INTERMEDIATE EXCEL

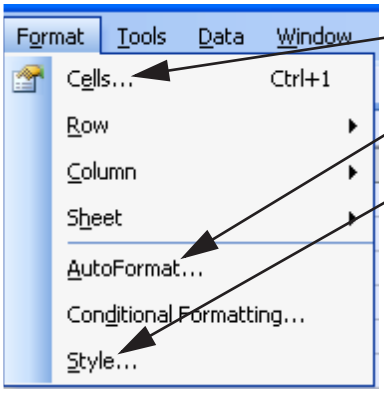
This document is for those who already know the basics of spreadsheets and have worked with either Excel for Windows or Excel for Macintosh. Many features of Excel are the similar on both platforms, so this document will be helpful for both. This document also describes how to execute procedures in Excel by using the pull-down menus. As you gain experience with the program, you will become familiar with the short-cut buttons as well.

By "Intermediate Excel," we refer to those functions of Excel that go beyond simple spreadsheet creation and management. Microsoft has built into this package many advanced capabilities that greatly enhance your ability to input, manipulate, and present information. If you have basic questions that are not covered here, please see the "Introduction to Excel" document. We also recommend that you become familiar with Excel's on-line help feature.

Formatting your Spreadsheet

While the basic formatting commands were covered in "Introduction to Excel," the software does allow you to go significantly further in making your spreadsheet more attractive and easier to read.

Using the Format Menu:



Format Cells - Allows you to control cell borders, colors, patterns, number format, alignment, and data protection.

AutoFormat - Allows you to select a pre-designed format for tables.

Style - Allows you to specify a combination of formats for selected cells.

Format Cells - Borders

Borders are useful for drawing attention to a particular row, column or cell. Borders are different from gridlines. By hiding the gridlines, you can make borders and shading stand out on your screen.

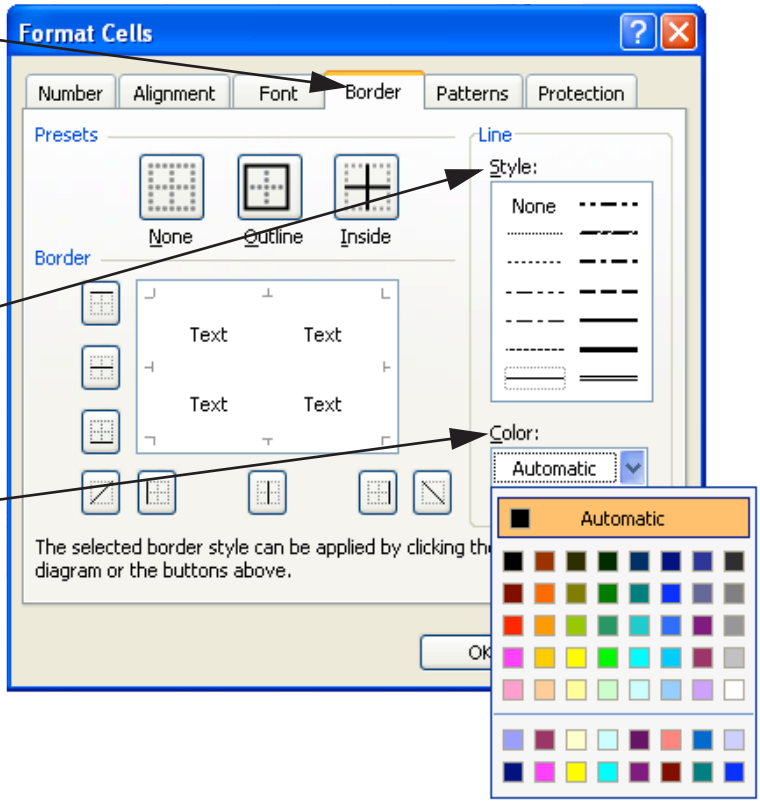
Note: Adjoining cells share borders. Therefore, putting a bottom border on one cell produces the same effect as putting a top border on the cell below it.

To make borders, select the cells for which you would like a border. Choose the Format Cells command, and then click on the tab called Border. Excel displays these options:

Border - Allows you to select whether you want an outline border (a box) or simply a line on one or several sides.

Style -Allows you to select line type -thin, thick double, etc.

Color - Sets the border line color. (In most cases, it is best to use the default black).

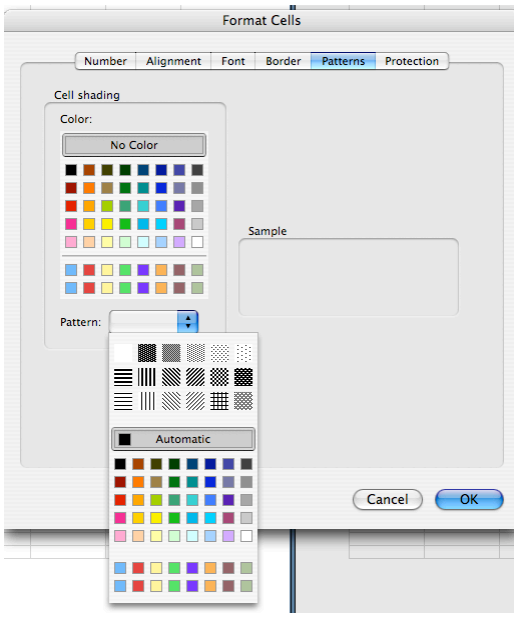


Format Cells - Patterns

This command is also useful for calling attention to a particular row, column, or cell. Shading the cell in gray or even in zebra stripes makes is much more noticeable.

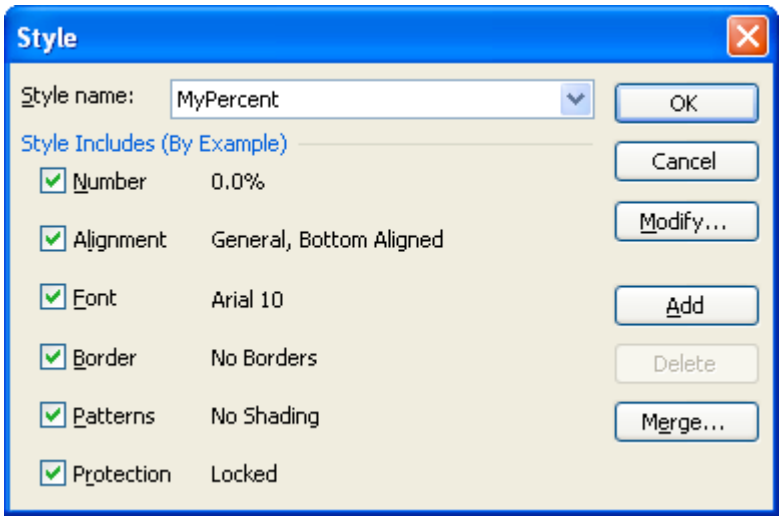
Warning: Make sure that your pattern does not make it difficult to read the content of your spreadsheet. Go to Tools/Options/Select View Tab and uncheck Grid View.

To add a pattern, select the cells for which you would like the pattern to fill. Choose the Format Cells command, and then click on the tab called Patterns. Excel displays these options:



Format/Style

With Excel, you can specify a combination of formats for a cell and then apply them all at once. After you have a style defined, you can use it in many different places within your worksheet. To define a style, select the Format Style command. A dialog box will pop up which looks like this:



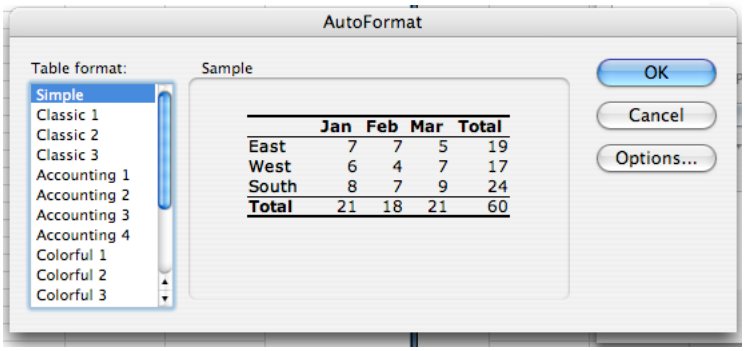
Give the new style the name of your choice, then click Modify. You will have the opportunity to format the number appearance, alignment, font, borders, patterns, and protection for your new style. When the style has been designed to your satisfaction, click on Add.

From now on, you can use your new style to apply your formatting choices to any part of your spreadsheet. Simply select the desired cells, then go to Format Style, and use the pull down arrow to find the name of your new style on the list. There are a number of default styles included in Excel to facilitate spreadsheet formatting, such as Percent and Currency.

Once styles have been created in one workbook, you can easily bring them into another. Let's say, for example, that you had created several styles in workbook BUDGET that you would also like to use in workbook FINANCE. Open both workbooks. While in FINANCE, use the Format Style command, then click on Merge. Select BUDGET from the list of available workbook names, and the BUDGET style names are now part of the FINANCE workbook.

Autoformat

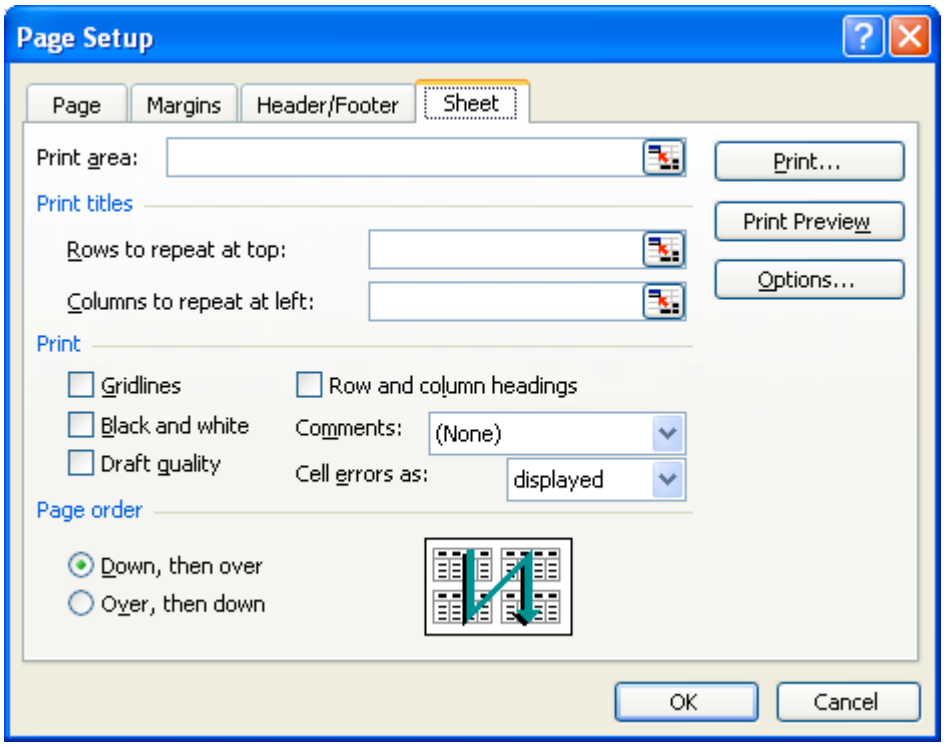
Excel has created several general formats to apply to a table of cells. Select all the cells in your table, then choose Format Autoformat. A menu will pop up that looks like the following:



On the left is a list of common table formats. As you select each format style, an example will be displayed in the center. (In the above demonstration, the "Simple" format is being displayed). When you find the style you prefer, click on the OK button.

