	5
9	EAM175	hmr_miR-320:bead_108-A	8.07585	9.53403
10	EAM168	hmr_let-7e:bead_109-A	5	5
11	EAM161	hmr_miR-28:bead_110-A	5	6.0629
12	EAM160	hmr_miR-26b:bead_111-A	9.05839	10.2495
13	EAM155	hmr_miR-136:bead_112-A	5	5
14	EAM289	hmr_miR-129:bead_113-A	5	5
15	EAM283	mr_miR-211:bead_114-A	5	5
16	EAM282	m_miR-199b:bead_115-A	5	5

	A	B	AS	AT	AU	AV	AW
1	Name	Description	T	SJ ALLT	SJ ALLT	SJ ALLT	SJ ALLT
205	EAM262	hmr_miR-24:bead_141-C	10.7642	10.4262	10.9083	10.0856	9.1264
206	EAM261	hmr_miR-23b:bead_142-C	9.78359	9.07827	9.96377	9.06908	8.25569
207	EAM260	hmr_miR-23a:bead_143-C	9.9768	9.20423	10.4016	9.3761	7.91577
208	EAM256	h_miR-220:bead_144-C	5	5	5	5	5
209	EAM255	hmr_miR-22:bead_145-C	6.6803	6.5739	6.9621	5	5.9253
210	EAM248	hmr_miR-213:bead_146-C	8.0225	7.9829	7.5319	5	7.8359
211	EAM244	hmr_miR-21:bead_147-C	11.4979	9.46292	10.7102	9.3822	10.8146
212	EAM240	hmr_miR-20:bead_148-C	11.6498	10.3633	10.6887	9.41711	11.3248
213	EAM237	hmr_miR-19b:bead_149-C	11.6564	10.5444	11.0681	10.6726	11.409
214	EAM233	hmr_miR-196a:bead_150-C	6.1346	5.6395	5	5	5
215	EAM214	hm_miR-148a:bead_151-C	5	5	5	5	7.1788
216	EAM212	hmr_miR-145:bead_152-C	5	5.0443	5	5	5
217	EAM211	hmr_miR-144:bead_153-C	5.31221	6.73872	5	5	5
218	EAM210	hmr_miR-143:bead_154-C	5	5.233	5	5	5
219	EAM208	hmr_miR-141:bead_155-C	5.3976	5	6.0522	5	5

Sort by more than 3 columns

If you want to sort by columns A B C D E, select the whole spreadsheet, than sort by C D E, than A B. This will result in all five columns being sorted.



About Formula

Problem: values changed due to an extra column inserted

	A	B	C	D	E	F
1	Name	exp1	exp2	exp3	AVERAGE	
2	COL1A2	1	1	1	1	
3	POSTN	1	1	1	1	
4	COL3A1	1	1	1	1	
5	COL3A1	1	1	1	1	
6	COL4A1	2	2	2	2	
7	COL3A1	2	2	2	2	
8	SPARC	2	2	2	2	
9	SPARC	2	2	2	2	
10	POSTN	5	5	5	5	
11	IGFBP7	5	5	5	5	
12	TGFBI	5	5	5	5	
13	IGFBP7	5	5	5	5	
14	IGFBP7	5	5	5	5	
15	TGFBI	5	5	5	5	

	A	B	C	D	E	F
1	Name	exp1	exp2	NEW	exp3	AVERAGE
2	COL1A2	1	1	10	1	3.25
3	POSTN	1	1	10	1	3.25
4	COL3A1	1	1	10	1	3.25
5	COL3A1	1	1	10	1	3.25
6	COL4A1	2	2	10	2	4
7	COL3A1	2	2	10	2	4
8	SPARC	2	2	10	2	4
9	SPARC	2	2	10	2	4
10	POSTN	5	5	10	5	6.25
11	IGFBP7	5	5	10	5	6.25
12	TGFBI	5	5	10	5	6.25
13	IGFBP7	5	5	10	5	6.25
14	IGFBP7	5	5	10	5	6.25
15	TGFBI	5	5	10	5	6.25

Solution: Copy → Paste Special → Values

	A	B	C	D	E	F
1	Name	exp1	exp2	exp3	AVERAGE	
2	COL1A2	1	1	1	1	
3	POSTN	1	1	1	1	
4	COL3A1	1	1	1	1	
5	COL3A1	1	1	1	1	
6	COL4A1	2	2	2	2	
7	COL3A1	2	2	2	2	
8	SPARC	2	2	2	2	
9	SPARC	2	2	2	2	
10	POSTN	5	5	5	5	
11	IGFBP7	5	5	5	5	
12	TGFBI	5	5	5	5	
13	IGFBP7	5	5	5	5	
14	IGFBP7	5	5	5	5	
15	TGFBI	5	5	5	5	

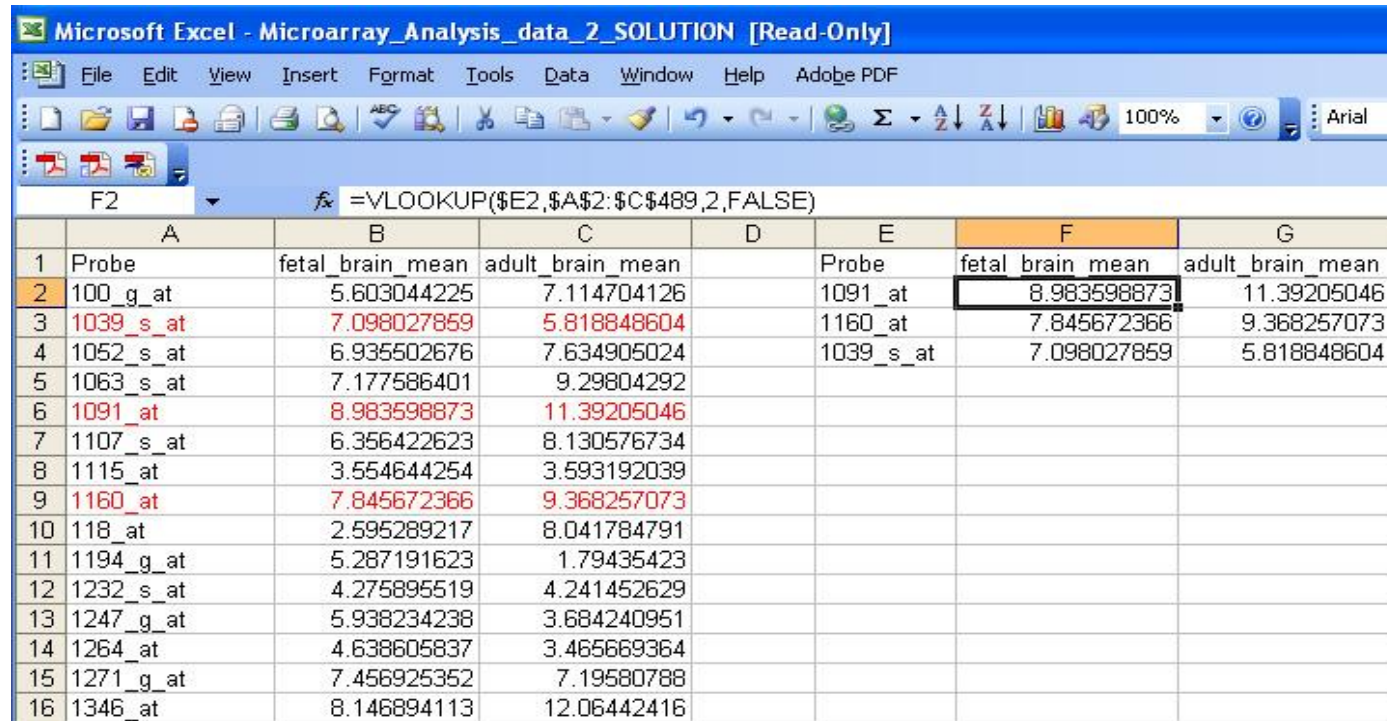
	A	B	C	D	E	F
1	Name	exp1	exp2	exp3	AVERAGE	
2	COL1A2	1	1	1	1	
3	POSTN	1	1	1	1	
4	COL3A1	1	1	1	1	
5	COL3A1	1	1	1	1	
6	COL4A1	2	2	2	2	
7	COL3A1	2	2	2	2	
8	SPARC	2	2	2	2	
9	SPARC	2	2	2	2	
10	POSTN	5	5	5	5	
11	IGFBP7	5	5	5	5	
12	TGFBI	5	5	5	5	
13	IGFBP7	5	5	5	5	
14	IGFBP7	5	5	5	5	
15	TGFBI	5	5	5	5	
16	COL1A2					10
17	MFAP5					10
18	LUM					10
19	MFAP5					10

vlookup

Look up values in data list: vlookup

- **VLOOKUP(lookup_value,table_array,col_index_num,not_exact_match)**
 - **lookup_value**: The value to search in the first column of the table array
 - **table_array**: The table to search (containing the value to search for in the first column)
 - **col_index_num**: the column number from which the matching value is returned
 - **not_exact_match**:
 - True/omitted, an exact/approximate match
 - False: an exact match

