

EXCELSIOR

An Excel lab manual for CS 101 at Furman University, by Chris Healy, Fall 2009

Table of Contents

Lab	1	2	3	4	5
Topics	Editing Formulas Cell ranges Formatting Absolute ref. Quick sum	Functions Mixed ref. Statistical fun's If Vlookup Cond. formatting	Charts Sort Filter Print Save Web page	Subtotals Pivot tables Sumif 3-D reference	Importing text Text fun's Web query Scenarios Goal seek Solver
Example files	calorie furman cost gradedemo foreign lang f08 checkbook alert2005	loan, calorie mult table loan table climate interview ttt gradedemo tax table mass ed	furman cost canada calorie bedroom gradedemo city education schools	employee mileage data college2006 ag college sales	gradedemo2 NHL win-loss gradedemo loan solver texas tax table

Excel is a program by Microsoft Corporation. The purpose of Excel is to manipulate and analyze data. With Excel, you can effortlessly perform thousands of calculations. The type of file used by Excel is called a *spreadsheet*, and its file name typically ends in “.xlsx”. Spreadsheet programs like Excel have been around for about 30 years. More traditional ways of doing lots of computations were much more tedious. Today the use of spreadsheets is widespread in business and research around the world.

Excel is a particularly useful program because it has so many built-in functions to simplify our work. For example, it is straightforward to sort data, compute subtotals and produce professional-looking charts. More importantly, Excel allows us to ask probing or subtle questions about our data, and draw conclusions that would have been very difficult to make just by scanning the data ourselves manually. People use Excel to help them make important quantitative decisions. Although this document contains information that deals specifically with Excel, it also contains general quantitative concepts and techniques that should be valuable to you for many years to come.

Before we begin the labs, you need to *download* the folder containing all the necessary Excel example files to your memory stick or account. In lab, I can show you where you can find these files.

There's no need to rush through the labs. Read the directions slowly and carefully. Write your answers to questions as you come to them. You may find it convenient to check off each example as you move on to the next. When you are finished with a lab, please have the instructor check your work before you leave.

Excel lab #1

This first lab will cover the most fundamental skills: how to enter values and formulas into a spreadsheet. We will also do a little formatting.

Introduction

When you first open a spreadsheet, it almost looks like a sheet of graph paper. Data is arranged into boxes called *cells*. The cells are arranged into rows (horizontal) and columns (vertical). A spreadsheet file consists of one or more pages or “worksheets”.

As an example, consider a spreadsheet containing baseball statistics. You might have one page devoted to batting statistics, another page for pitchers, and finally a third page for statistics about overall team performances. Each page would show data arranged into neat rows and columns. All these data are related to baseball, but they are arranged into different pages because you would not need to see all of it simultaneously. If you also wanted to have Excel keep track of data about the stock market, it would make sense for this information to go into an entirely different spreadsheet file. It’s generally a good idea for a spreadsheet not to have too many pages. Otherwise it may be difficult to find your data.

For our first lab, our spreadsheets will use just one page.

The rows of a spreadsheet are numbered 1, 2, 3, etc. The columns are identified by letter: A, B, C, etc. After the letter Z, the columns are given multiple-letter names such as AA, AB, AC, etc., although rarely do we need so many columns. In the long run you will find that spreadsheets tend to use many rows, but not many columns. The name of a cell is given by the column and row that it is in. For example, E7 is the cell in the fifth column and the seventh row. It turns out that column letter names are not case sensitive. This means that it does not matter if you refer to them using capital or lowercase letters.

Into each cell, we can type a *value* or a *formula*. A value is either a number or text. A formula is a mathematical expression starting with an equals sign (=) in which you ask Excel to perform a calculation. In addition, you may leave cells blank, and it does not matter to Excel where you begin typing your data. You should arrange your data so that you find it easy to read, organized into rows and columns as you see fit.

How do we use Excel? Here is the general routine: First we enter some raw data manually. Then, we ask Excel to calculate the results of some formulas. Next, we may insert more data and ask Excel to update the results. Once we have enough data in our spreadsheet, we can study the results to look for interesting trends or conclusions. We can ask Excel to draw charts based on our data. Finally, when we are finished, we can format the spreadsheet using font and/or color effects to make it appear more professional.

The version of Excel we will be using is Excel 2007. If you have used an earlier version of Excel in the past, you may notice a few differences. The most obvious difference is the “ribbon” at the top of the screen, with a different menu organization from the 2003 version. Spreadsheet file names now end in

.xlsx instead of .xls. If you have an old spreadsheet in .xls format, you should probably save it in the new .xlsx format. It turns out that the new spreadsheet format is much more space-efficient. In many cases the new spreadsheet is less than half the size of the old .xls format. Another nice feature of Excel 2007 is that it can handle larger spreadsheets than before. The maximum number of rows on a page has increased from 65,536 to 1,048,576 (i.e. from 2^{16} to 2^{20}). The maximum number of columns has also increased, but we won't need to worry about that in this course. Finally, Excel 2007 has some attractive new options for presenting and formatting data that we will see in later labs.

One thing to keep in mind about Excel and other software is that there are usually several ways to accomplish the same task, such as how to open or save a file. You have probably noticed this when using other applications.

Editing a spreadsheet

Editing data in Excel is straightforward. You just need to remember that everything that you type goes into some cell. The first cell in the worksheet is called A1. Usually, we reserve row 1 and column A for header information, as you will see in our examples. Excel gives you the freedom to use whichever rows and columns you wish. We will discuss formatting later.