File/Page Setup

Many of the commands used to format your worksheet are actually accessed by going to the File menu, Page Setup command. You will see a dialog box that looks like the one below. Notice that you alter four different aspects of the page setup by clicking on the Page, Margins, Header/Footer or Sheet tabs.



Group Editing of Worksheets

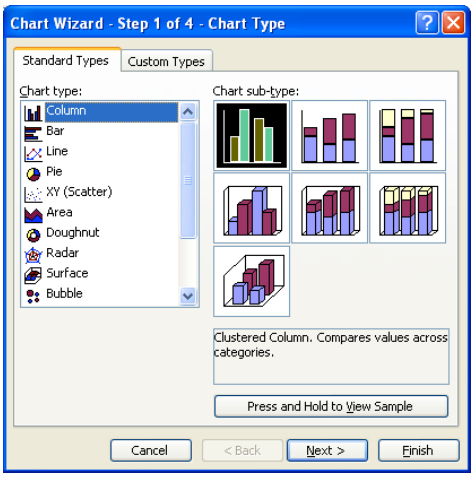
It is possible to perform the same formatting or editing operations on more than one worksheet at a time. First, select the worksheets that you want to edit by clicking on their worksheet tabs, located at the bottom of the screen. Hold down the CTRL key to allow you to select more than one at a time. You will notice that the word "Group" appears in the Title Bar at the top of the spreadsheet. You can then proceed to format or edit as you wish-the changes will appear on every worksheet you have selected.

When you are finished editing, you must "ungroup" the worksheets. You can do this by clicking on one of their tabs with the right mouse button. On the menu that pops up, choose Ungroup Sheets. This will allow you to again work with the worksheets one by one.

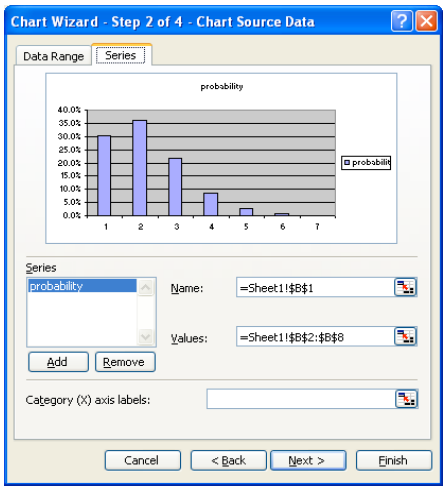
The Basics of Graphs and Charts

Before starting this section, you should be aware of the fact that the graph and chart capabilities of Excel are extensive, and could almost be considered an entirely separate program. This handout will get you started. You can then refer to Excel's on-line help or manuals for more information.

To make a chart, you first need a spreadsheet and data. Put your data into the spreadsheet, then highlight the cells containing the information that you would like to see presented graphically. At this point, you can begin to use the Chart Wizard.



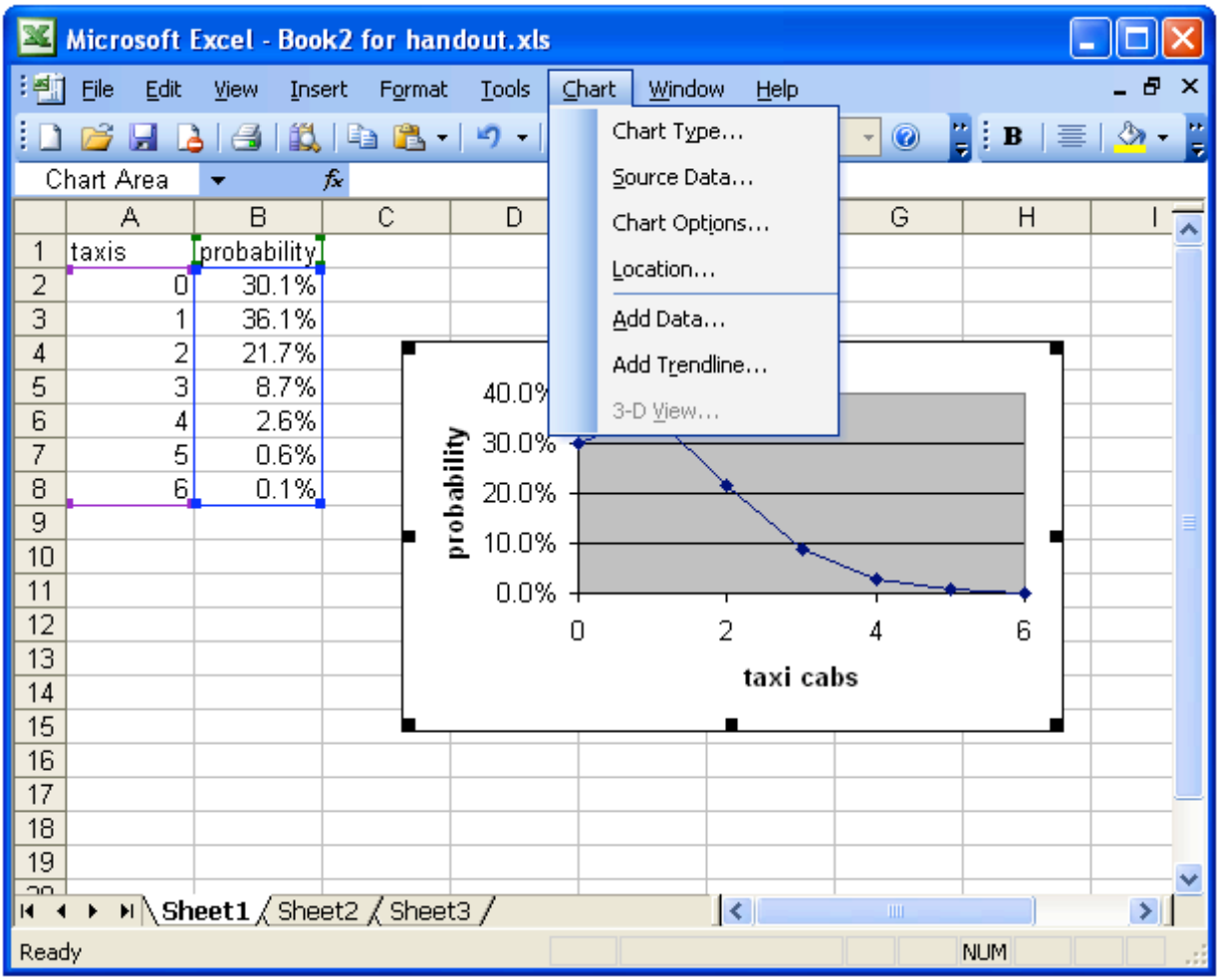
Your cursor will now be transformed from an arrow to a cross-hair. Click and drag this cross-hair across your worksheet; it's not particularly important where, because you can change the chart location later. When you let go, a dialog box will come up which will lead you through the process of creating a chart. Follow the instructions and click on Next when you have completed each step. As an example, Step 2 of the process is shown below.



When you reach Steps 4 and 5, you will see a small display on the left that is meant to serve as a sample. don't worry if it doesn't look exactly like your envisioned final chart.

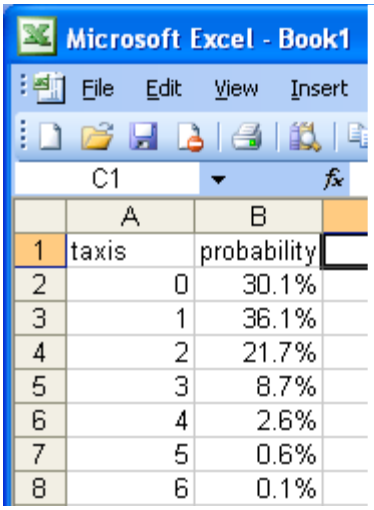
After the last step, you will be returned to your worksheet. The chart will be located in a small graphic box on the spreadsheet, where your crosshair cursor is.

To edit your chart, double click on it. You will then enter the Excel Charting module. You can note this change because the pull-down menus will be different. The Insert and Format menus in this Module contain functions that are especially useful for editing. Your screen will now look something like this:

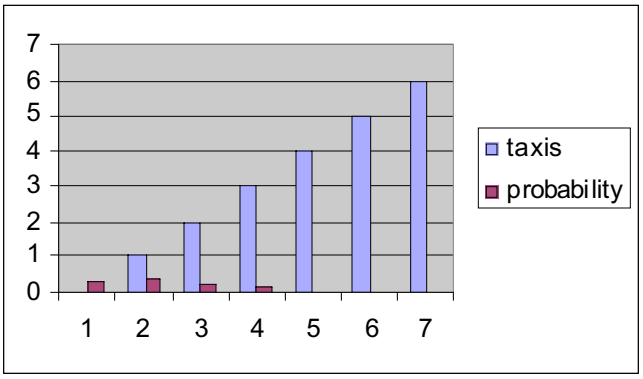


Useful Things To Know About Charts

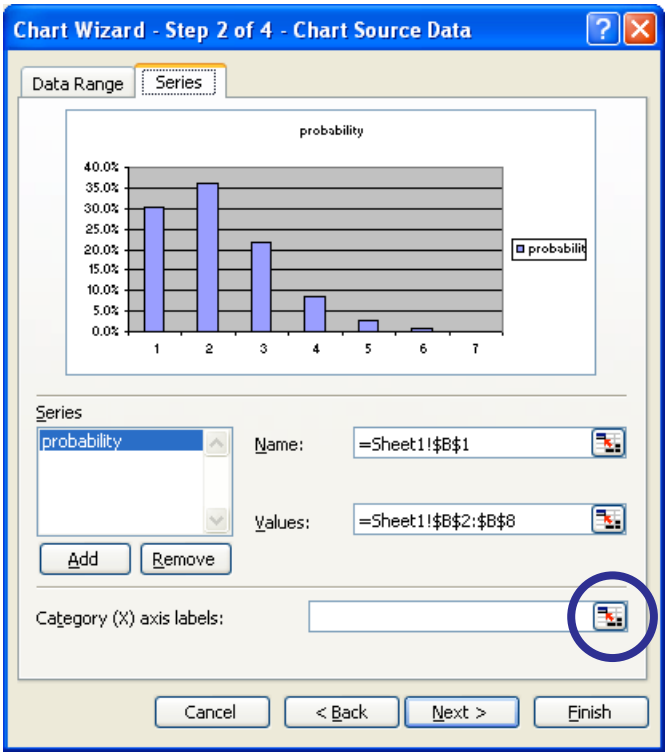
Excel prepares nice graphs, but to get them to appear as you want requires a little finesse. Here are two common tricks. First, you have a relative frequency table for the number of taxi-cabs that stop by a hotel in a five- minute interval. The first column lists the number of taxis, and the second column lists the probability of exactly that many taxis stopping outside the hotel in a five minute interval. You want to make a bar graph to represent this. If you select both columns of data, you will obtain the following graph:



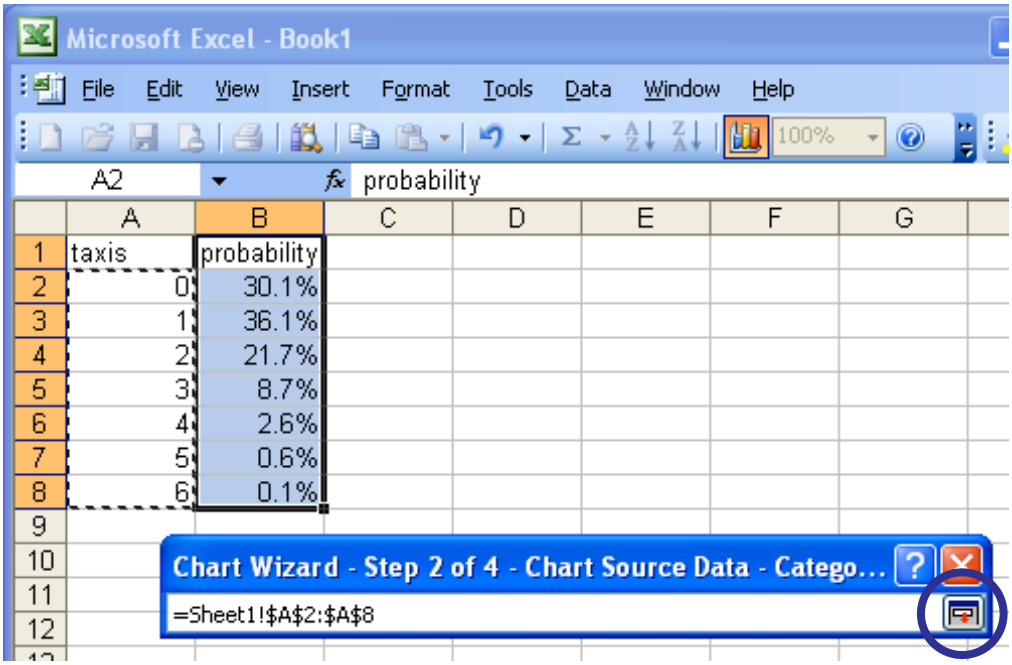
	A	B
1	taxis	probability
2	0	30.1%
3	1	36.1%
4	2	21.7%
5	3	8.7%
6	4	2.6%
7	5	0.6%
8	6	0.1%



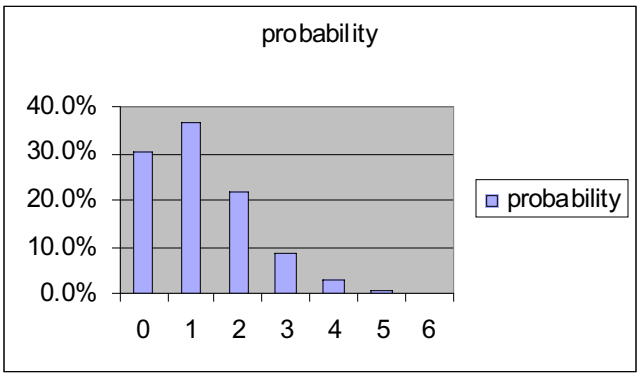
What has happened is Excel thought that the values in the first column were also data values. It did not recognize that they were labels for the x-axis. There are two ways to fix this, the first is to enter the values in the first column as text (before typing the number, type a single-quotation mark " ' "). Or, start by selecting only the second column, then hit the chart-wizard button. At the second step of the chart wizard, move to the "Series" tab at the top of the dialogue box.



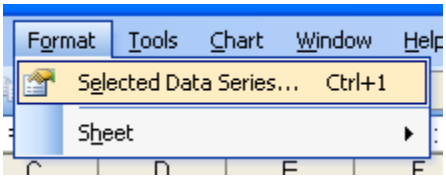
Click on the "Category (X) axis labels:" return-to-data-button (highlighted in dialogue box). Now, selected the first column (excluding the header row) and hit the return-to-dialogue-box-button:

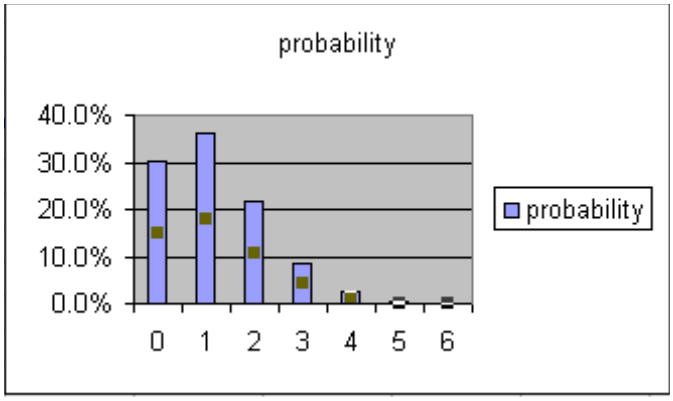


Now, the graph appears as expected, with the appropriate labels:

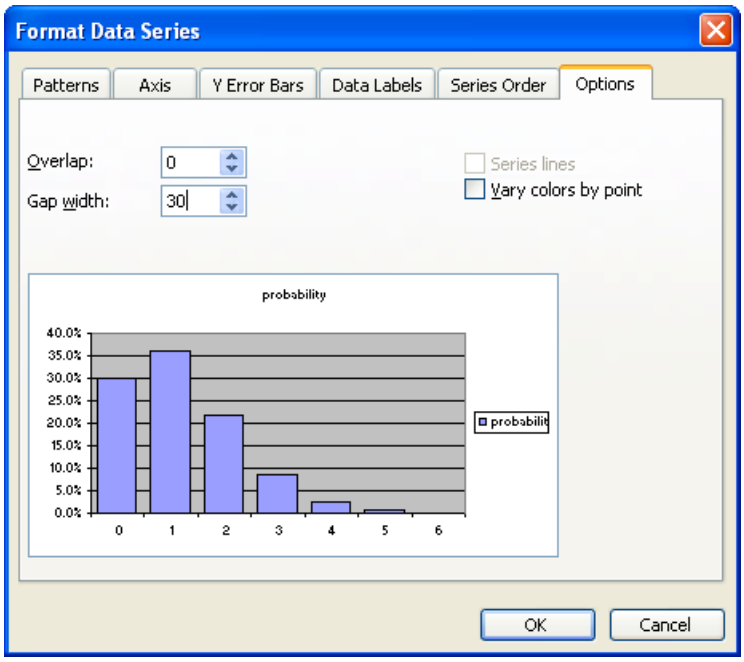


Another hidden feature in Excel's graphing capability is changing the width of the bars. First, select the chart by clicking the mouse anywhere on the chart. Next, click any one of the bars to select the data series represented by the bars. (See example to the right.) Next, from the file menu choose: Format > Selected Data Series…

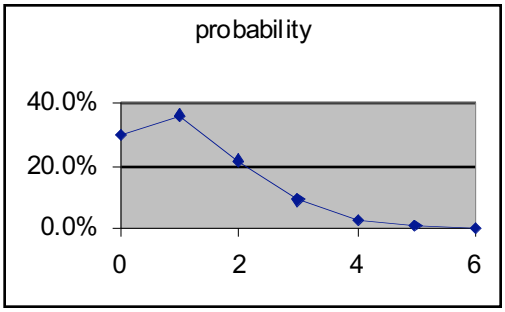




This will bring up a dialogue box. Select the last tab ("Options") at the top of the box. In this box, you can change Gap width (represented as a percentage of the size of the bar). In this case, the gaps are set to be 30% of the width of the bars. The resulting graph is pre-displayed in the dialogue box.



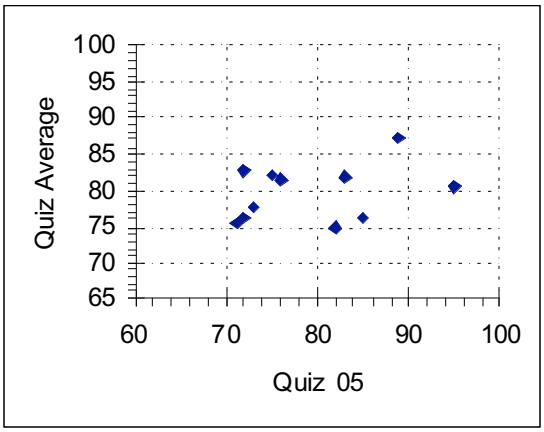
If you want a polygonal line graph for this data, you need not worry about these issues. Instead, selected the entire data set, click the chart wizard button, and select XY (Scatter) Chart Type. Here, Excel knows to interpret the first column as values for the x-axis.



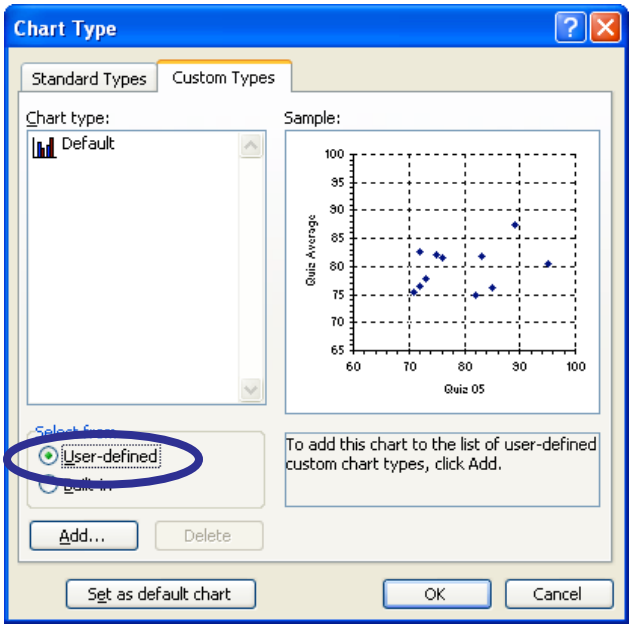
User-Defined Custom Chart Types

If you just spent an hour getting a chart to look a certain way, you may want to have Excel store the format in a user-defined custom chart type. For example, the chart to the right was obtained by specifying all of the following:

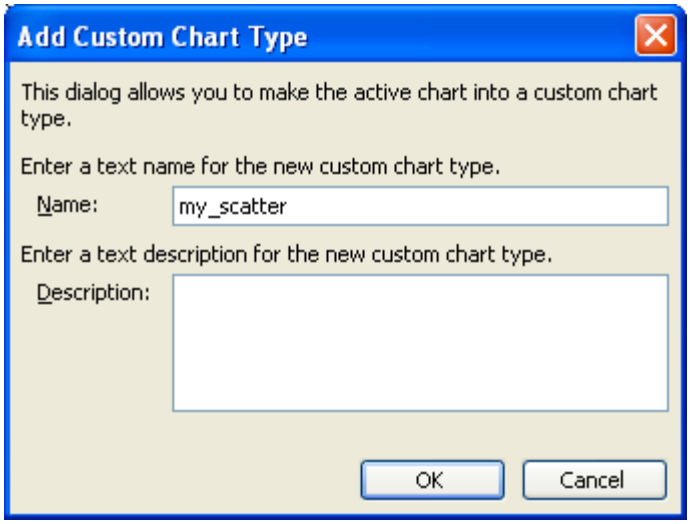
- titles for the axes were provided, title was turned off
- major gridlines were turned on
- the legend was turned off
- the shading and outline of the plot area were removed
- the style of the gridlines (both horizontal and vertical) was changed
- the weight of the axes, tick-marks and scales for the axes were changed
- the Auto scale feature for the font for the plot area was turned off



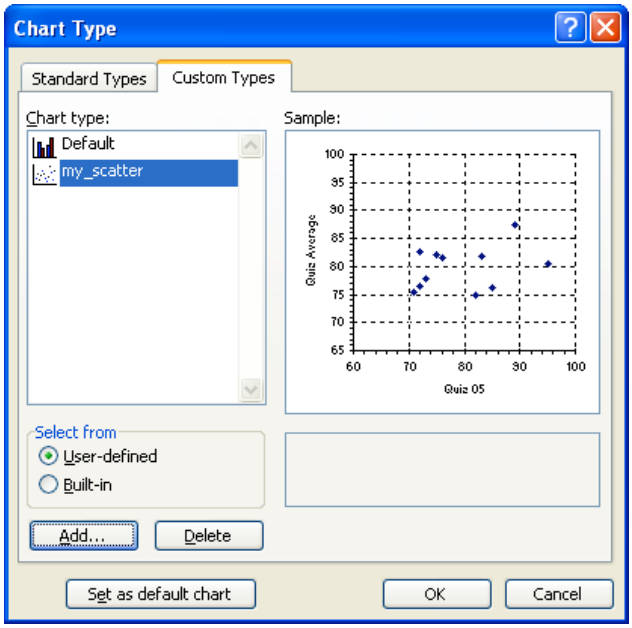
To get Excel to memorize this chart's formatting, click the mouse anywhere on the chart. From the menu bar, select Chart > Chart Type… and select the "Custom Types" tab at the top of the dialogue box. Click the User-defined radio-button (indicated on the diagram). Finally, hit the Add... button.