Vlookup example



The screenshot shows a Microsoft Excel window titled "Microarray_Analysis_data_2_SOLUTION [Read-Only]". The formula bar displays the formula `=VLOOKUP($E2,$A$2:$C$489,2,FALSE)`. The active cell is F2, which contains the value 8.983598873. The table below shows the data being used for the VLOOKUP operation.

	A	B	C	D	E	F	G
1	Probe	fetal_brain_mean	adult_brain_mean		Probe	fetal_brain_mean	adult_brain_mean
2	100_g_at	5.603044225	7.114704126		1091_at	8.983598873	11.39205046
3	1039_s_at	7.098027859	5.818848604		1160_at	7.845672366	9.368257073
4	1052_s_at	6.935502676	7.634905024		1039_s_at	7.098027859	5.818848604
5	1063_s_at	7.177586401	9.29804292				
6	1091_at	8.983598873	11.39205046				
7	1107_s_at	6.356422623	8.130576734				
8	1115_at	3.554644254	3.593192039				
9	1160_at	7.845672366	9.368257073				
10	118_at	2.595289217	8.041784791				
11	1194_g_at	5.287191623	1.79435423				
12	1232_s_at	4.275895519	4.241452629				
13	1247_g_at	5.938234238	3.684240951				
14	1264_at	4.638605837	3.465669364				
15	1271_g_at	7.456925352	7.19580788				
16	1346_at	8.146894113	12.06442416				

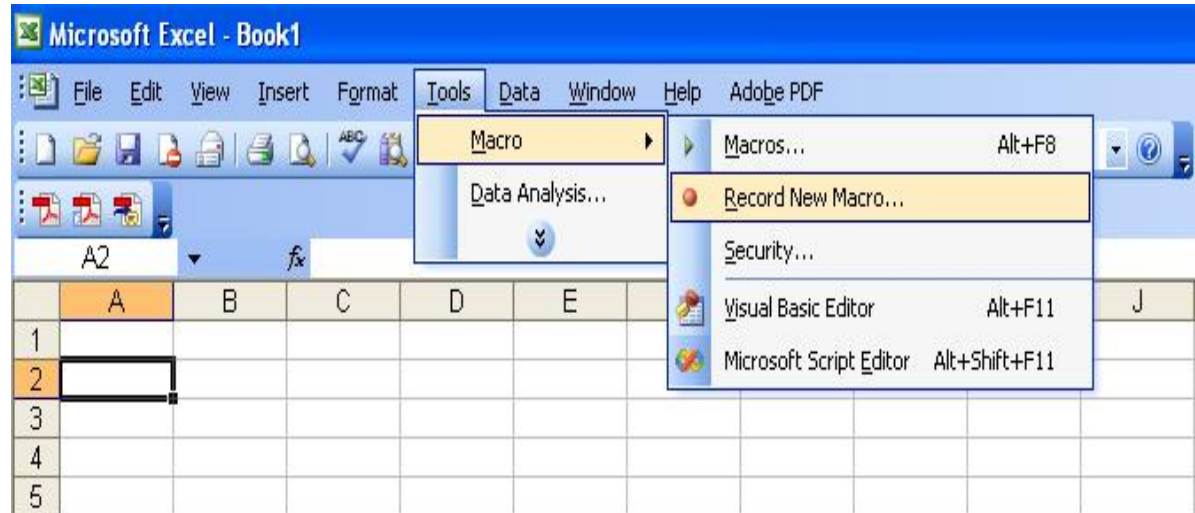
Too complicated?

Try BaRC Submatrix Selector

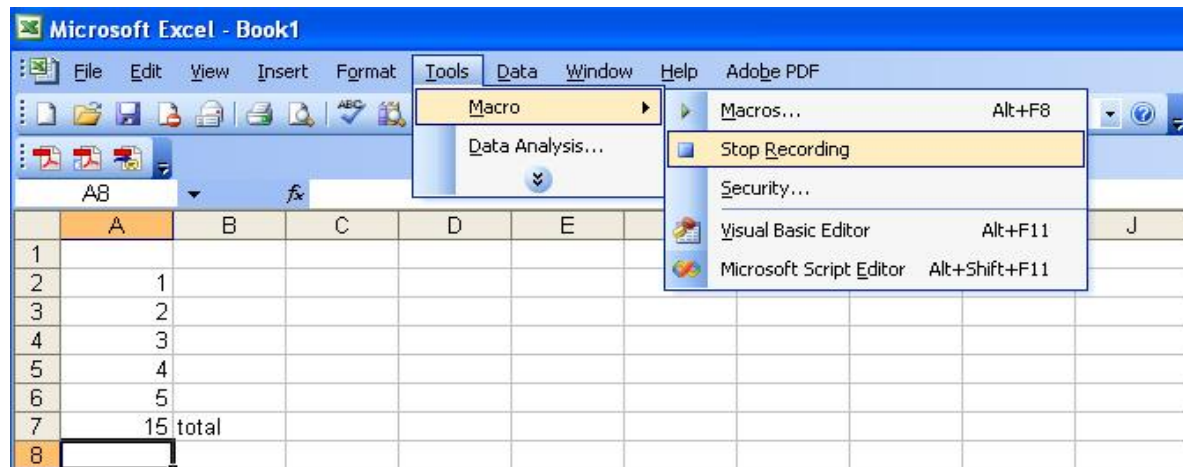
http://iona.wi.mit.edu/bell/submatrix_selector