To navigate with a spreadsheet, we commonly use the arrow keys. To move rapidly up and down in a large spreadsheet, we can use the Page Up and Page Down keys. Scroll bars can also be used, as needed.

Example #1 – calorie counter

Open the spreadsheet [calorie.xlsx](#). This file shows how many calories have been consumed by someone during the course of a week. The data that you see was simply typed in manually. Right now, this spreadsheet shows only raw data, and we need to enter formulas that compute the total number of calories for each day, and the average number of calories for each type of meal. (See the to-do list at the bottom of the spreadsheet.)

Column widths

First, we have a little problem. The word “Wednesday” appears cut off. It turns out that, by default, all cells are given the same size. But Wednesday is a long word. We can widen column D so that we can see the entire word. Position the cursor between the letters D and E at the top of the spreadsheet. The cursor should look like a vertical bar with arrows pointing both left and right. Now, drag the boundary of columns D and E to the right to accommodate the word Wednesday. Once you've done that, you might look at the other columns and notice they seem a little too *wide*. For example, the word Friday does not need a lot of space. Thus, you can drag the boundary of columns F and G to allocate a little less space for column F.

There is actually a shorter way to format column widths. If you double click on the boundary of columns F and G, this will automatically format column F to just the right size so that it will fit the largest entry in that column. Try this technique again: double click the border between columns B and C. You should notice that column B now fits the word “Monday” perfectly. You can streamline this process one step further: we can select all five columns and format them all at once. To select all 5 columns, click on the letter “B” and drag the mouse until you reach column “F”. You should notice that the 5 columns have become highlighted. Release the mouse button, and now double click the boundary between any two columns in this range. This has the effect of formatting the widths of all the selected columns.

Note that we did not want to format column A in this manner because at the bottom of the spreadsheet, cells A9 and A10 contain long phrases, and we don’t want these to dictate the size of the columns in our calorie spreadsheet. However, we do want column A to be wide enough to fit the phrase “total per day” in cell A6. Manually set the boundary between column A and B.

Column G, which is blank, probably does not need to be so wide. We can manually shrink its width. We can double click the boundary of H and I to make H just the right width as well.

Formulas

Let’s calculate calorie totals per day. Row 6 is where we will enter our formulas. Let’s start with Monday. We want to add the numbers 555, 650 and 900 and put the result in cell B6. Formulas always begin with an equals-sign (=) followed by a mathematical expression involving the cells we want to compute. We could enter formula this way: =555+650+900. This would be a legal formula in Excel. But it’s generally not a good idea to use constant values this way. Suppose later you realized that your dinner calories were actually 910 instead of 900. Not only would you need to change cell B4, but you’d also need to change the formula in B6. It is much better if your formula can make use of cell references. Thus, the formula for our Monday calories should be: =b2+b3+b4. (The period is not part of the formula.) Type this formula into B6. Notice that as you enter this formula, Excel gives you a visual cue: colored boxes temporarily appear around the cells being referred to in your formula. This helps you guard against possibly entering the wrong cell reference in your formula. When you are done typing a formula, hit enter, and the computation is performed automatically.

We see that we ate 2105 calories on Monday. But where did our formula go? It’s still there! Click in the cell B6 again. At the top of the spreadsheet there is a “formula bar” which tells you what information was literally entered in that cell. The formula bar should say: =B2+B3+B4. If you wanted to change the formula, you could edit the information in the formula bar, or you could double click on cell B6 to re-enter the formula.

Now, how about formulas for the rest of the week? For Tuesday, we could again type in a formula that adds up breakfast, lunch and dinner on Tuesday. C6 would contain a similar formula that we entered for B6. And so on for cells D6, E6 and F6. Does this mean that we have to type in 4 more formulas, one in each of these cells? No: there is a general rule in Excel that if something we need to do seems repetitious, there is probably an easier way to do it. We know that the general way to calculate the total calories per day shouldn’t really change from day to day. We already have in B6 a formula to add

up breakfast, lunch and dinner calories for Monday. All we need to do is apply it to the other days of the week. So, we simply need to *copy the formula*.

Here is how you copy a formula. Note that in this case we have a formula in cell B6, and we want to apply it to the 4 cells to the right. Click cell B6. Notice that in the lower right corner of this cell there is a small black square. This is called the *fill handle*. Using the mouse, drag the fill handle to the right until you reach cell F6. Notice that as you drag your mouse, the cursor looks like a plus sign. Then, release the mouse button. This will copy your formula to cells C6 through F6. You can verify that the formulas are correct: for example, click in cell E6, and you'll see the formula for Thursday, which is `=E2+E3+E4`.

Entering a function

The last thing we need to do for this spreadsheet is to calculate the average number of calories for each type of meal: breakfast, lunch and dinner. These formulas should go in cells H2, H3 and H4, respectively. How do you calculate an average? For instance, in H2 we want the average of the cells B2, C2, D2, E2 and F2. We could write a formula that adds up all five of these numbers and then divide by 5. The formula would look like this: `=(b2+c2+d2+e2+f2)/5`. However, could you imagine how tedious this formula would be if we were averaging a whole month of breakfasts? Fortunately, there is an easier way. Averaging values is such a common operation, that Excel has a built-in function for it. It's called the "average" function, and here is how we would use it. The formula is: `=average(b2:f2)`. Type in this formula in cell H2. Basically this function says "average the values in the cells B2 through F2."

Use the fill handle to copy this formula to cells H3 and H4. Now you have the average number of calories of lunch and dinner as well.