You will be asked to give the new chart type a name and an optional brief description.

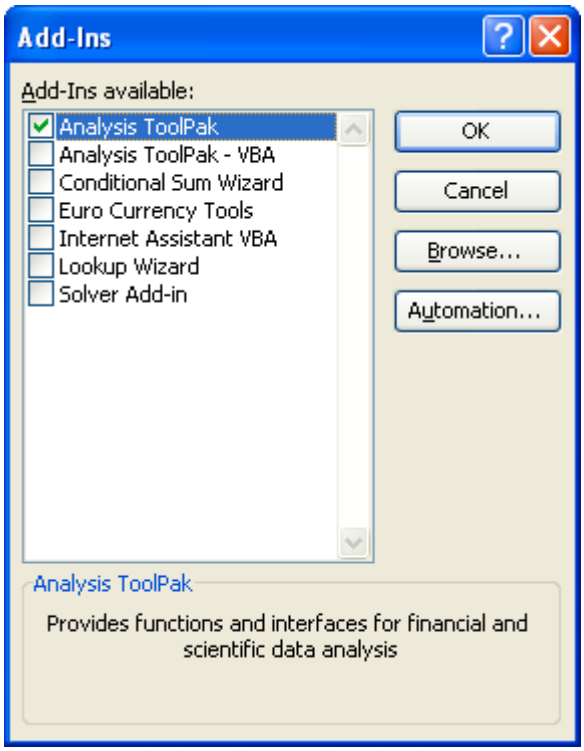


Now, this chart type will be available for future charts under the User-defined Custom Types:

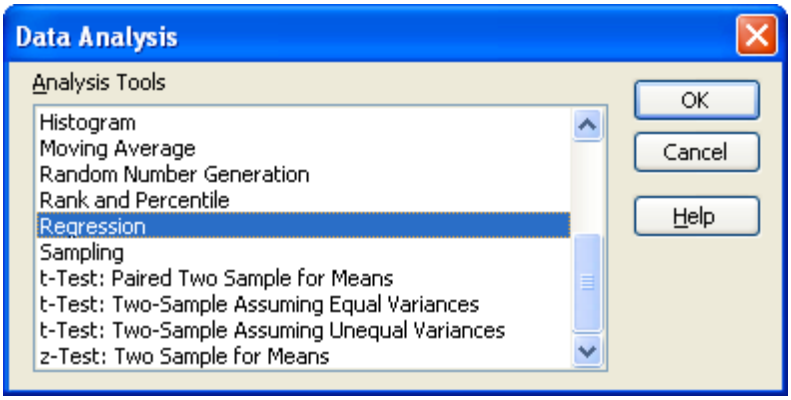


Analysis ToolPak

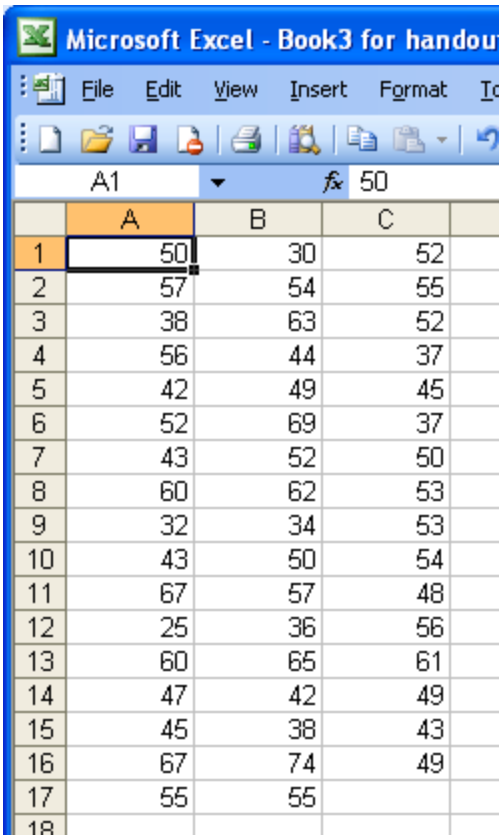
To use the Analysis ToolPak, the Add-In must be initialized. From the menu bar, select Tools > Add-Ins…. In the resulting dialogue box (to the right), check the Analysis ToolPak and hit OK.



Once this is done, a new option under the Tools menu bar is available: Data Analysis…. If you select this from the menu bar, you will see a dialogue box of various statistical procedures:

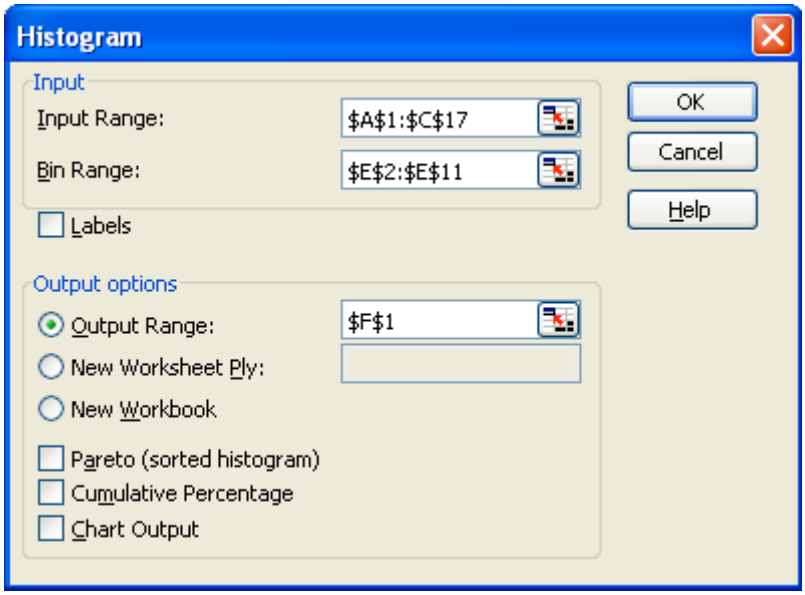


We will use the Analysis ToolPak to generate the frequency distribution table for these values (see right). First, we must tell Excel the largest value in each class. The values from 30 to 75 will be entered in $E$2:$E$11 counting by 5.

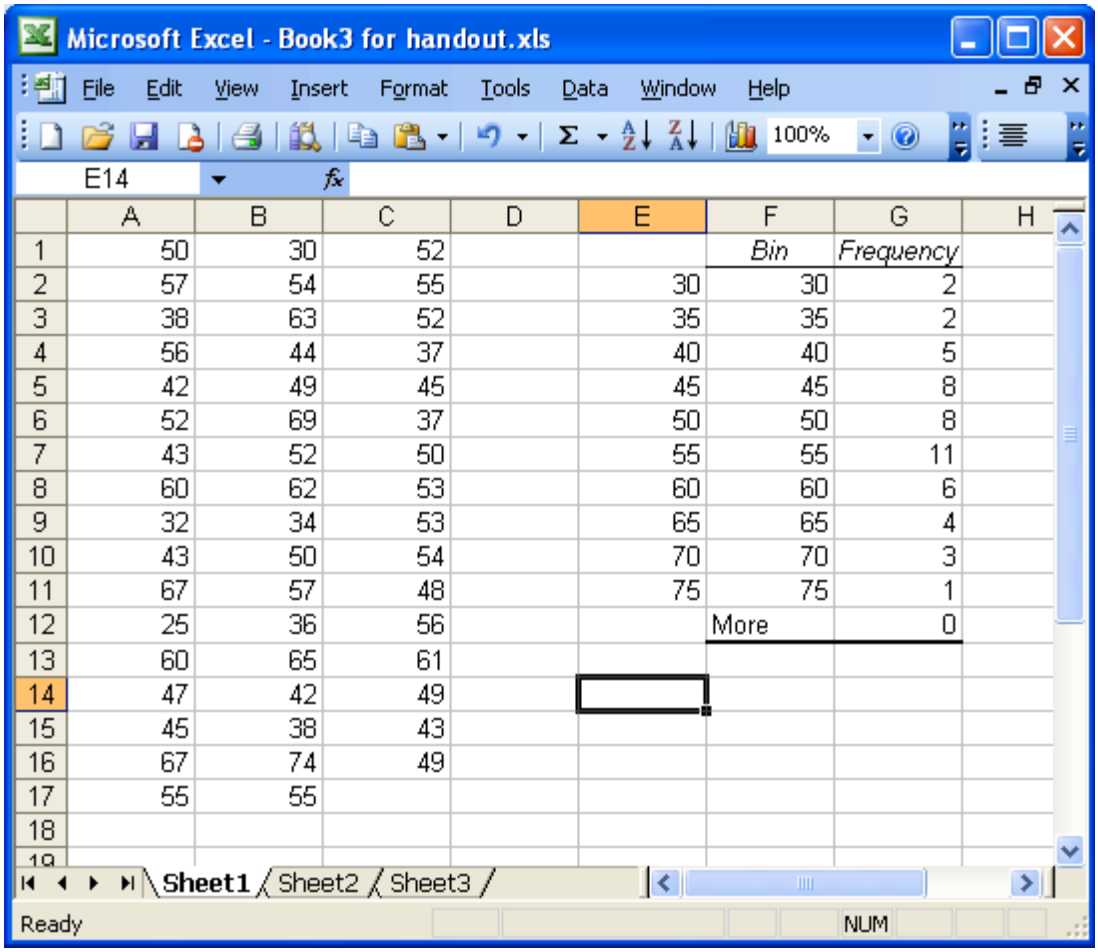


	A	B	C
1	50	30	52
2	57	54	55
3	38	63	52
4	56	44	37
5	42	49	45
6	52	69	37
7	43	52	50
8	60	62	53
9	32	34	53
10	43	50	54
11	67	57	48
12	25	36	56
13	60	65	61
14	47	42	49
15	45	38	43
16	67	74	49
17	55	55	

This means we want to know how many values in our data set are less than or equal to 30, greater than 30 but less than or equal to 35, etc. From the menu bar, select Tools > Data Analysis… > Histogram and hit OK.



The appropriate cells have been entered into the Input Range:, Bin Range:, and Output Range:. The input range is the collection of data. The bin range includes the largest values for each class. The output range is where Excel will put the frequency distribution table. Here is the final result.