Macro

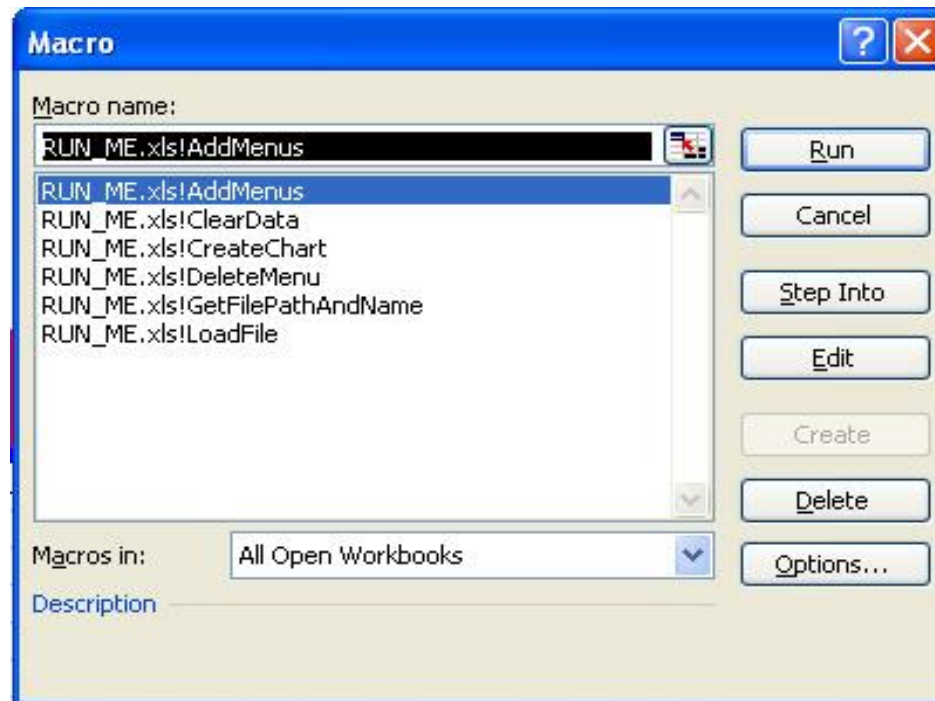
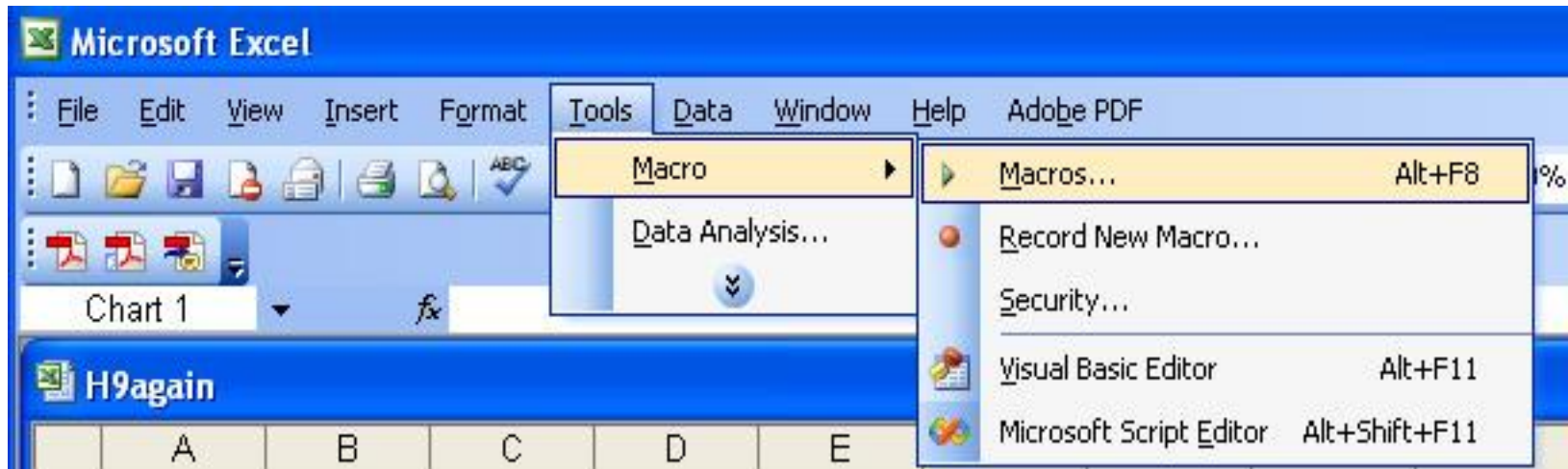
- Record macro



- Stop Recording macro



Run macro



Task one

- Open a new page in Microsoft Excel.

	A	B	C	D
1	x7			
2		1		
3		2		
4		3		
5		4		
6		5		
7		6		
8		7		
9		8		
10		9		
11		10		
12		11		
13		12		
14				
15				
16				
17				
18				

1. Enter this data as shown in the right cells.
2. We are going to make up some Multiplication tables. Click on cell B2.
3. Enter the number 7
4. Highlight the column from B2 to B13
5. On the Menu bar. Select: Edit - Fill - Down. The whole column will fill with 7's.
6. Click on cell C2. We are going to enter the calculation.

What do we need to enter first?

Task one

	A	B	C	D
1	x7			
2		1	7	
3		2	7	
4		3	7	
5		4	7	
6		5	7	
7		6	7	
8		7	7	
9		8	7	
10		9	7	
11		10	7	
12		11	7	
13		12	7	
14				
15				
16				
17				

$$=a2*b2$$

This is because these are the cells with the numbers we want to multiply.

Note what happens when you enter the formula.

Now use the same method that you used to FILL column B with 7's to repeat the formula in column C.

Can you remember how?

Task one

	A	B	C	D
1	x7			
2	1	7	7	
3	2	7	14	
4	3	7	21	
5	4	7	28	
6	5	7	35	
7	6	7	42	
8	7	7	49	
9	8	7	56	
10	9	7	63	
11	10	7	70	
12	11	7	77	
13	12	7	84	
14				
15				
16				

If you have done it right, which you will have done!

Your spreadsheet should look like this.



Task Two:



In a new spreadsheet, can you create the 8x table using the techniques we have learned so far.


Task Three

	A	B	C	D	E	F	G	H
1	Kit	Quantity	Cost	Total for 1	Total for 10	Total for Y6		
2	T-Shirt	2	£ 4.25					
3	Shorts	2	£ 5.50					
4	Trainers	1	£ 22.00					
5	Socks	2	£ 1.50					
6	Plimsolls	1	£ 9.00					
7			Grand Total					
8								
9								
10								
11								
12								

With the information shown above you can calculate the cost of your new PE Kit for yourself.

Fill in cells D2, D3, D4, D5 and D6.

Now select cell D7 and click on this symbol on the toolbar . Press RETURN and the column total appears 

Use multiplication to fill column E and F, and use  to total up the columns.

Task four

- Creating Graphs
- You can use spreadsheets to create graphs of your data.
- Of course you will need some data to make graphs of!

	A	B	C	D
1	YEAR GROUP	BOYS	GIRLS	TOTAL
2				
3	1			
4	2			
5	3			
6	4			
7	5			
8	6			
9				
10	TOTAL			
11				

Open a new spreadsheet and set up the columns and rows as shown here.

Now you need to find out how many boys and how many girls are in each year group.

Find out and enter the data.

In column D use a formula to add up the totals.

Click on B10

Then click on  on the toolbar.

Task four

- Creating Graphs

- This tool  allows you to add up column totals.

Click on B3, hold down the shift key and click on B8.

The column total should appear in B10.

Use this method to complete the rest of this spreadsheet.

Now you are ready to create a graph!

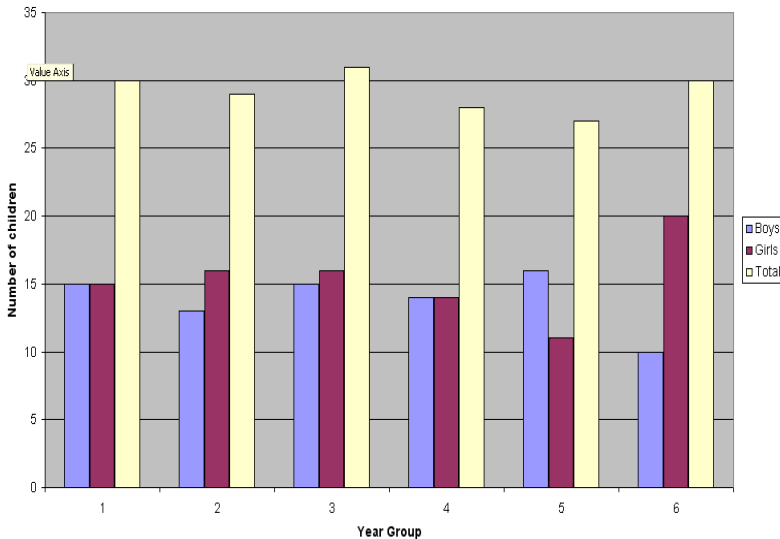
It is easy - but watch carefully!!