Cell ranges

Notice that when we used the average function above, we used a colon, as in `b2:f2`. The colon is used to specify a range of cells. It's much more convenient to refer to a range of cells than to list them one by one. We'll use cell ranges with functions quite often. And cell ranges can be two dimensional. All you need to do is identify the first and last cell in the range. For example, the range `a1:c7` would refer to 21 cells (7 rows and 3 columns). What formula would we write if we wanted to calculate the average number of calories of all 15 meals eaten during the week? Write your answer here:

Now that you know about the average function, I should tell you about another function: sum. Earlier we were adding calories by day. For instance, B6 had the formula `=b2+b3+b4`. It turns out we could have used the sum function like this: `=sum(b2:b4)`. In future labs, we'll see that there are many more functions besides average and sum that can do very useful computations: functions that can do statistical, financial, logical, text, as well as arithmetic computations.

We are now finished with the calorie spreadsheet, so you can save and close the file. To save the file, hit CTRL-S, or you can select Save in the Windows button menu at the top left corner of the Excel screen. To exit Excel, you can click the X button at the top right of the window.

Example #2 – the cost of attending Furman

Now, open this file: [furman_cost.xlsx](#). We will practice with simple Excel operations on this spreadsheet, which shows a brief history of the tuition, room and board charges at Furman since 1920. We want to perform some summary calculations so we can draw conclusions from our data. A to-do list appears at the bottom of the file. However, we will save step 5 (charts) for a future lab.

The first thing to do is to compute the total cost per year. Where should this information go? Based on how this spreadsheet is arranged, it appears logical that we should put the yearly totals on the right side of the spreadsheet, probably in column E or F. You might like to put the totals in column F and leave column E blank for readability. It doesn't matter; we can easily change locations later. Let's put the total in column E. In cell E3, type the word "total" (without the quotes). In cell E4, we need a formula that will give us the total cost for 1920-21. Enter this formula in E4: `=sum(b4:d4)`. Does this formula make sense to you? Use the fill handle to copy this formula all the way down the column to cell E12.

Next, we need to determine, for each year, what percent of the total yearly cost comes from tuition. We'll put the results in column F, next to the total costs. In cell F3, enter the phrase "tuition percentage". This is a long phrase, so make sure to widen column F appropriately.

How do we calculate a percent? The way to do this in Excel is to divide a number by the total, and then *format* the result as a percent. In particular, we do not multiply by 100. For example, if you divide 3 by 10, the result is 0.3. This number formatted as a percent would appear as 30%, which is numerically the same as 0.3.

In cell F4, enter the formula `=b4/e4`. In other words, we want to divide the tuition by the total cost. Now it's time to format this result as a percent. Right click on cell F4. When the menu appears, select "Format cells..." A Format Cells dialog box appears. In it, under the Number tab, you will see a list of categories. Choose "Percentage." Off to the right we can specify the number of decimal places we want. For this exercise we don't need much precision, so make the number of decimal places zero, and click the OK button to dismiss the dialog box.

You have now calculated the percentage of the 1920-21 cost that is represented by tuition. Copy the formula in F4 to the other cells in that column down to F12 (using the fill handle). Notice that the copied formulas also incorporate the percentage format you specified. This is a convenient feature of Excel. We'll discuss a little more about formatting later.

Inserting/deleting rows or columns

So far, our spreadsheet on Furman cost has useful information. But all of a sudden, we realize that we're missing a year's worth of data. For 1962-63, the tuition, room and board were 750, 300 and 500, respectively. These numbers need to be added to the worksheet. Where do we put them? The most logical place would be in sequence between rows 6 and 7. We need to insert a new row between existing ones. In Excel, it's easy to create new rows or columns as we need them. To insert this new row, right click any cell in row 7. When the menu appears, choose Insert.... When the Insert dialog box appears, select Entire row, and click the OK button. This will insert a new row above row 7. This new row is the "new" row 7, and all the rows underneath move down by 1 row. Excel automatically updates any formulas as needed. If you had selected a new column, it would have appeared to the left of the current cell.

Now, you are ready to enter the appropriate data for 1962-63: enter these values in cells A7 through D7. The next thing to do is to copy formulas into cells E7 and F7. We can simply copy them from the cells immediately above. Highlight (select) both cells E6 and F6: you can do this by clicking on cell E6 and dragging the mouse until you reach F6. You have now selected both cells. Both these cells contain a formula which you will copy. Use the fill handle and drag into row 7. You may want to check with me or a lab assistant to make sure you did this operation correctly. At this point, the formulas should have simultaneously copied into cells E7 and F7. This technique of copying multiple formulas can save you time and effort.

If we can insert a row or a column, how about deleting one? For example, we may decide that the row containing values for 2006-07 is redundant given the fact we have data for 2008-09. To delete row 12, right click on the row number 12 (i.e. the actual numeral "12" to the left of the row). This highlights the entire row, and when the menu appears, select Delete. The row vanishes, and all the rows below it move up in position. Any formulas in those lower rows have been changed automatically to reflect their new positions. Now, you may have second thoughts about the deletion. In Excel, you can almost always *undo* an operation you just performed: at the top of the Excel window there is an undo button, labeled with a curved arrow pointing left. Notice that there is a *redo* button as well, and little triangle buttons next to both the undo and redo buttons allow you to select how many actions need to be undone or redone. Let's click the undo button to reinstate the 2006-07 data. An alternative to clicking on the undo button is the keystroke sequence CTRL-Z.