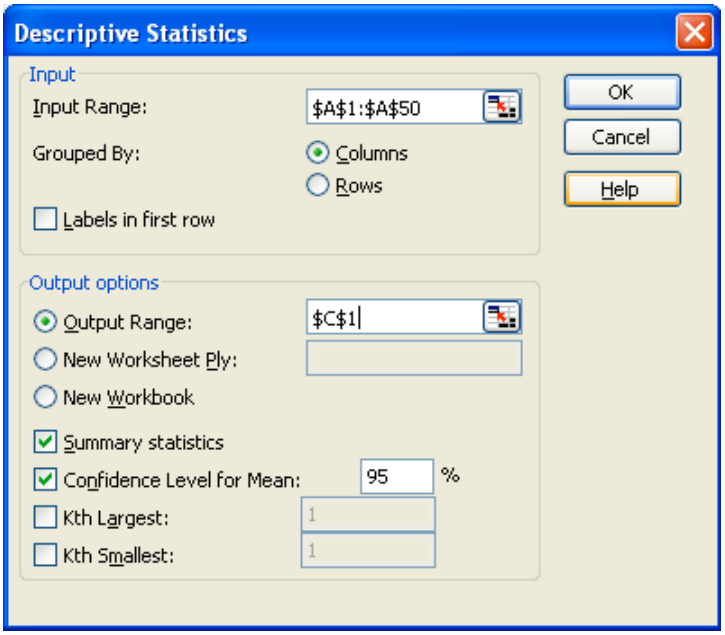


	A	B	C	D	E	F	G	H
1	50	30	52			*Bin*	*Frequency*	
2	57	54	55		30	30	2	
3	38	63	52		35	35	2	
4	56	44	37		40	40	5	
5	42	49	45		45	45	8	
6	52	69	37		50	50	8	
7	43	52	50		55	55	11	
8	60	62	53		60	60	6	
9	32	34	53		65	65	4	
10	43	50	54		70	70	3	
11	67	57	48		75	75	1	
12	25	36	56			More	0	
13	60	65	61					
14	47	42	49					
15	45	38	43					
16	67	74	49					
17	55	55						
18								

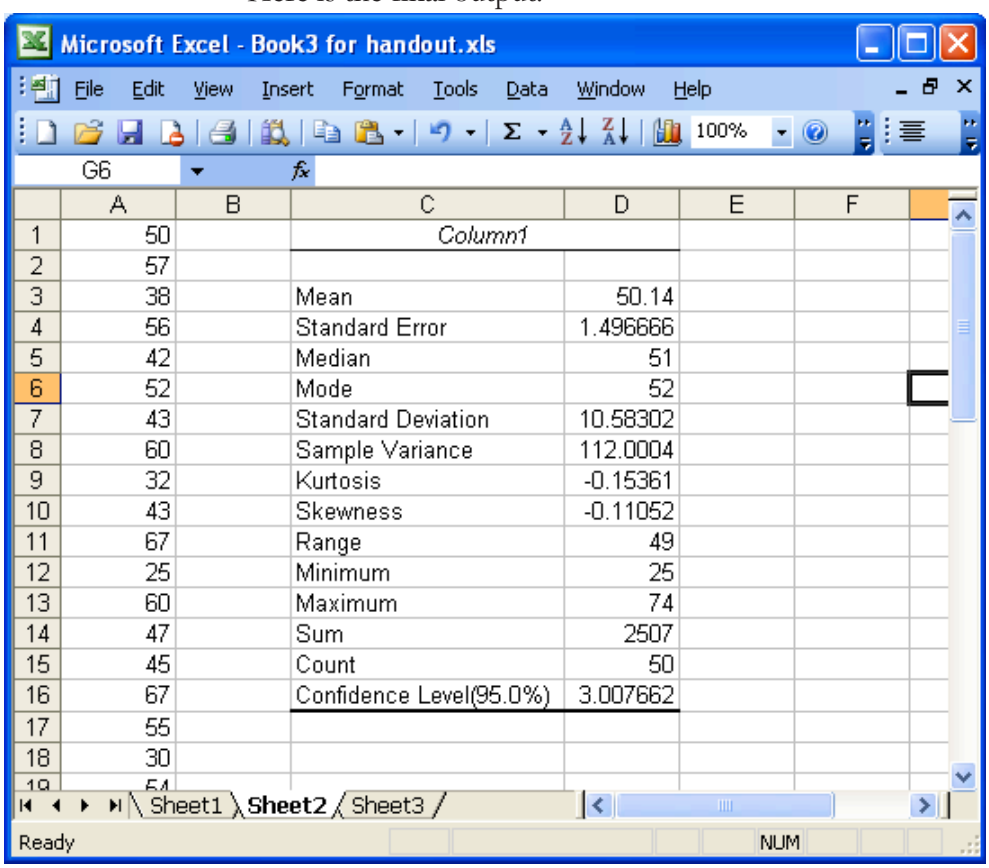
Note: If you want to draw the histogram/bar-chart for this data set, you will need to follow the directions from the Useful Things To Know About Charts section.

Remember: The histogram feature requires the user to provide the largest values in each class.

Next, we will use the Analysis ToolPak to generate the descriptive statistics for a data set along with the 95% confidence interval of the mean. The same data set as above has been copied to Sheet2, and entered into one column (A1:A50). From the menu bar, select Tools > Data Analysis… > Descriptive Statistics. The input range and output range have been entered into the dialogue box shown below. Also, the Summary statistics and Confidence Level for Mean have been checked.



Here is the final output.

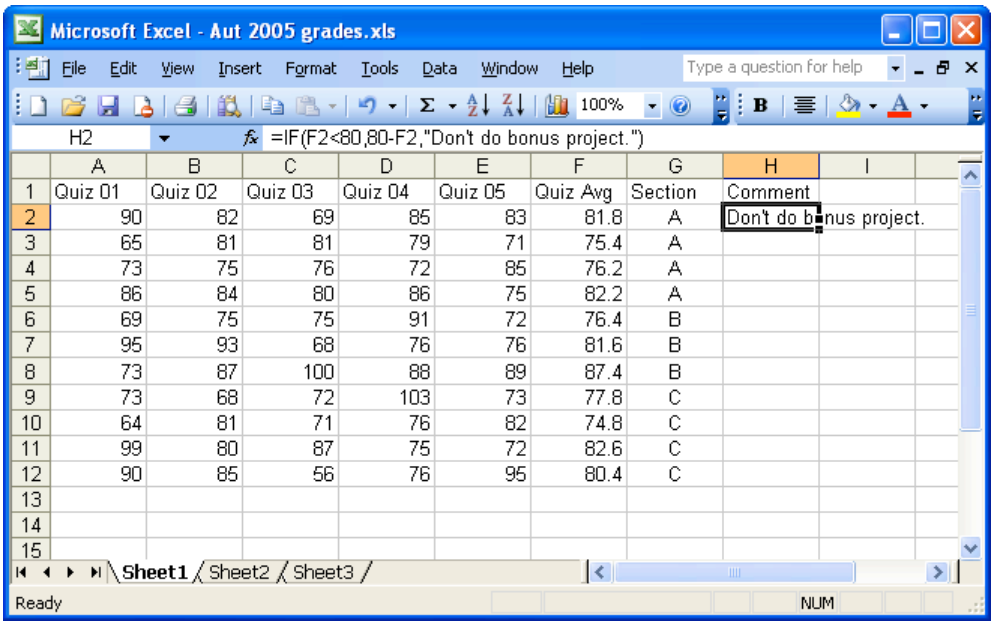


	A	B	C	D	E	F
1	50		*Column1*			
2	57					
3	38		Mean	50.14		
4	56		Standard Error	1.496666		
5	42		Median	51		
6	52		Mode	52		
7	43		Standard Deviation	10.58302		
8	60		Sample Variance	112.0004		
9	32		Kurtosis	-0.15361		
10	43		Skewness	-0.11052		
11	67		Range	49		
12	25		Minimum	25		
13	60		Maximum	74		
14	47		Sum	2507		
15	45		Count	50		
16	67		Confidence Level(95.0%)	3.007662		
17	55					
18	30					
19	54					

To calculate the confidence interval, add and subtract the confidence level (i.e., margin of error) from the mean.

The IF() Function

One of the most useful function in Excel is the IF() function. The following is a simple example demonstrating it. Our professor wants students that have an average below 80 to do a bonus project. Those students with an average above 80 are not to do the project. He also wants to tell the students how many points they will need to earn if the do need to do the project. The first IF() function is displayed in the screen capture below.



	A	B	C	D	E	F	G	H	I
1	Quiz 01	Quiz 02	Quiz 03	Quiz 04	Quiz 05	Quiz Avg	Section	Comment	
2	90	82	69	85	83	81.8	A	Don't do bonus project.	
3	65	81	81	79	71	75.4	A		
4	73	75	76	72	85	76.2	A		
5	86	84	80	86	75	82.2	A		
6	69	75	75	91	72	76.4	B		
7	95	93	68	76	76	81.6	B		
8	73	87	100	88	89	87.4	B		
9	73	68	72	103	73	77.8	C		
10	64	81	71	76	82	74.8	C		
11	99	80	87	75	72	82.6	C		
12	90	85	56	76	95	80.4	C		
13									
14									
15									

The IF() function has three arguments (each separated by a comma and contained in the parentheses):

=IF(F2<80,80-F2,"Don't do bonus project.")

The structure of this function is

If (test) is true, then do (first task), else do (second task).

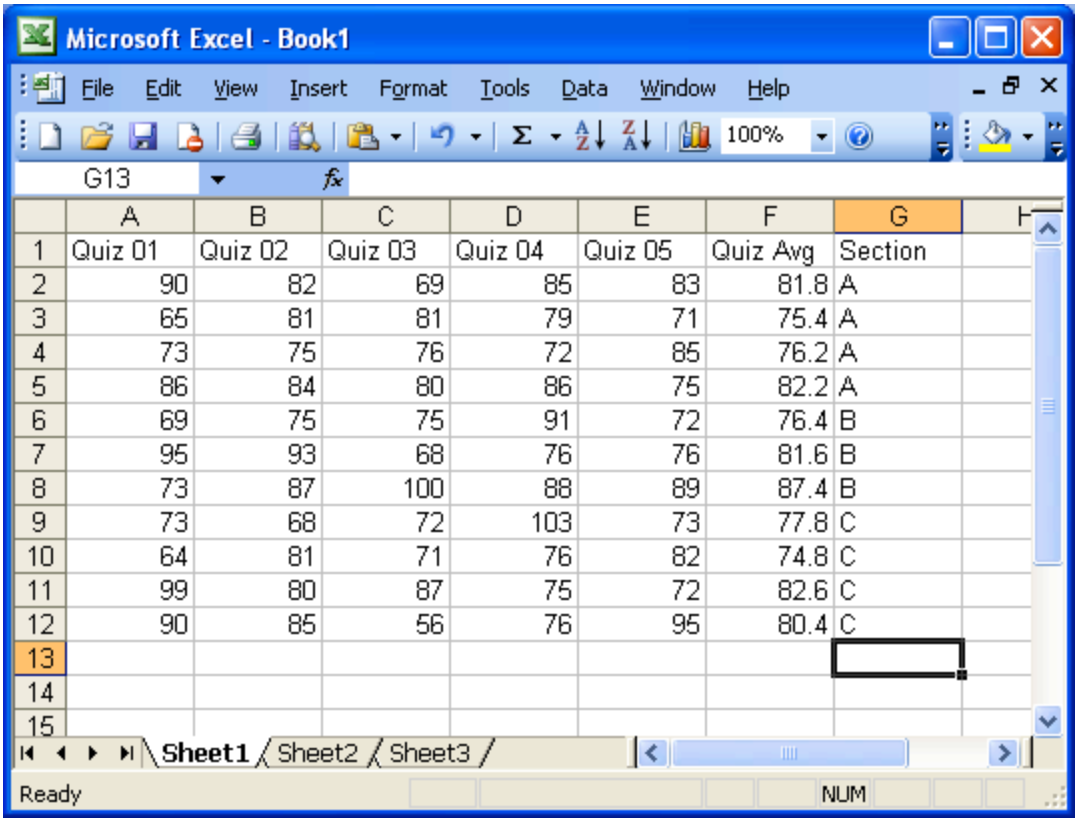
So, if the value is cell F2 is less than 80, Excel will calculate 80-F2. If the value in cell F2 is not less than 80, Excel will print the statement "Don't do bonus project." Thus, the test for the IF() function is the first argument (F2<80). The action for a true statement is the second argument (80-F2). And the action for a false statement is the third argument ("Don't do bonus project.").