	A	B	C	D
1	YEAR GROUP	BOYS	GIRLS	TOTAL
2				
3	1			
4	2			
5	3			
6	4			
7	5			
8	6			
9				
10	TOTAL			
11				

Task four

	A	B	C	D
1	YEAR GROUP	BOYS	GIRLS	TOTAL
2				
3	1	15	15	30
4	2	13	16	29
5	3	15	16	31
6	4	14	14	28
7	5	16	11	27
8	6	10	20	30
9				
10				
11	TOTAL	83	92	175
12				

Graph to show number of children in each class

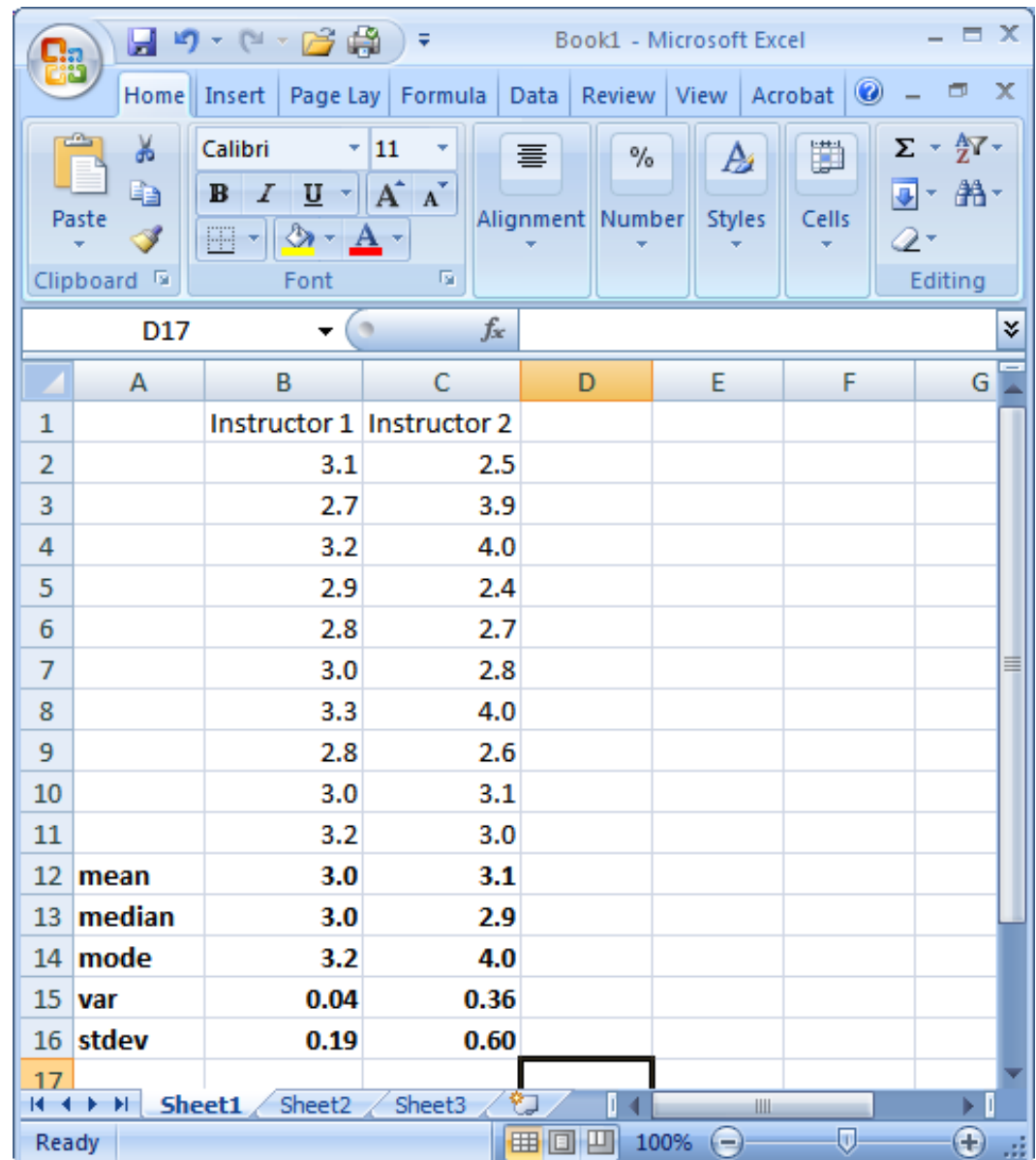


1. Select the data you are going to graph.
2. Now you are ready to create a graph! Find this icon and click it. 
3. Select the column graph and click NEXT.
4. Click the 'Series' tab. Change each series, Series 1 - Boys, Series 2 - Girls, and Series 3 - Total.
5. Click NEXT and give your chart a title and label the X and Y axis.
6. Click NEXT, choose to place chart in new sheet, then click FINISH.
7. Now you can print your chart.

It's that simple!!

Functions for Descriptive Statistics

- Your Excel spreadsheet should now look like this:



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G
1		Instructor 1	Instructor 2				
2		3.1	2.5				
3		2.7	3.9				
4		3.2	4.0				
5		2.9	2.4				
6		2.8	2.7				
7		3.0	2.8				
8		3.3	4.0				
9		2.8	2.6				
10		3.0	3.1				
11		3.2	3.0				
12	mean	3.0	3.1				
13	median	3.0	2.9				
14	mode	3.2	4.0				
15	var	0.04	0.36				
16	stdev	0.19	0.60				
17							

Part 2: Correlations and Scatterplots

Correlations

A quick review:

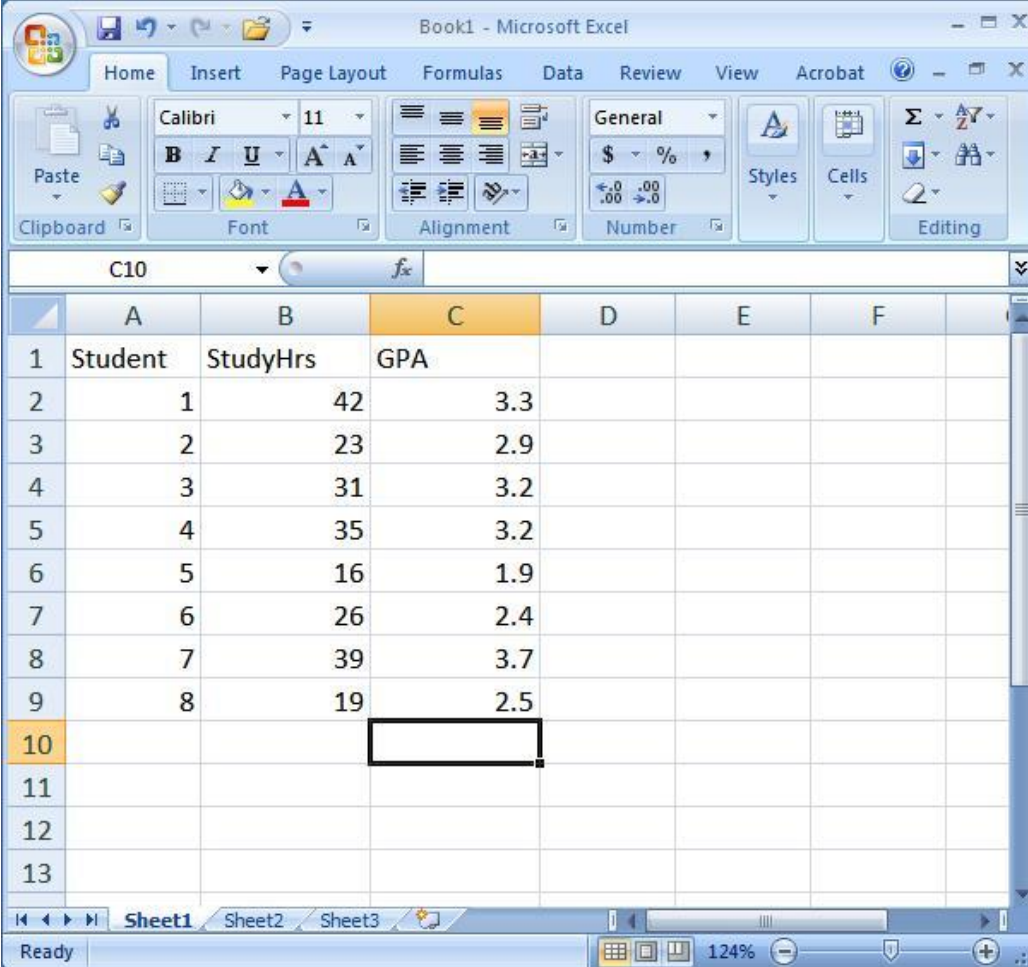
- Every correlation has a *direction* (positive or negative):
 - + correlation: **high** scores on one variable are associated with **high** scores on another variable.
 - - correlation: **high** scores on one variable are associated with **low** scores on the other variable.
- Every correlation has a *magnitude or strength*:
 - The closer the correlation coefficient is to +1.00 or -1.00, the **stronger** it is.
 - The closer the correlation coefficient is to 0.00, the **weaker** it is.