By the way, other control key sequences also work in Excel. For example, CTRL-C means "copy", CTRL-X means "cut" and CTRL-V means "paste." If you don't like using control keys, you can accomplish these same actions by right clicking and selecting the appropriate command from the menu that pops up.

Formatting

There are a few things we can do to improve the appearance of the spreadsheet. First, at the top of the file there is a title given in row 1. We can boldface this title and slightly enlarge the font. You should see the appropriate buttons for accomplishing this in the ribbon under the Home tab.

Next, we'd like to format the numbers in the spreadsheet. The percentages in column F are right justified, but this might not be appropriate when the column is wide. It may look better to center these percentages. At the top of the Excel window, make sure the Home menu is selected. Highlight all the cells in column F that contain percentages, and then click the centering button in the ribbon. (Alternatively, you can format by right clicking, and choosing "Format Cells", but this entails a few more steps.)

On second thought, it doesn't look good at all to have such a wide column in the first place. Undo the previous action of centering the percentages. What we should do instead is make column F narrower. Recall the reason why it was so wide was because the column heading has two words. In Excel, it's sometimes convenient to be able to "wrap text" when you have several words in a cell. That way, you can have more than one line of text in a cell and the column won't need to be so wide. Right click on cell F3 (the cell containing the phrase "tuition percentage"). When the menu appears, select Format Cells. When the Format Cells dialog box appears, choose the Alignment tab, and check the box called Wrap text. Click the OK button. Now you are ready to adjust the size of row 3 and column F. Arrange column F so that it is wide enough for the word "percentage", and fit row 3 so that the word "tuition" is directly above the word "percentage".

It's often a good idea to center our column headings. Select cells A3 through F3, and click on the center button located in the ribbon under the Home tab.

Finally, let's format the money. There are two ways to format numbers as money in Excel: *Accounting* and *Currency*. Remember that whenever we need to format some cells, we need to select those cells first. Which cells contain monetary values? It turns out they are in the range B4:E13. Use the mouse to highlight these cells. Then, right click, and when the menu appears, choose Format cells. When the Format Cells dialog box appears, select the Number tab. Under the list of categories notice we have both Accounting and Currency. We'll experiment with each.

First let's try Currency. Select Currency from the list. To the right, we have some options: we can set the number of decimal places, change the monetary symbol, and specify how negative amounts are to be displayed. All we really need to do here is specify 0 decimal places. Then click OK. You should see dollar signs before all of our monetary figures.

The Currency style is certainly useful, but one drawback is that it always puts the dollar sign immediately next to the number. From a distance, a dollar sign might look like a '5' and be confusing. I generally prefer the Accounting style instead. Check it out: Once again select cells B4:E13, go to "Format Cells", and this time choose Accounting instead of Currency. The decimal places again should be 0. Click OK. Do you see the difference? Now, the dollar signs are vertically aligned at the left edge of each cell, not touching the numbers. I think this is a more attractive format.

Here's a rule of thumb about money: If all of your monetary figures are in whole dollars, especially if they are large amounts of money, then you should not display cents. In other words, the number of decimal places should be zero.

Before we leave this spreadsheet, we should take a moment to interpret what information we have found here. A spreadsheet is more than just numbers – people use Excel to draw conclusions and make decisions. Take a look at the percentages that we calculated, and quickly scan the columns of data to see if you notice trends. What do you think are some important conclusions one could take from this spreadsheet? _____

Save and close furman cost.xlsx.

Example #3 – our first grade sheet

Let's look at an example spreadsheet of student grades. Open the file [gradedemo.xlsx](#). This spreadsheet shows a class in which we have already recorded test grades. We'd like to compute the total (i.e. average) grade earned by each student.

However, the first thing we need to do is adjust the column widths. Do you notice the problem? Column A is too narrow, and the other columns are perhaps too wide. Use the techniques we saw earlier to change column widths.

We are ready to calculate students' test averages. But first, we forgot to enter the exam weights. They need to be entered into row 12 of the spreadsheet. In cell A12 type the word "weights." In cells B12, C12 and D12 enter these percentages: 30%, 30% and 40%, respectively. These represent the weights assigned to each test.

Next we need to enter a formula to calculate a student's total grade, which will be a weighted average of the 3 test scores. Click in cell E2. Each of the test scores in cells B2, C2 and D2 needs to be multiplied by the corresponding percentage weight given in cells B12, C12 and D12. Enter this formula carefully into cell E3: $=b2*b12 + c2*c12 + d2*d12$. Yes, you may put spaces inside a formula to make it more readable. Hit enter, and make sure the result makes sense. Marianne's grade should be 83.9. Okay, now we are ready to copy this grade formula to the other students in the class. Use the fill handle to copy the contents of cell E2 down to cells E3 through E10.

Uh-oh! Everybody else has a total grade of 0! Do you know what happened? Click on cell E3 to see how Neil's grade got calculated. Do you see something wrong with the formula? It may help to double click on this formula so that you can better see which cells are being referred to in this formula.

Absolute references

Recall that Marianne's grade formula was: $=b2*b12 + c2*c12 + d2*d12$. We would have expected Neil's formula to be similar, except the 2's would be changed to 3's because his grade data is in row 3. Row 12 contains the exam weights, and these 12's should not change to 13's. These exam weights are *constants*, to be used for the grade formulas for every student. We need to tell Excel that when we copy a formula, the constants should not change. We do this by specifying the exam weights as absolute references.