Of course, the nice feature of this function is that the professor need not first determine who does and does not need to do the project, and the needed points are calculated automatically. Here are the results of copying the H2 cell to the H2:H12 range.

F	G	H	I
Quiz Avg	Section	Comment	
81.8	A	Don't do bonus project.	
75.4	A	4.6	
76.2	A	3.8	
82.2	A	Don't do bonus project.	
76.4	B	3.6	
81.6	B	Don't do bonus project.	
87.4	B	Don't do bonus project.	
77.8	C	2.2	
74.8	C	5.2	
82.6	C	Don't do bonus project.	
80.4	C	Don't do bonus project.	

Subtotals

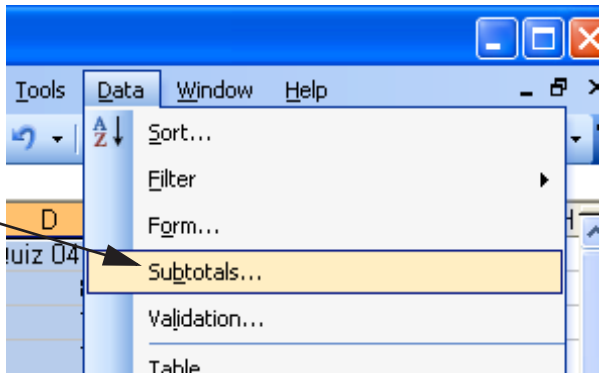
Excel provides an automated feature to obtain subtotals from a table of data. For this example, we will work with the following collection of grades.



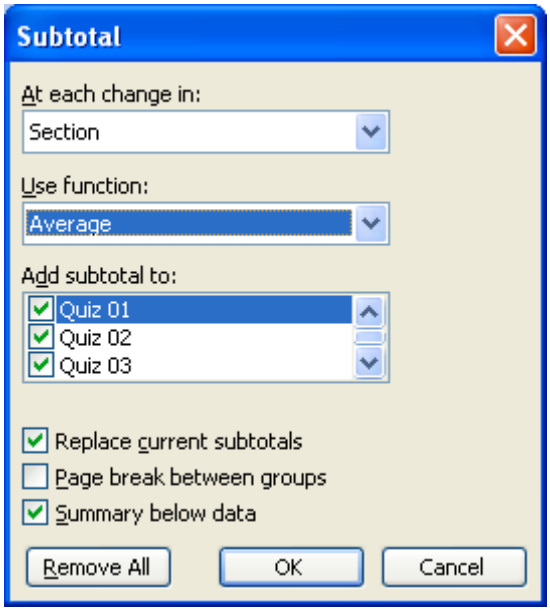
	A	B	C	D	E	F	G
1	Quiz 01	Quiz 02	Quiz 03	Quiz 04	Quiz 05	Quiz Avg	Section
2	90	82	69	85	83	81.8	A
3	65	81	81	79	71	75.4	A
4	73	75	76	72	85	76.2	A
5	86	84	80	86	75	82.2	A
6	69	75	75	91	72	76.4	B
7	95	93	68	76	76	81.6	B
8	73	87	100	88	89	87.4	B
9	73	68	72	103	73	77.8	C
10	64	81	71	76	82	74.8	C
11	99	80	87	75	72	82.6	C
12	90	85	56	76	95	80.4	C

This professor has three sections (A, B & C) and has given five quizzes. Each row represents a different student. This information has been entered in Excel starting at cell $A$1.

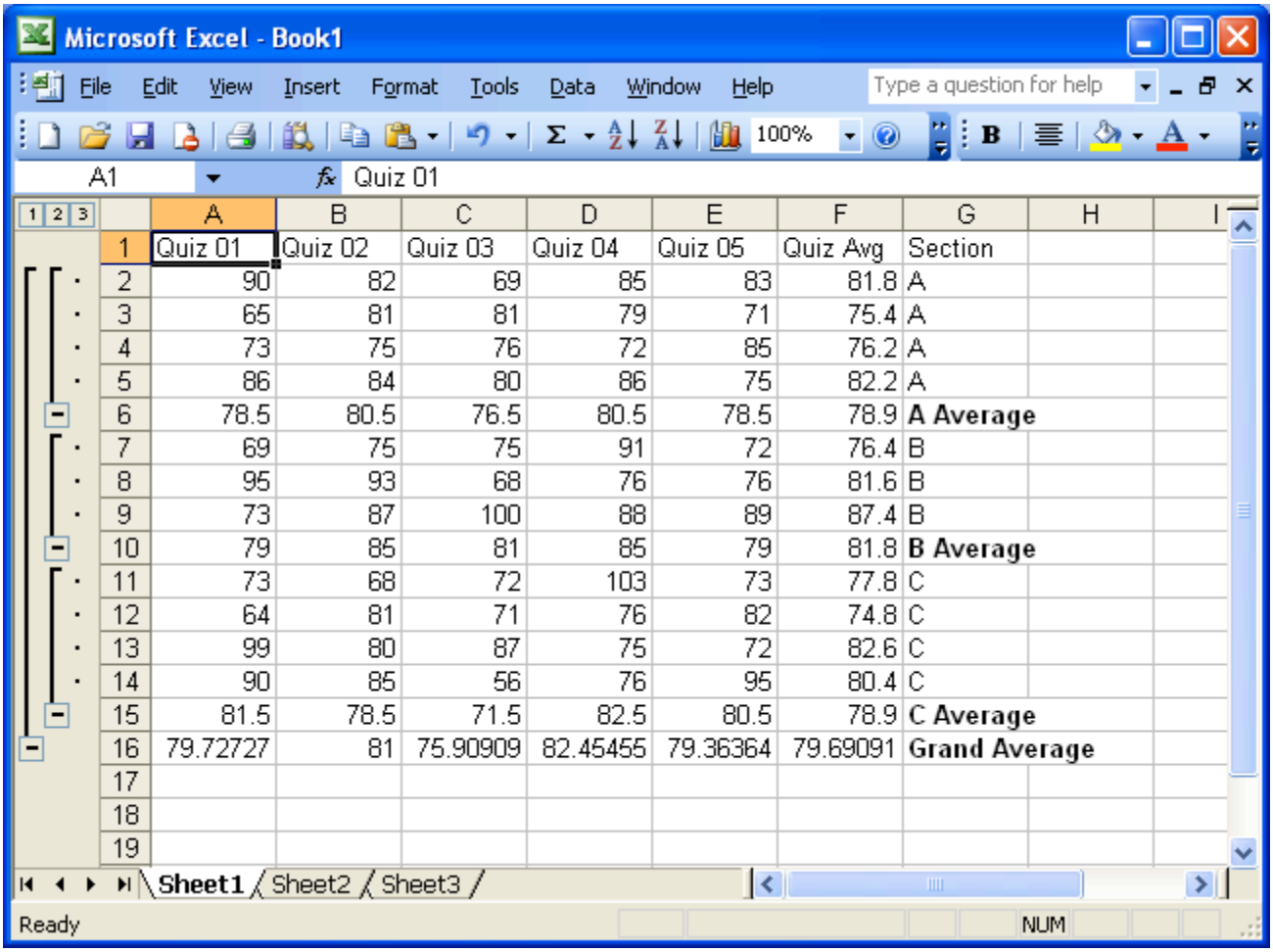
To use the Subtotals feature, select the table (which must have labels in the first row) and from the menu select Data > Subtotals…



This will bring up the following dialogue box:

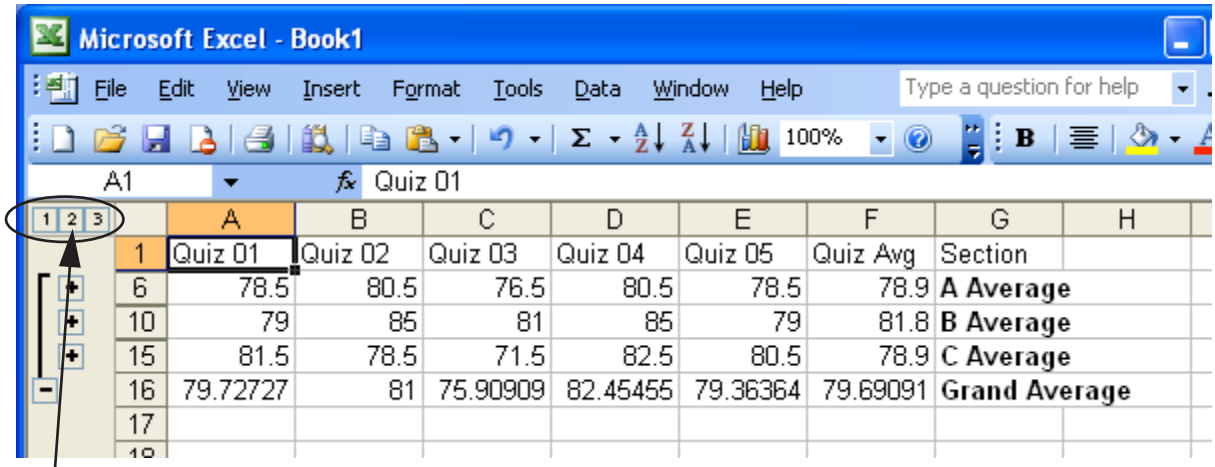


Because the professor wants the subtotals for each section, choose "Section" from the first drop down menu. Of the functions available in the second drop down menu (Sum, Count, Average, Max, etc.), the most useful function to our professor is "Average." Finally, select all of the columns for which you would want the subtotal (or sub-average, in this case). While only three selected columns are displayed, everything from Quiz 01 to Quiz Avg has been checked. Select OK Here is the output.



	A	B	C	D	E	F	G
1	Quiz 01	Quiz 02	Quiz 03	Quiz 04	Quiz 05	Quiz Avg	Section
2	90	82	69	85	83	81.8	A
3	65	81	81	79	71	75.4	A
4	73	75	76	72	85	76.2	A
5	86	84	80	86	75	82.2	A
6	78.5	80.5	76.5	80.5	78.5	78.9	**A Average**
7	69	75	75	91	72	76.4	B
8	95	93	68	76	76	81.6	B
9	73	87	100	88	89	87.4	B
10	79	85	81	85	79	81.8	**B Average**
11	73	68	72	103	73	77.8	C
12	64	81	71	76	82	74.8	C
13	99	80	87	75	72	82.6	C
14	90	85	56	76	95	80.4	C
15	81.5	78.5	71.5	82.5	80.5	78.9	**C Average**
16	79.72727	81	75.90909	82.45455	79.36364	79.69091	**Grand Average**

The structure of the worksheet has been changed to represent the three levels of the data set. Level 3 represents the students, Level 2 represents the sections, and Level 1 represents all of the professors grand total. One of the advantages of this display, is that you can hide the student information by selecting level 2.