Calculating Pearson's r

- Correlations are described using the Pearson Product-Moment correlation statistic, or r value.
- In Excel, there are many functions that can calculate a correlation statistic, however, we will only use =PEARSON in this class.

Let's say we want to determine if there is a relationship between number of hours spent per week studying for Psych 209 and GPA earned in the class at the end of the quarter. To do so, we can calculate Pearson's r for our two variables.

Enter the following data into Excel:



The screenshot shows the Microsoft Excel interface with a spreadsheet containing the following data:

	A	B	C	D	E	F
1	Student	StudyHrs	GPA			
2	1	42	3.3			
3	2	23	2.9			
4	3	31	3.2			
5	4	35	3.2			
6	5	16	1.9			
7	6	26	2.4			
8	7	39	3.7			
9	8	19	2.5			
10						
11						
12						
13						

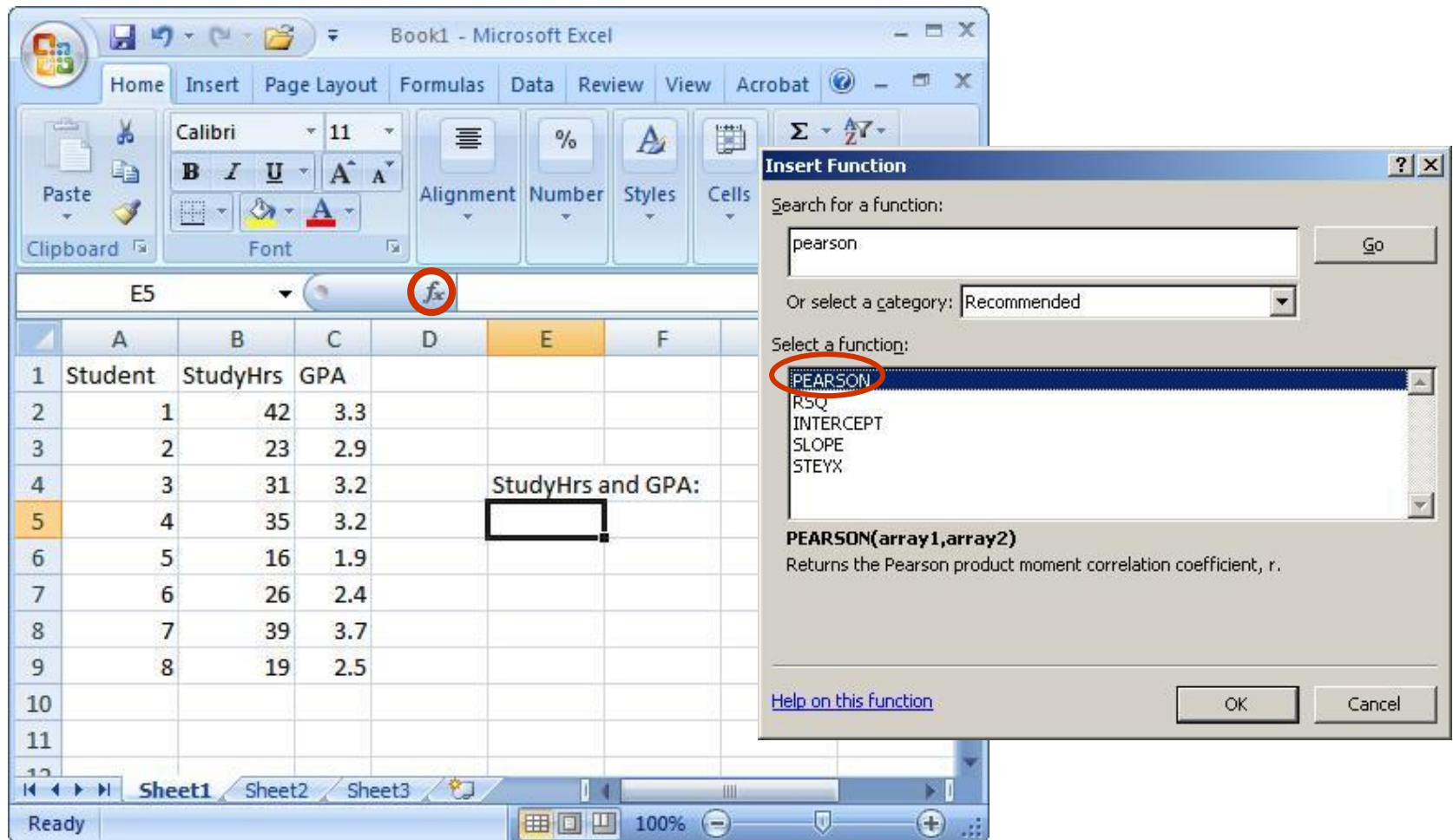
StudyHrs = average number of hours spent per week studying for 209

GPA = grade-point average earned in 209 at the end of the quarter

Step 1: Select the cell where you want your r value to appear (you might want to label it).

Step 2: Click on the function wizard  button.

Step 3: Search for and select PEARSON.



The screenshot shows the Microsoft Excel interface with the 'Insert Function' dialog box open. The dialog box is titled 'Insert Function' and has a search field containing 'pearson'. Below the search field, there is a dropdown menu for 'Or select a category:' set to 'Recommended'. A list of functions is displayed, with 'PEARSON' selected and circled in red. Other functions visible in the list include RSQ, INTERCEPT, SLOPE, and STEYX. Below the list, the function signature 'PEARSON(array1,array2)' is shown, along with a description: 'Returns the Pearson product moment correlation coefficient, r.' At the bottom of the dialog box, there are buttons for 'OK' and 'Cancel', and a link for 'Help on this function'.

	A	B	C	D	E	F
1	Student	StudyHrs	GPA			
2	1	42	3.3			
3	2	23	2.9			
4	3	31	3.2		StudyHrs and GPA:	
5	4	35	3.2			
6	5	16	1.9			
7	6	26	2.4			
8	7	39	3.7			
9	8	19	2.5			
10						
11						
12						

Step 4: For Array1, select all the values under StudyHrs.
For Array2, select all the values under GPA.

The screenshot shows Microsoft Excel with a data table and the Function Arguments dialog box for the PEARSON function. The data table is as follows:

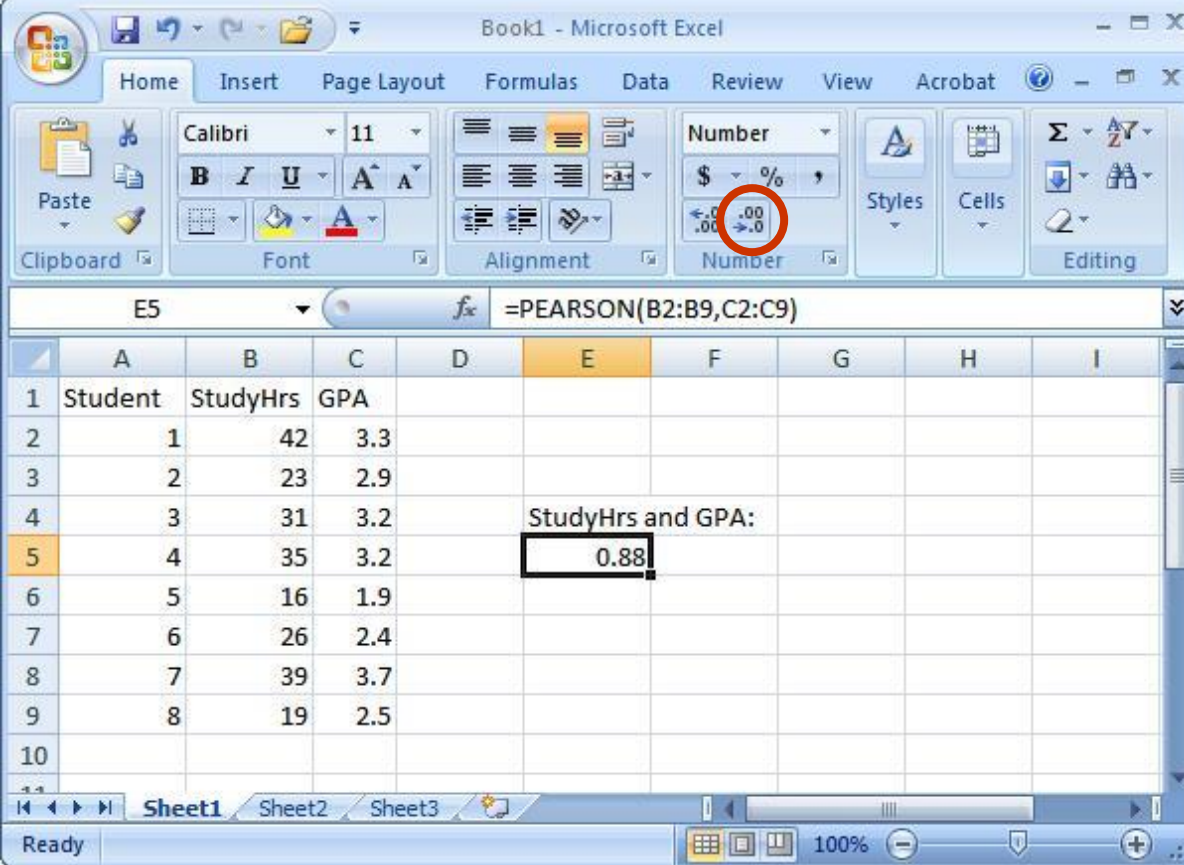
	A	B	C	D
1	Student	StudyHrs	GPA	
2	1	42	3.3	
3	2	23	2.9	
4	3	31	3.2	
5	4	35	3.2	
6	5	16	1.9	
7	6	26	2.4	
8	7	39	3.7	
9	8	19	2.5	