An absolute reference is a cell reference where we put a dollar sign symbol (\$) before both the column and row names. So B12 becomes \$B\$12, for example. An absolute reference does not change the value of a cell; it does not have anything to do with money. It simply tells Excel not to change this reference whenever it appears in a formula we want to copy. At this point, we need to re-enter Marianne's grade formula. Go back to cell E3 and enter this formula: $=b2*\$b\$12 + c2*\$c\$12 + d2*\$d\12 . Do you see the difference between this formula and the one we had originally? The '\$' symbols indicate an absolute reference, which is a special marker to Excel that these references will stay constant. Now, copy the formula from cell E2 down through E10 again. This time the average grades should be computed correctly. 😊 And you can click on the various formulas to verify they make sense. For example, Neil's formula came out like this: $=b3*\$b\$12 + c3*\$c\$12 + d3*\$d\12 . Notice that the 2's have changed to 3's, but the 12's have NOT changed to 13's. This is the difference between a *relative* and an *absolute* reference.

Oh, there's one more thing we need to do for this spreadsheet. The instructor who created it has realized that the actual test weights need to be: 25%, 30% and 45%. Make the appropriate updates to the cells in row 12. Notice how the total grades update automatically. But, the problem we have now is that they are not formatted to a consistent number of decimal places. Let's fix this: highlight cells E2:E10. Right click, and choose Format Cells. Under the Number tab, choose the Number category, and specify that we want 1 decimal place. Now, the results look more attractive.

Absolute references are a powerful and important feature of Excel formulas. If you are having trouble understanding what they mean, please pause and ask me or a lab assistant to help you. The reason why we use absolute references is that we generally want to avoid using explicit constants in formulas. Consider the problem we just had with the exam weights. Just glancing at the spreadsheet, we know exactly what exam weights we are using. They are not buried in some formula. And when we needed to change the exam weights, all we needed to do was update some numbers in row 12. We didn't need to painstakingly re-enter grade our formulas.

Finally, it's a good idea to improve the appearance of the worksheet. Let's center the data in columns B-E. Highlight columns B through E, and click the center alignment button.

Save and close gradedemo.xlsx.

Example #4 – percentages of a total

In fall 2008, the student enrollment in various foreign language classes at Furman was as follows: 71 in Chinese, 171 in French, 23 in Greek, 79 in German, 1 in Hindi, 36 in Japanese, 53 in Latin, and 611 in Spanish. Let's create a spreadsheet that will hold this data and we'll do some routine calculations. On your memory stick or account, create a new spreadsheet file called foreign lang f08.xlsx. You can do this by right clicking in the background of the folder, and selecting New → Microsoft Office Excel Worksheet. Double click on this file icon in order to open Excel.

Of course, since this is a new file, the spreadsheet is initially blank. In cell A1, let's put in a title: "Foreign language enrollment at Furman, Fall 2008". Format this title as boldface. In row 3, we will enter some column headings. In cell A3, type "language". In cell B3 type "enrollment" and in cell C3, enter "percent". Eventually, we'd like to know what percentage of all foreign language enrollments is represented by each language. In column A, beginning with cell A4, enter the names of the foreign languages listed above, with A4 = "Chinese", A5 = "French", and so on. In column B, you should enter the corresponding enrollments of each language. Once you have done this, you have finished entering the raw data. Please make sure that columns A and B are large enough to hold all the words you entered. However, column A should not be so wide as to fit the title. It's okay for the title to "spill" into other columns of row 1, because they will not be used.

We are now ready to enter formulas. In row 13, we will report the total. In cell A13, type "total", and in B13, enter a formula that will determine the sum of the enrollment numbers you entered earlier in column B. In column C, we'd like to compute percentages of the total enrollment for each language. Select cell C4. Here, we need to divide the Chinese enrollment by the total enrollment. But make sure to specify the total enrollment as an *absolute reference*. Format cell C4 to hold a percentage format, with 1 decimal place of precision. Finally, copy the formula in C4 to the other cells C5 through C11 using the fill handle. Please don't hesitate to contact me or a lab assistant if this step is unclear to you.

Quick sum feature

It would be a good idea to verify that our percentages add up to 100%. There is a quick and easy way to do this without entering a formula. Highlight all of your percentages. Look at the bottom of the Excel window. In the status bar, it should read, "Average 12.5% Count: 8 Sum: 100.0%". This is Excel's Quick Sum feature. Anytime you highlight several cells containing numbers, Excel gives you some summary information about them, without you having to enter a formula into some cell. This is a very convenient feature to do some quick calculations that you don't really need to store explicitly in your spreadsheet. Once you click somewhere else in the worksheet, these displayed values will disappear.

As another example, suppose you quickly wanted to know how many students were taking languages other than Spanish. You could simply highlight cells B4 through B10, and look to the bottom of the Excel window for your answer.

In your own words, what do the figures and percentages in this spreadsheet tell us? What are the two most important conclusions you would like to draw from the data? _____

Before we continue, I want to mention something else about percents. Occasionally you may wish to compute a percent but not show the '%' sign. In this case, you would put a '%' character at the end of your formula, and this would have the effect of multiplying your expression by 100. So 0.3 would be shown as 30, without the '%' sign.

Example #5 – checkbook

A friend of yours needs help balancing her checkbook. At the beginning of the month, she had \$ 80 in her checking account. During the month, she made the following transactions.