	A	B	C	D	E	F	G
1	Quiz 01	Quiz 02	Quiz 03	Quiz 04	Quiz 05	Quiz Avg	Section
6	78.5	80.5	76.5	80.5	78.5	78.9	**A Average**
10	79	85	81	85	79	81.8	**B Average**
15	81.5	78.5	71.5	82.5	80.5	78.9	**C Average**
16	79.72727	81	75.90909	82.45455	79.36364	79.69091	**Grand Average**

The three level buttons

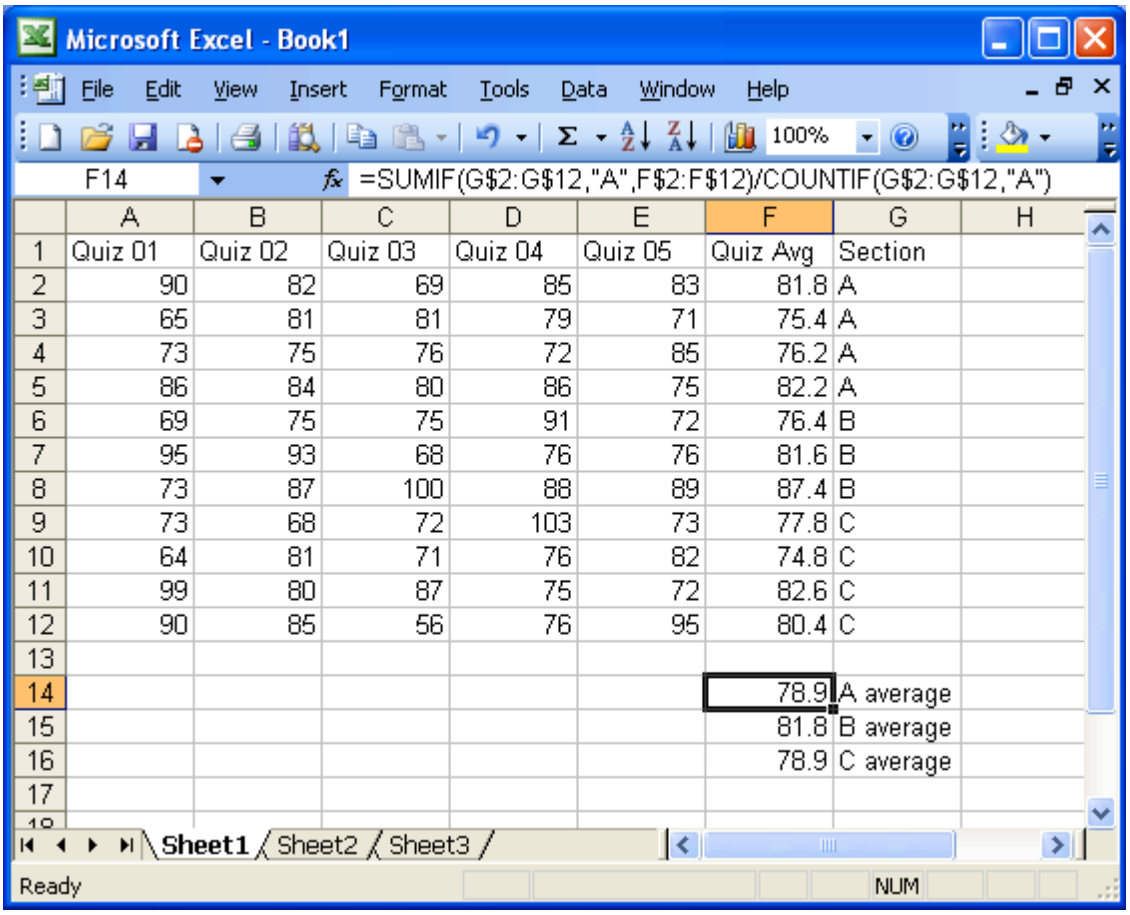
This hides the students, but displays the average for each section, and the Grand Average for all of the students. Before you click the level one button, predict what will be displayed.

The subtotal feature assumes two things. First, there is a header row with titles for each column in your table. Second, the data has already been sorted by the column for subtotal identification (in this case, the Section). (If the data needs to be sorted, the fastest way is to select the table, hit the Tab key until the appropriate column is highlighted, then use the sort ascending or sort descending buttons.)

Some people like the ability to hide the lower levels of the data set. Other people do not like having the subtotals reported inside of the data set. The next section introduces the COUNTIF() and SUMIF() functions as an alternative to subtotals.

COUNTIF() & SUMIF()

The COUNTIF() function will count how many times a particular value appears in a range of values. The SUMIF() function will sum those values, or corresponding values in another column. Here is a display of a similar result from that obtained using Subtotals:



	A	B	C	D	E	F	G	H
1	Quiz 01	Quiz 02	Quiz 03	Quiz 04	Quiz 05	Quiz Avg	Section	
2	90	82	69	85	83	81.8	A	
3	65	81	81	79	71	75.4	A	
4	73	75	76	72	85	76.2	A	
5	86	84	80	86	75	82.2	A	
6	69	75	75	91	72	76.4	B	
7	95	93	68	76	76	81.6	B	
8	73	87	100	88	89	87.4	B	
9	73	68	72	103	73	77.8	C	
10	64	81	71	76	82	74.8	C	
11	99	80	87	75	72	82.6	C	
12	90	85	56	76	95	80.4	C	
13								
14						78.9	A average	
15						81.8	B average	
16						78.9	C average	
17								

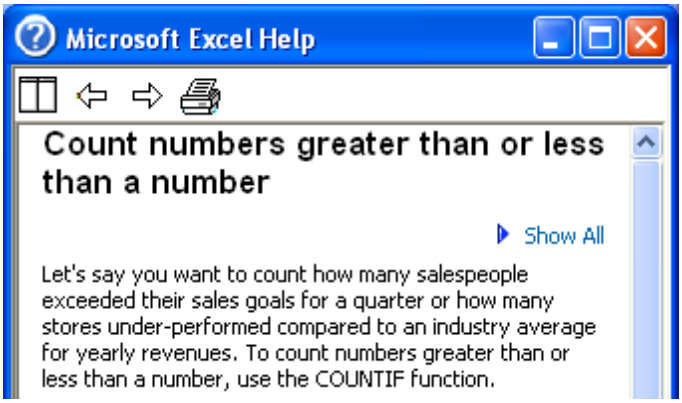
The formula that has been entered into cell $F$15 is

=SUMIF(G$2:G$12,"B",F$2:F$12)/COUNTIF(G$2:G$12,"B")

The SUMIF() function has three arguments. The first input is the section column (G$2:G$12). The second input (separated by a comma) is the match criteria ("B"); this tells Excel that you are only interested in those values associated with section "B". Finally, the third input is the column of values to be summed (F$2:F$12).

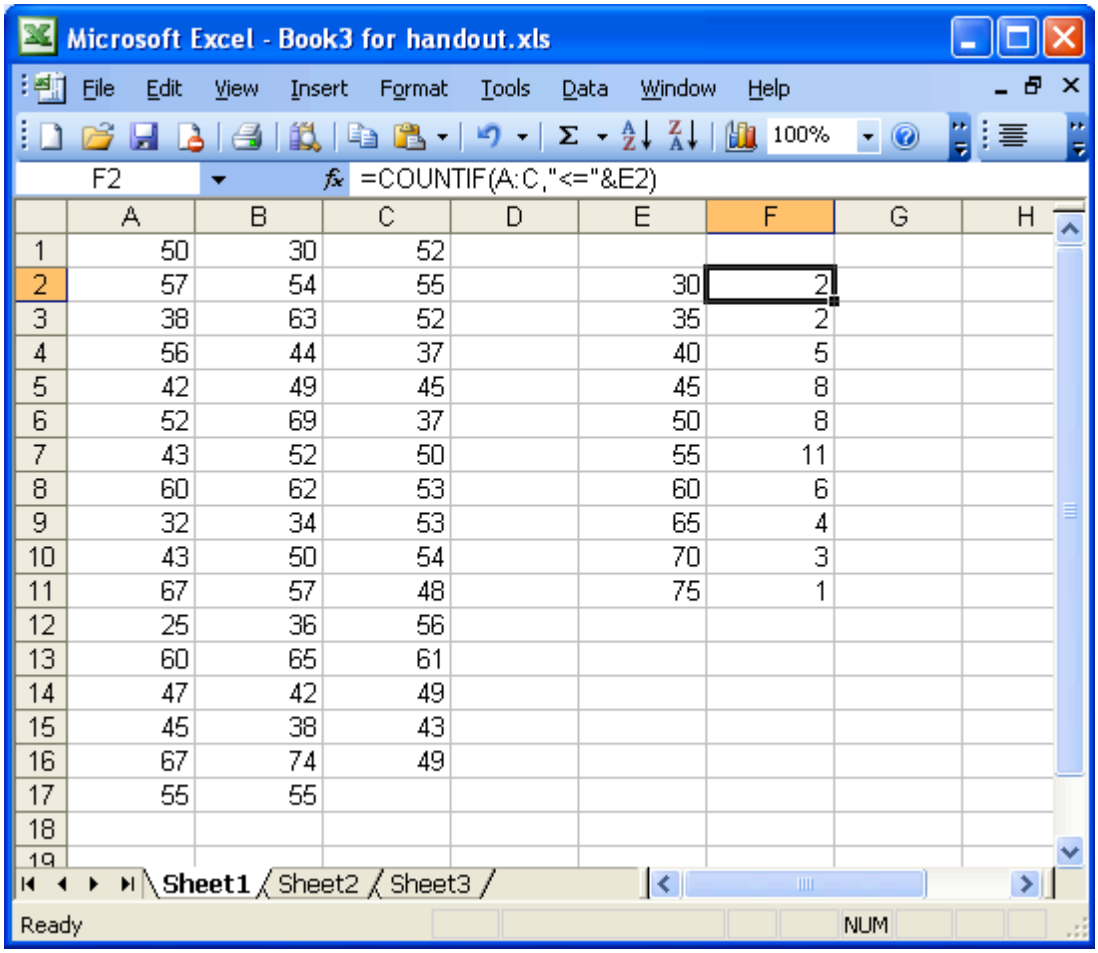
The COUNTIF() function has two arguments, and is structured much the same way as the SUMIF() function. The first input is the column to be searched, and the second input is the match criteria. Thus, this COUNTIF() will report how many times "B" appears in the "Section" column. Because this is the sum of all section "B" students' scores divided by the number of section "B" students, this is just the average for section "B". Note that in the screen capture, the formula for cell $F$14 is visible; the only difference is the match criteria for each function has been changed to "A". (Relative referencing was used for the rows so that the cells could be copied and only the match criteria would need to be changed.)

Please be sure to check out the Excel help menu to learn how to use the COUNTIF() function to count how many values are less than or greater than a given value.



One aspect of the Analysis ToolPak is unpopular with some Excel users: the results are not connected to the data. That is to say, if you change a value in the data set, the Analysis ToolPak results will not update automatically. This section provides an example to generate a frequency distribution table using the COUNTIF() function. It is a little more advanced, and is provided for the eager Excel user.

Working with the data values in the previous two examples, we will generate the same frequency distribution table as the Analysis ToolPak.