The Function Arguments dialog box for the PEARSON function is open, showing the following details:

- Function: PEARSON
- Array1: B2:B9 (circled in red)
- Array2: (empty)
- Formula result: =
- Help on this function: [Help on this function](#)
- Buttons: OK, Cancel

The dialog box also includes the following text: "Returns the Pearson product moment correlation coefficient, r." and "Array1 is a set of independent values."

Step 5: That's it! Once you have your r value, don't forget to round to 2 decimal places.



The screenshot shows the Microsoft Excel interface with the following details:

- File name: Book1 - Microsoft Excel
- Active tab: Home
- Font: Calibri, 11
- Number format: .00 (highlighted with a red circle)
- Formula bar: $=PEARSON(B2:B9,C2:C9)$
- Cell E5: 0.88 (highlighted with a black box)
- Worksheet: Sheet1

	A	B	C	D	E	F	G	H	I
1	Student	StudyHrs	GPA						
2	1	42	3.3						
3	2	23	2.9						
4	3	31	3.2						
5	4	35	3.2		0.88				
6	5	16	1.9						
7	6	26	2.4						
8	7	39	3.7						
9	8	19	2.5						
10									

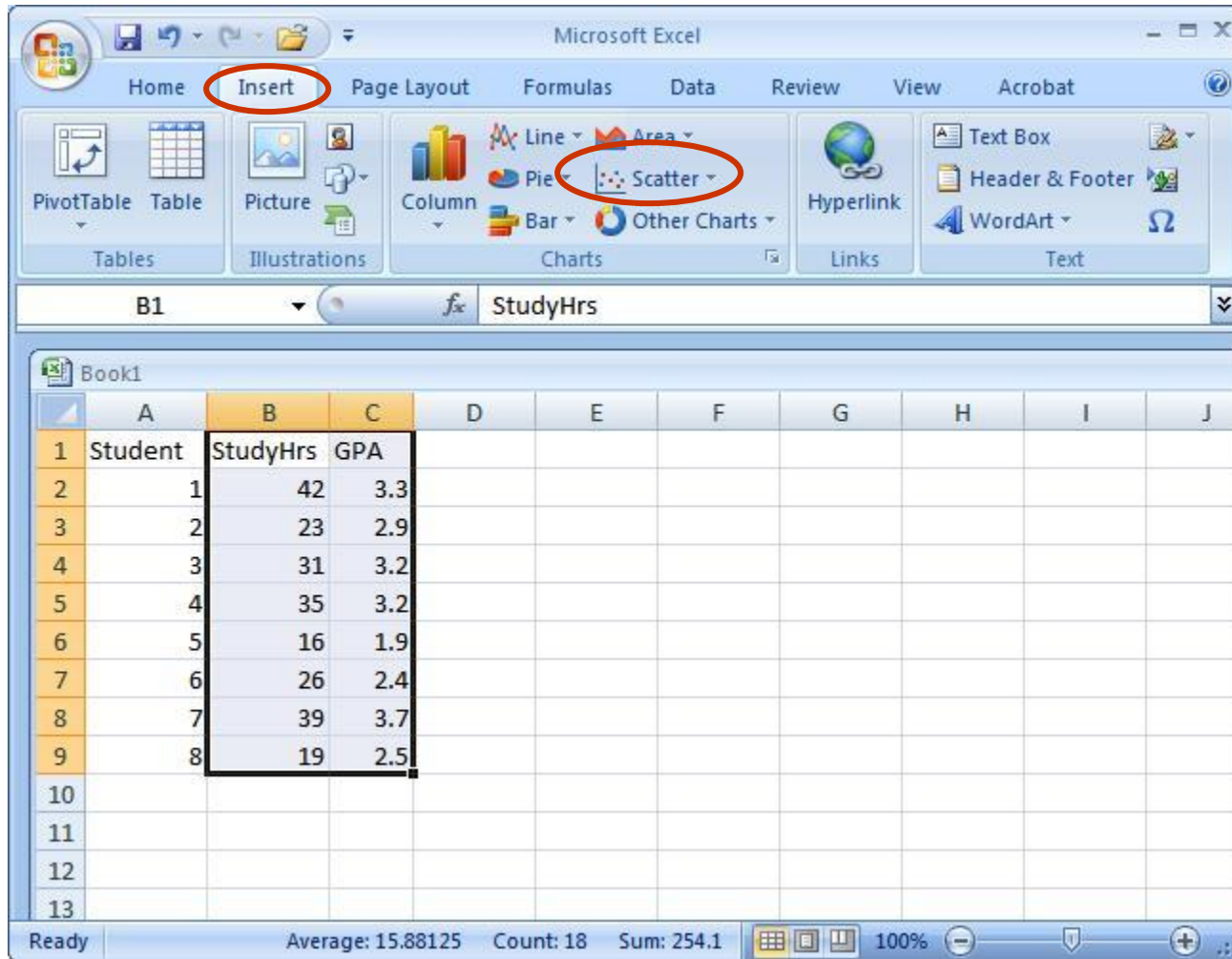
Knowledge check: What does the r value of 0.88 tell you about the strength and direction of the correlation between StudyHrs and GPA?

Scatterplots

- A scatterplot is an excellent way to visually display the relationship (correlation) between two variables.
- Each point on the scatterplot represents an individual's data on the two variables.
- We will now create a scatterplot for StudyHrs and GPA.

Step 1: Select both columns of variables you wish to plot (StudyHrs and GPA).

Step 2: Click on the tab labeled 'Insert', and then select 'Scatter' in the 'Charts' menu.