June 5, Check #101, Target, \$24.87
 June 7, Check #102, Shell gas, \$22.85
 June 8, Check #103, cash, \$25
 June 11, deposit pay check, \$160
 June 13, Check #104, Ingles, \$15.67
 June 17, Check #105, Dillard's, \$39.53
 June 25, deposit pay check, \$195
 June 30, Check #106, Ingles, \$6.29

Let's enter this information into a spreadsheet so that we can see a running balance in her checking account. This time, I'll show you another way to create a new spreadsheet. Open Excel from the Start menu. (Start → All Programs → Microsoft Office → Microsoft Office Excel 2007). Once Excel is open, click on the Windows logo button at the top left corner of the window. Select New → Blank Workbook. Now, a new spreadsheet is ready to begin. But we need to save it under a specific name. Click on the Windows logo icon again, and Save As this file name: checkbook.xlsx. (Make sure this file is located in the proper folder.)

In row 1 of this spreadsheet, we will enter column headings. In cells A1 through F1, enter these headings: "Date", "transaction", "description", "add", "subtract", "balance".

In row 2, we will start the checkbook with our original balance. Enter 80 in cell F2.

Starting in row 3, enter the individual transaction information that was given above. Fill in the appropriate columns A-E. For checks, enter the amount under the subtract column. For deposits, enter the amount in the add column. Leave column F blank for now.

Now, we are ready to compute the running balance for the checking account. Click in cell F3. Enter this formula: =f2+d3-e2. In other words, we are adding and/or subtracting an amount from the previous line in the checkbook. Copy the formula from F3 all the way down column F. The last number in the column would give the current balance in the account.

The nice thing about having a running balance is that we can see that the account is never overdrawn. We could also use it if we wanted to compute the daily interest earned. What was the lowest balance the account reached? _____

This is our first spreadsheet that has featured a date. We can format dates in Excel. Highlight all the dates in column A. Right click, and select "Format Cells". Under the Number tab, choose the Date category. Off to the right, we can specify exactly what date format we want to see. Make sure the first type is selected, the one showing the example as 3/14/2001, because we also want to show the year.

Example #6 – cold weather

How cold is the North Pole? The closest weather station to the North Pole is the Canadian Forces station in Alert, located near the north end of Ellesmere Island at 82 degrees north. Open the spreadsheet [alert2005.xlsx](#), which contains a year's worth of temperature data for Alert. The data is in Celsius, so the first thing we'll do is convert the high and low temperatures to Fahrenheit. Then, we will calculate the averages of the high and low temperatures.

The column headings are in row 1. In cells D1 and E1 we want to indicate that these columns are going to hold the Fahrenheit high and low temperatures. For example, you could call them "max (F)" and "min (F)" for brevity.

In cell D2, we need to enter a formula that converts the Celsius temperature in B2 into Fahrenheit. Enter this formula: $=9/5*b2 + 32$. Copy the formula from D2 to E2. Next, highlight both cells D2 and E2. *Double click the fill handle*. This is a nice trick: double clicking the fill handle copies your formulas all the way down the spreadsheet as needed. Now, in row 368, we'll enter formulas to compute the average high and low temperatures. Click in cell D368. Ask Excel to average the values in cells D2 through D366. Copy this formula into cell E368 as well.

Findings: What are the average temperatures for Alert? _____
Brrrrr! You can see why Santa Claus decided to move to Montreal.

Summary and concluding thoughts:

- Data is arranged in cells, which are organized by row and column.
- A spreadsheet file may consist of multiple pages (worksheets).
- It's often useful to adjust the width of our columns.
- It's desirable for a spreadsheet to be long rather than wide. Over time, we tend to add data to a spreadsheet. You should plan its design so that the growth is vertical rather than horizontal.
- When we insert or delete columns, formulas change automatically to reflect the new position of the data.
- In general, it's better for a spreadsheet to have more rows than columns.
- Formulas often use built-in functions like "average".
- In general, we should not type literal constants in formulas, because this makes editing later more tedious.
- Cell ranges
- Usual rules of thumb: Column headings should be centered. Other text should be left justified. Numbers should be right justified. Excel usually handles this by default.
- Wrap around text

- Double-clicking the fill handle will copy a value/formula down the column until Excel detects a break in the data, such as a blank line.
- For displaying money, often we don't want to show cents if they are always .00, or if the monetary values are so large that the cents are insignificant.
- If you are navigating a spreadsheet and you inadvertently start modifying a formula, hit the ESC key to cancel the editing. If the errant change has already been entered, you can click the "undo" button or hit CTRL-Z.
- Anytime you encounter a feature that you are not sure about in Excel, you can use the Help utility. This is the question mark inside a blue circle located near the top right corner of the Excel window. If you are connected to the Internet, Excel can find additional help topics from Microsoft.

Review Questions – Ideally, you should be able to answer these questions without looking at Excel. But feel free to use Excel to check your answers.

1. What are the two ways to format money? Name one advantage that each has over the other. If you are not sure, go back to a spreadsheet that had monetary values in it, and check out the two possible formats.
2. Write a cell range that contains 3 rows and 3 columns, where the first cell is in the fourth row and fifth column of the spreadsheet.
3. How many cells are in the range Z5:BB14 ?
4. Suppose that cell H10 contains this formula: =sum(c2:f9). What will happen to this formula if we insert a new column between A and B? What about a new row between 5 and 6?
5. How could we simplify this formula? $=(c2+c3+c4+c5c6+c7+c12+c13+c14+c15+c16+c17)/12$
6. Suppose a formula's result is 0.555555556. What would you have to do in Excel so that this result instead looks like "55.6%"?

Excel lab #2

In this lab we will look at absolute versus relative and mixed references, plus some more functions.