	A	B	C	D	E	F	G	H
1	50	30	52					
2	57	54	55		30	2		
3	38	63	52		35	2		
4	56	44	37		40	5		
5	42	49	45		45	8		
6	52	69	37		50	8		
7	43	52	50		55	11		
8	60	62	53		60	6		
9	32	34	53		65	4		
10	43	50	54		70	3		
11	67	57	48		75	1		
12	25	36	56					
13	60	65	61					
14	47	42	49					
15	45	38	43					
16	67	74	49					
17	55	55						
18								
19								

Take note of the use of the concatenation operator (&) in the formula for cell $F$2. This is the same as having entered the following formula:

=COUNTIF(A:C,”<=30”)

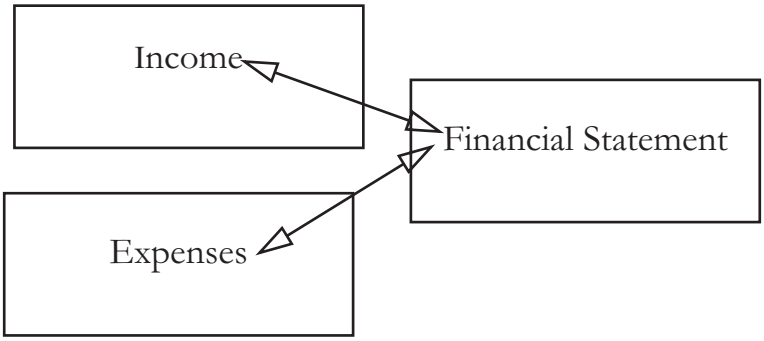
The formula for the next cell, $F$3, is

=COUNTIF(A:C,”<=”&E3)-SUM(F$2:F2)

This formula starts the same as the formula in cell $F$2, but it subtracts all the values that occur above it. (Note the combined use of absolute and relative references.) The formula in $F$3 can now be copied and pasted into the remaining cells ($F$4:$F$11).

Linking Worksheets

With Excel, you can dynamically link separate worksheets, either within a workbook or between workbooks. Any changes you make in one worksheet are automatically reflected in any other worksheets that are linked to it. Linking is especially convenient for managing formulas that utilize data from more than one worksheet. A classic example of this comes from accounting: you can create one worksheet as an income record and a separate one for expenses, and then create an financial statement that draws totals from both. If you need to edit your income spreadsheet, your final financial statement will automatically update the totals.



Link two Excel documents by creating a formula in one document that refers to a value in the other. The linking formula contains an external reference. An external reference consists of the document name and a cell reference separated by an exclamation point. =Worksheet2!$B$3

Like a regular internal reference formula, you always begin with "=" Then you name the supporting worksheet, put in an exclamation point (indicating an external reference), and finally the cell to which you would like to refer.

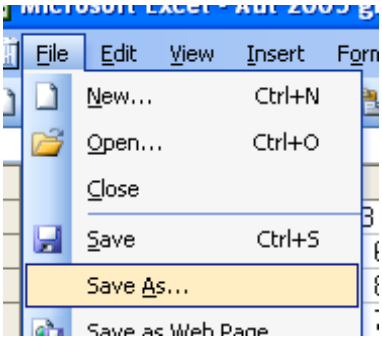
Alternatively, if you have both the income worksheet and the expenses worksheet open at the same time, you can simply type "=" and then go to the other worksheet and click in the desired cell. Excel will automatically create an external reference equation. Chart creation using external references follows the same principle.

Warning: The dollar signs ($) in the formula above indicate that this is an absolute reference. This means that the reference refers to the exact column letter and row number of the cell, so that if the cells in the supporting worksheet shift, your reference could suddenly point to a different value. If you wish to use a relative reference, in which the reference floats as cells shift, remove the dollar signs. Both reference styles have their advantages - you will need to determine which is best for your particular needs

For more information about formulas, or about worksheets vs. workbooks, please see the "Introduction to Excel" document.

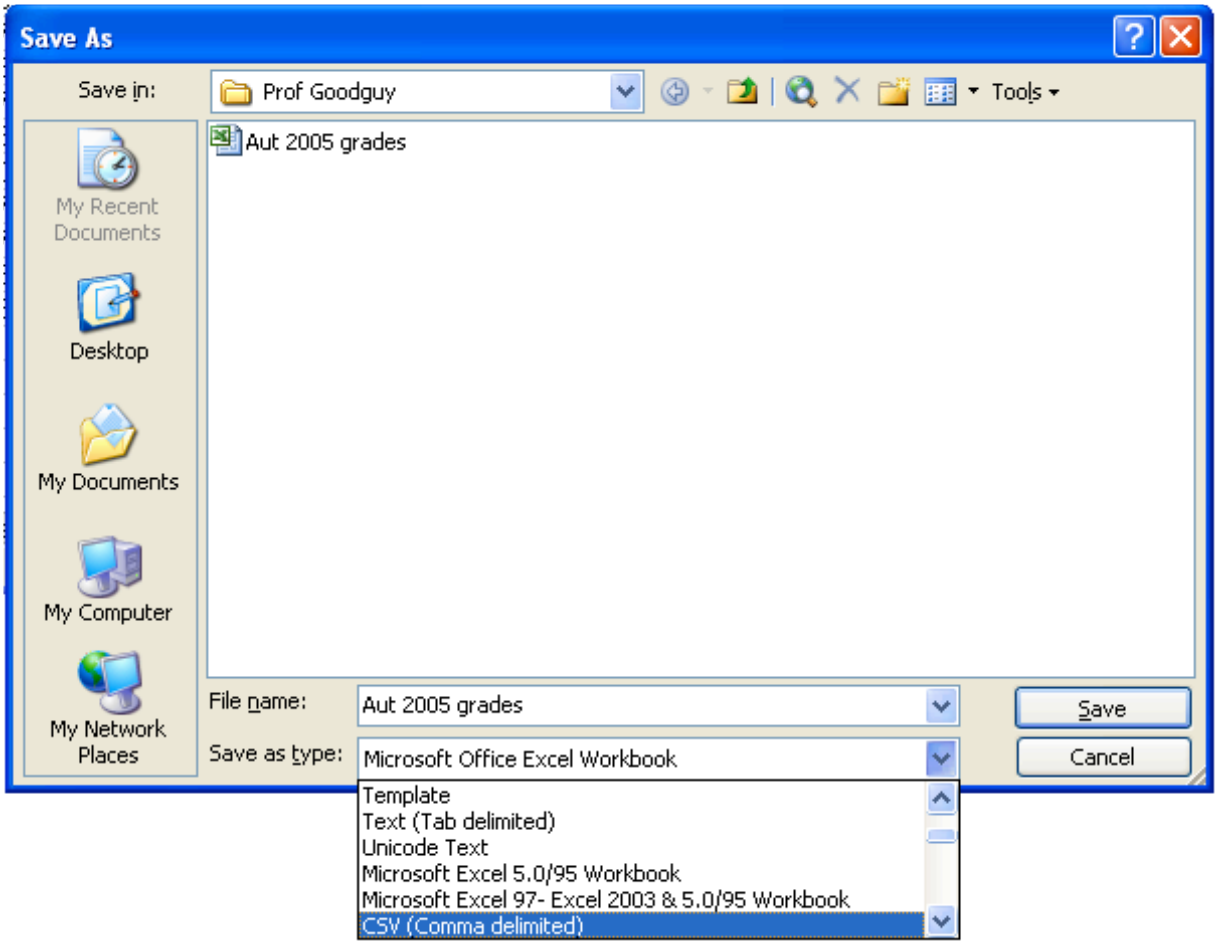
Exporting Data

Excel can convert workbooks into various formats. In that many people want to use data in other programs (such as SPSS, SAS, STATA, and R, to name just a few), there are a number of useful options for exporting data. Note, while most of the data will be saved to other formats, many formatting features available in the most recent version of Excel cannot be transported to other formats. Some of the available formats to which Excel files can be saved are HTML/XML, older versions of Excel, Lotus 1-2-3, QuatroPro, and dBase. However, one of the most useful file formats is the Comma Separated Values (*.csv) file format. First, this file can easily be read by most programs. Second, it produces a relatively small file for emailing and web-posting. And, third, it can easily be edited with a text editor or word processor (you need not use Excel).



Let's save our professor's grade book in a *.csv format. It is always recommended that you first save the original Excel file as an Excel file (*.xls). Next, select File > Save As…

This will produce the dialogue box shown below. From the "Save as type:" drop-down selection box, choose CSV (comma delimited)"

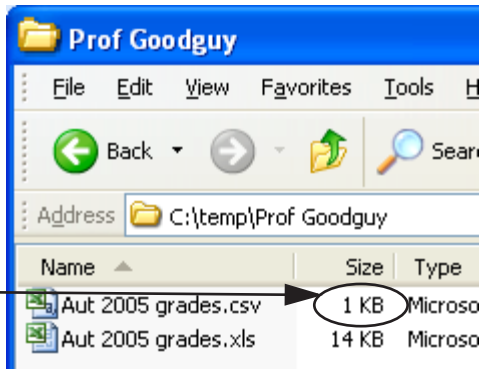


Normally, one or two warning dialogue boxes will open. In each case select the option to proceed with saving the file in the selected format.

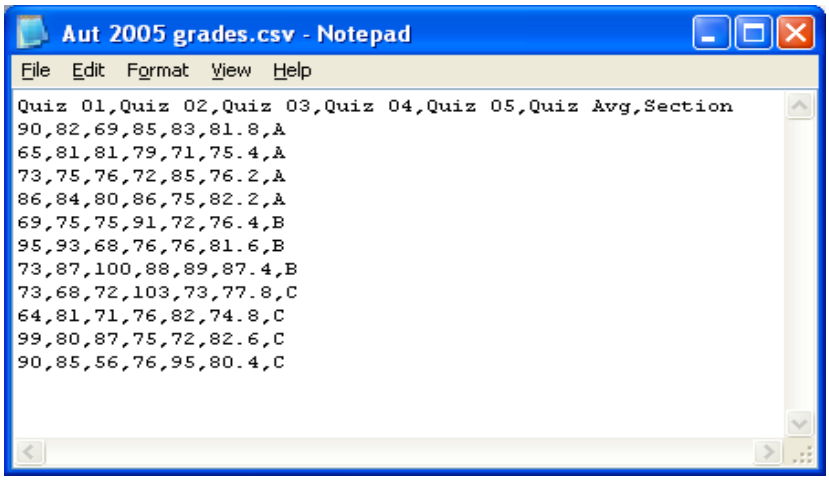
Now, if you check the file directory, the file size has been reduced by tenfold:

If you double-click on the *.csv file, it will naturally open Excel. However, if you use the right mouse button, you can choose to open with Notepad (a simple text editor).

And, this is the much smaller, much more portable file:

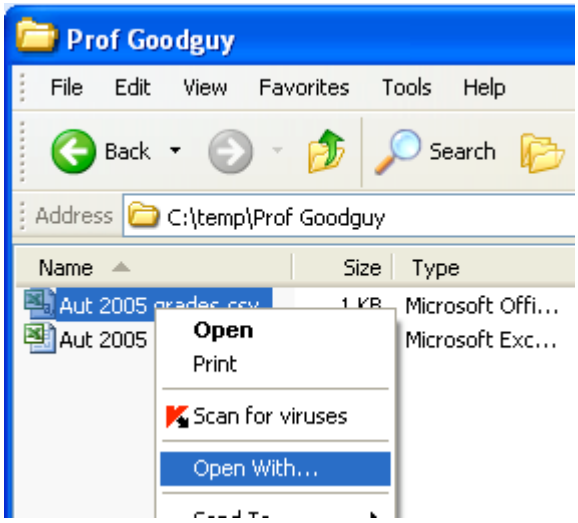


Most programs can use these *.csv files, however, you may need to change the "csv" extension to "dat" (for data file).



Excel also offers a tab-delimited option.

Note: Another useful advantage to the *.csv file is that it is very convenient to tag missing data. Say you want to code all missing data as the number 999. However, your data set has been entered by multiple people. Some people used the dash (-) for missing data, some people entered "n/a" and other people left the cells blank.



Intermediate Excel

Written by Gregg Harbaugh

Center for Social Science Computation & Research
145 Savery Hall
University of Washington
Seattle WA 98195 U.S.A.
(206)543-8110

December 2005

http://julius.csscr.washington.edu/pdf/excelint.pdf