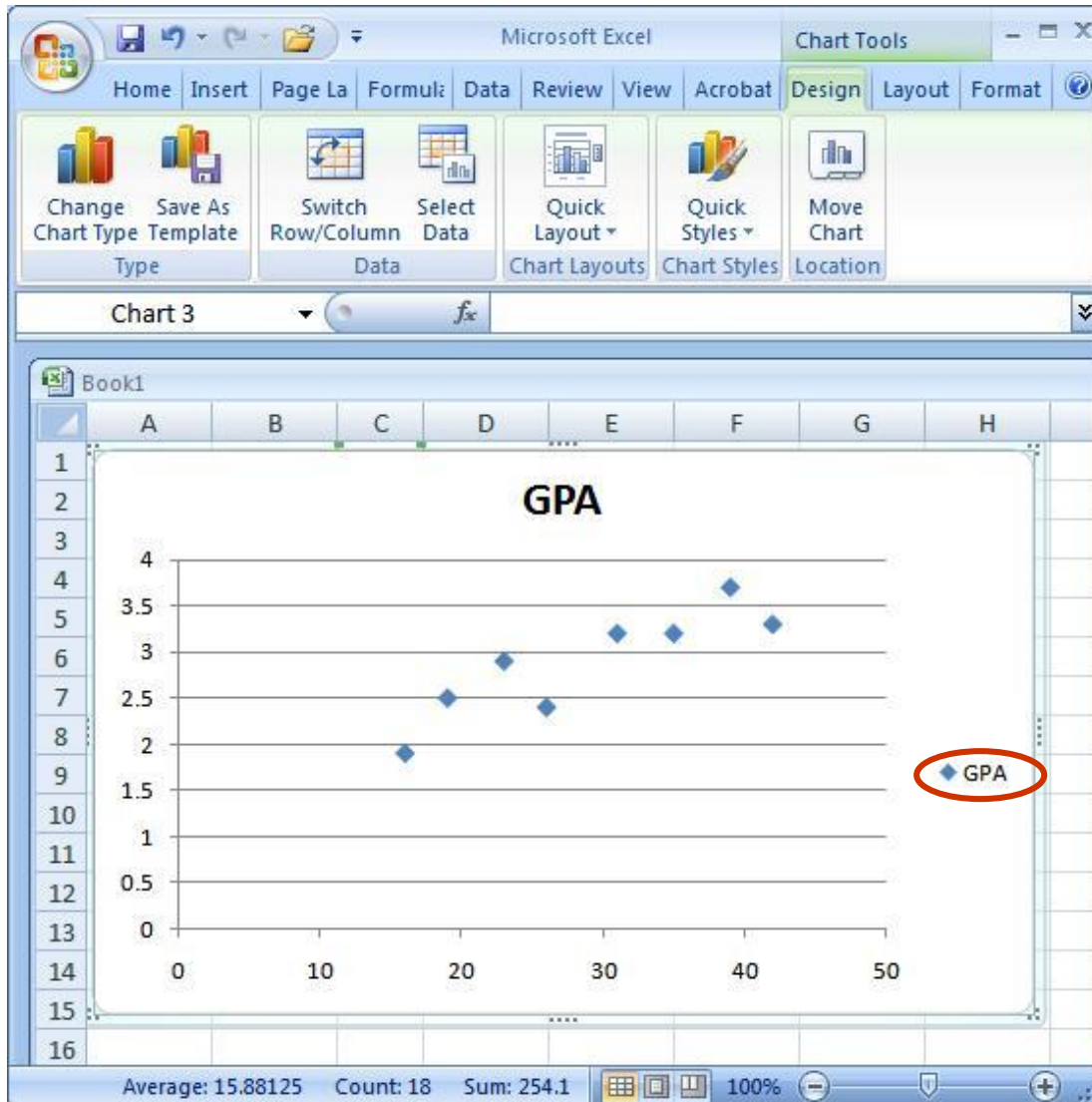


The screenshot shows the Microsoft Excel interface. The 'Insert' tab is selected and circled in red. Within the 'Charts' group, the 'Scatter' option is also circled in red. The spreadsheet below shows a table with columns labeled 'Student', 'StudyHrs', and 'GPA'. The data is as follows:

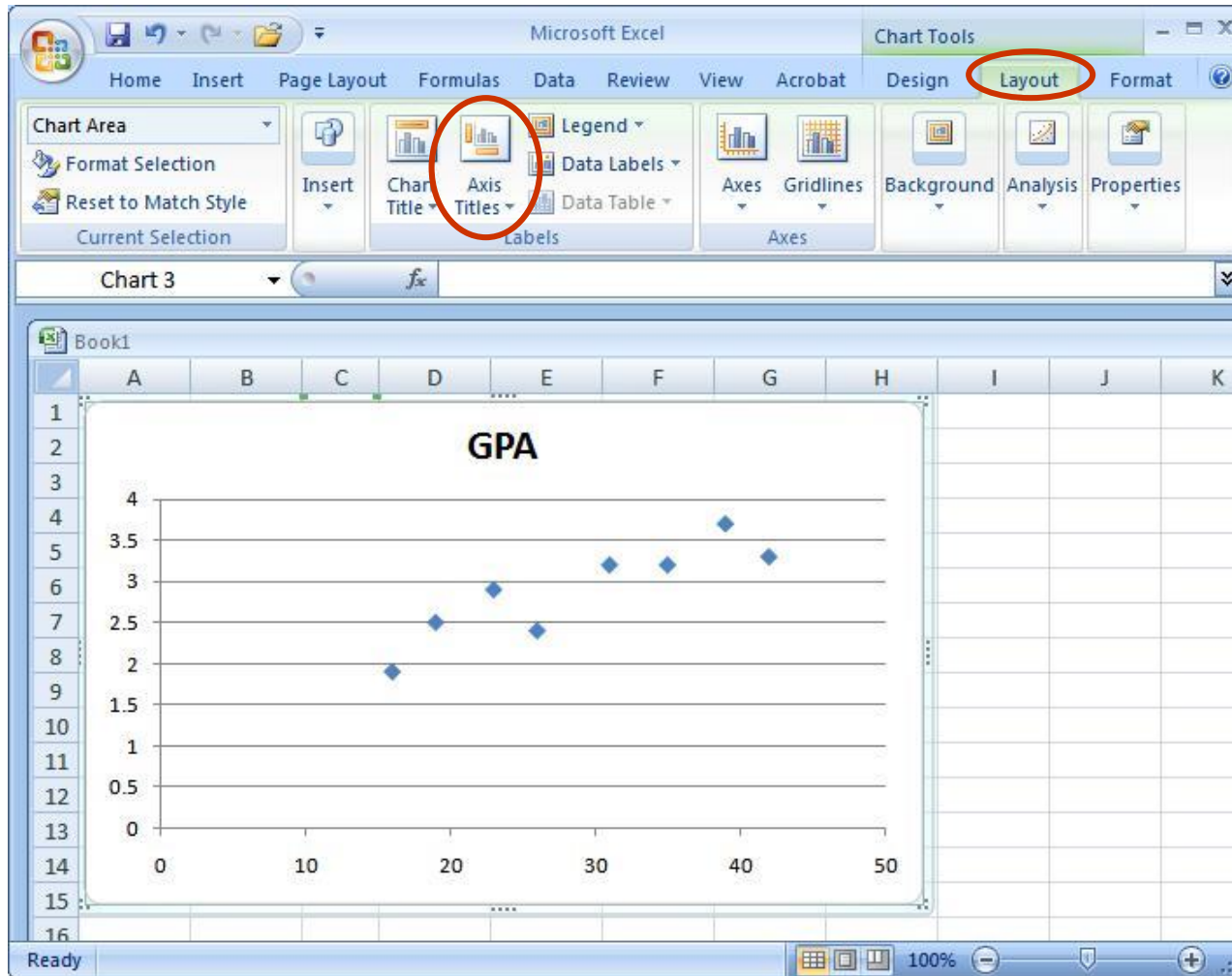
	A	B	C	D	E	F	G	H	I	J
1	Student	StudyHrs	GPA							
2	1	42	3.3							
3	2	23	2.9							
4	3	31	3.2							
5	4	35	3.2							
6	5	16	1.9							
7	6	26	2.4							
8	7	39	3.7							
9	8	19	2.5							
10										
11										
12										
13										

The status bar at the bottom shows: Ready, Average: 15.88125, Count: 18, Sum: 254.1, 100% zoom.

Step 4: Remove the legend by clicking on it and pressing Delete.



Step 5: Add axis titles by selecting the 'Layout' tab and clicking on 'Axis Titles.' For the horizontal title, you want it below the x-axis. For the vertical title, you want the 'Rotated Title' option.