Remember that when we work with spreadsheets, our data resides in cells that are arranged in rows and columns. We can type *values* or *formulas* into the cells. Formulas are extremely useful in that they perform computations, and they automatically update when you change data, insert new rows, etc. Many times a formula utilizes a built-in Excel function. For example, the formula =sum(a1:a4) is a lot easier than =a1+a2+a3+a4 because it's more concise and less error prone to enter.

Example #1 – loan payments

Many Excel functions are a little more sophisticated than sum. For example, consider the PMT function, which determines loan payments. To use this function requires 3 pieces of information: the interest rate, the number of payments, and the amount of money borrowed. If you think about it, this should make sense: if you wanted to calculate a loan payment, you would really need to have all of this information. If you only knew that you borrowed \$ 5,000, that single fact would not be enough. Because some functions like PMT require several parameters, it is not convenient to have to remember the *order* in which these parameters should be specified. It turns out there is an easy way to set up our function expression.

"Insert Function" command

Let's set up a simple spreadsheet for calculating a loan payment. Open Excel from scratch. (Start → All Programs → Microsoft Office → Microsoft Office Excel 2007). When the blank spreadsheet appears, save it as the name loan.xlsx. At the top of the Excel window, click on the Formulas tab. In the ribbon, notice that the first button on the left is called "Insert function." We'll be using this shortly.

Enter the following raw data about our loan:

A1 = "Amount to borrow", B1 = 75000

A2 = "Years", B2 = 15

A3 = "Interest Rate", B3 = 6

A5 = "Loan payment is"

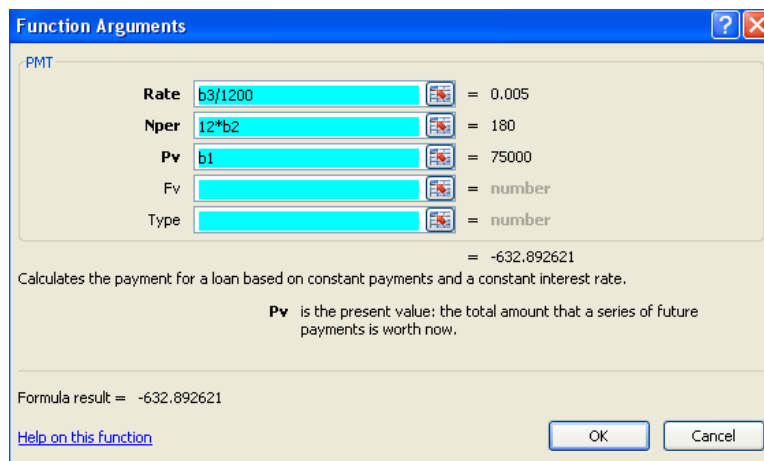
Make sure to widen column A so that all of the words are visible. The cell B5 is where we want to enter our PMT function to calculate our loan payment. This function is too detailed for us to memorize the exact format of the input, so let's use Excel's Insert Function command. Select cell B5. Then, at the top of the Excel window, from the Formulas tab, select the Insert Function button. A dialog box appears. The first text box invites us to enter a description of the type of function we want to use. Fortunately in this case we already know the name of the function we want: PMT. Enter "PMT" in the text box, click Go, and you'll see PMT listed in the suggested list of functions below. Make sure PMT is selected from the list, and click the OK button.

A new dialog box appears where we enter the PMT function arguments – in other words the input values or parameters. We need to enter the 3 necessary pieces of information for determining the loan payment. (Note that the 4th and 5th boxes will be left blank: they would only be needed in special loan situations.) It is essential to enter the input values correctly, in the format that the function expects. Pay close attention to how we enter the information.

Select the first box. It's called Rate. Near the bottom of the dialog box there is a description of what this means: "Rate is the interest rate per period for the loan." There is an important detail here: Excel wants the interest rate to be quoted per payment rather than annually. Most loans are monthly, so we need to divide the interest rate by 12. Also, note that the interest rate has to be entered as a percent. The number 6 would literally mean 600% percent interest, which is not correct. Our interest rate is in cell B3: we need to divide this number by 100 to convert it to a percent, and we need to divide by 12 to convert to months. Thus, we need to enter $b3/1200$ in the Rate box. Note that off to the right side of the Rate box it says "= 0.005" which is the monthly interest rate.

Select the second box, called Nper (number of periods). This is the number of payments of the loan. Cell B2 is the number of years, and we are going to pay monthly, so enter $12*b2$ in this box. Note that off to the side Excel responds with the total number of payments, 180.

Finally, select the third box, called Pv (present value). This is the amount of money we're borrowing. This value is in cell b1, so type b1 in this box. Excel confirms that it's 75000. At this point, your dialog box should like the following picture.



Excel has enough information to determine the answer, and if you look along the right side of the dialog box it displays a value: -632.892621. Later we are going to format this to 2 decimal places, but there is another problem: this number is negative. The reason why is because you are "losing" this much money per month: it's a debit rather than income. It's counterintuitive to tell people that their loan payment is negative. We can eliminate this problem by making the present value itself negative, so change the 3rd input box to -b1. Now our answer is positive. Click the OK button, and our formula will appear in cell B5.

Here are some finishing touches: select cells B1 and B5, and format them using the Accounting style with 2 decimal places. Format cell B3 as a percentage also to 2 decimal places.