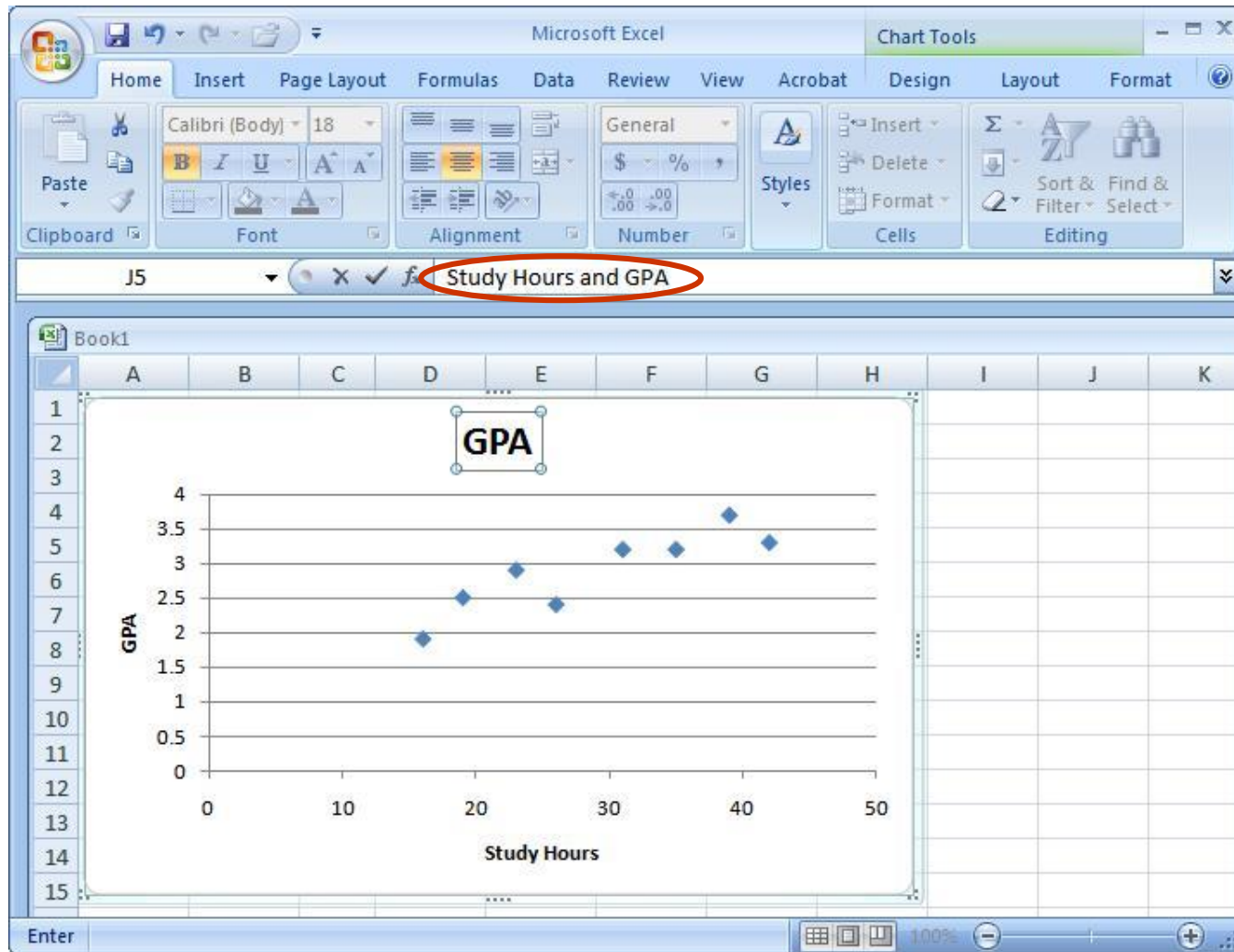


NOTE: Your chart must be highlighted for the 'Layout' tab to appear under 'Chart Tools.'

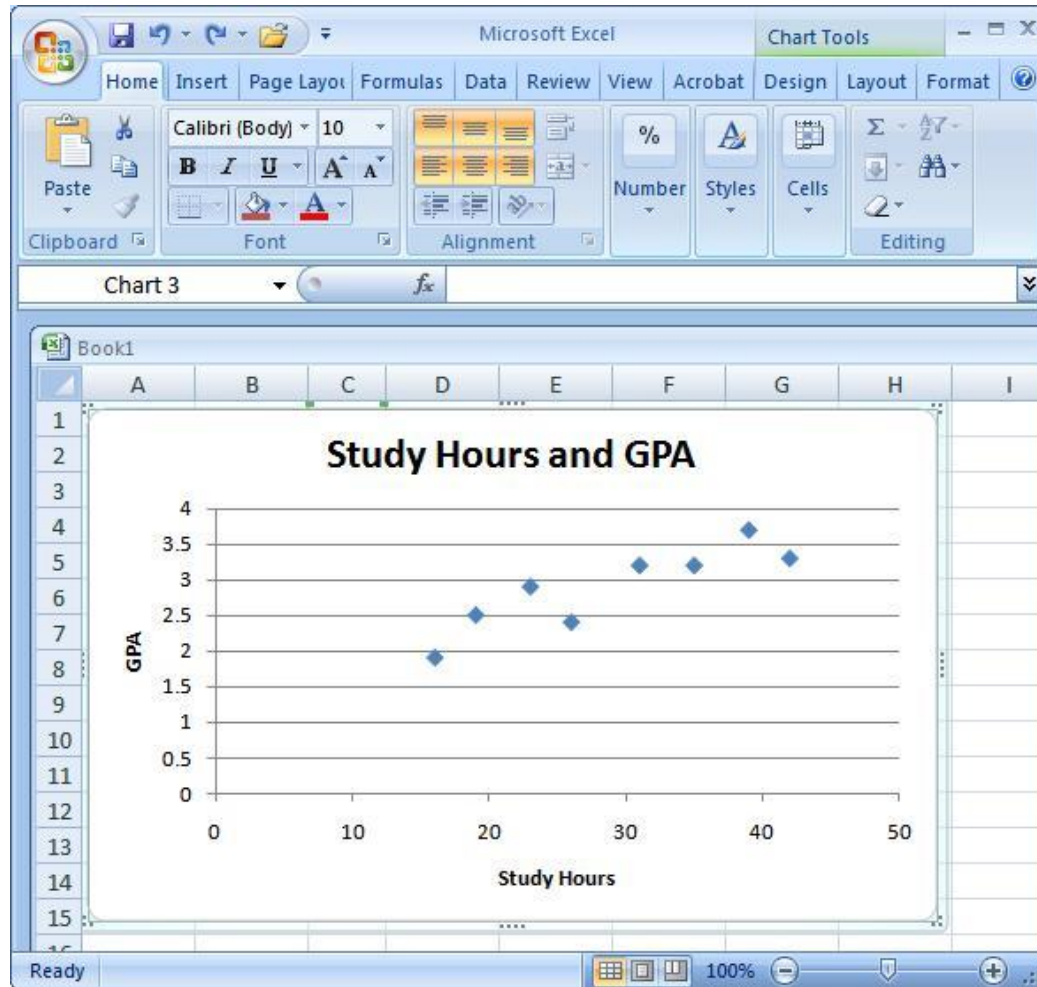
A note about x- and y-axes:

- For scatterplots, it does not matter which variable goes on each axis (this is NOT true for other types of charts).
- However, you need to make sure you label your axes with the proper variable name.
- In this example, GPA is on the y-axis and Study Hours is on the x-axis (we can tell this based on their different ranges of values).
- As a helpful hint, Excel will automatically put the first variable (left-hand column) on the x-axis, and the second variable (right-hand column) on the y-axis.

Step 6: Change the chart title by selecting it, typing a new one, and pressing Enter. Chart and axis titles may be altered by right-clicking on them.



Your scatterplot is now finished!



Remember: Each point in the scatterplot represents an individual's data.

Knowledge check: Identify Student 8 in the scatterplot.

Describing Correlations and Scatterplots

- Scatterplots and correlations are described:
 - As positive or negative.
 - As weak, moderate, or strong.
 - Using the r value.
 - Sentence 1: There is a strong, positive correlation ($r = 0.88$) between the number of hours studied and GPA.
- Then you want to describe the general relationship between the two variables:
 - Sentence 2: More hours of studying for Psych 209 was associated with a higher GPA earned in the class at the end of the quarter.
- NOTE: We cannot say “More studying led to a higher GPA” – this implies *causation*, which **cannot** be determined using correlational research.