Select cell B5 once more, and look at the formula bar. It shows that the formula entered for this cell is $=\text{PMT}(b3/1200, 12*b2, -b1)$. As you can see, this formula is rather complex, and the order in which the values are specified seems arbitrary to us. Rather than memorizing how to type in the loan payment function, it's a lot easier to use the Insert Function method that we just did. On the other hand, for a simple function like sum or average, it would have been reasonable for us to remember how to type that in. We generally use the "Insert Function" feature for the more complex functions, or when we cannot remember the name of a function.

Save and close loan.xlsx.

Mixed references

Many times in Excel we enter a formula, and then we want to copy this formula to other cells, usually to other cells down a column or across a row. For example, column B may contain the prices of some merchandise, and we want to add sales tax to these and put the resulting amounts in column C. We would type in one formula at the top of column C, and then copy this formula down the rest of the column. However, a formula usually contains references to other cells in the worksheet, and as we copy our formula down, we have to consider what is happening to each of those references. You saw this happen in the foreign language spreadsheet example we did in the previous lab. Take a moment to review what happened in that example.

There are 3 kinds of cell references in Excel: relative, absolute and mixed. A relative reference is by far the most common: this is when you simply type the column and row of a cell, as in E3 in a formula. If you copy this formula down to the next row of the spreadsheet, E3 will become E4 in the new formula. Also, if we had copied the formula to the next column to the right, E3 would become F3 in the new formula. And if we copy the original formula to a new cell that is 4 rows lower and 3 columns to the right, the E3 would become H7. Most of the time, this is the behavior we want. But as you saw in the previous lab, sometimes we have values in a formula that should not change. These constants need to be specified as absolute references because we don't want them to move. We would use dollar signs in front of the column and row like this: $\$E\3 . If you use an absolute reference, it will never change no matter where you copy the formula to. $\$E\3 will stay $\$E\3 . Absolute references are used for constant numbers we want to apply to all formulas, such as an exam weight, a minimum criterion value, a rate of tax, an equal bonus to give employees, a rate of commission, or some other constant.

On the other hand, a *mixed* reference is a cell reference that uses only one dollar sign, such as $\$E3$ or $E\$3$. Note that the dollar sign indicates what should not be changing if the formula gets copied. In the case of $\$E3$, we are saying that if we move the formula, the E stays fixed but the row number may change. Conversely, $E\$3$ means the 3 stays fixed and the column letter may change.

Example #2 – calories revisited

Open the file calorie.xlsx that we worked with previously. We'd like to know what percentage of each day's calories came from each meal. Let's place this information below the daily calorie totals. First, we need to make room for our percentages. If you haven't done so already, remove the To-do list at the bottom of the spreadsheet. Now, rows 8-10 should be blank.

In cells A8, A9 and A10 enter these row headings: "% breakfast", "% lunch" and "% dinner". Our percentages are going to reside in cells B8 through F10. Let's start with B8. Here is where we want the percentage of Monday calories that came from that day's breakfast. In cell B8, enter the formula =b2/b6. Format this cell to have the Percentage format with 0 decimal places.

What should we do next? We need to copy the formula in B8 to the other cells in B8:F10. We can copy this formula down to B10, and then copy all 3 formulas across through the Friday column. If you copy the formula in B8 to B9 and B10, you notice we have an error. As we copied the formula down, we should not have changed the reference to the total calories, B6. So, return to the formula in cell B8, and change it so that B6 is specified as an absolute reference, =b2/\$b\$6. Then, copy this formula again down to B9 and B10. The three percentages should be 26%, 31% and 43%.

Now we are ready to copy the Monday formulas to the rest of the week. Highlight cells B8 through B10. Drag the fill handle to the right until you are in column F, and release the mouse button. We have no error message, which is good, but on close inspection the percentages look funny. If you briefly highlight cells B8 through B10, and look at the Quick sum at the bottom of the Excel window, you can verify that these percentages add up to 100%. But do the same for Tuesday: highlight cells C8 through C10. This time the meals add up to only 86%. Do you know what's wrong?

The mistake is that we used an absolute reference when we should have used a mixed reference. Look at the formula in C8. It says that we divided the Tuesday breakfast calories by the total for Monday. As we had copied the formulas to the right, the reference \$B\$6 was not changing at all. Think for a moment what we actually wanted to do: we want the reference for B6 to stay the same for the various meals of the day (the rows), but we need it to change when we switch to a new day of the week (columns). This means that the dollar sign should only go in front of the row number: b\$6. Update the formula in cell B8 once again, and copy the formulas to the rest of the spreadsheet. Now, the column percentages should each add up to 100%. A mixed reference saved the day!

Save and close calorie.xlsx

Example #3 – multiplication table

Suppose a young relative of yours is in elementary school and is learning the multiplication table. You can use Excel to systematically create one of any size desired.

We'll start with a new blank workbook. In your folder of Excel files, right click in the background, and tell your system that you want to create a new Excel workbook called mult table.xlsx. Open this file, and Excel will automatically be invoked. In this blank spreadsheet, the first thing to do is to set up the values for the rows and columns. The cells A2 through A13 need to show the numbers 1-12. Rather than typing all 12 numbers, here is a shortcut using the fill handle: Type 1 in cell A2 and 2 in cell A3. Highlight both of these cells, and drag the fill handle down to cell A13. Voila! The fill handle can take an established pattern of values and extrapolate as far as you need. Similarly, let's also do the column headings. Type the numbers 1 and 2 in the cells B1 and C1. Highlight these two cells, and drag the fill handle to the right to column M. Now our multiplication frame is set up. Note that we don't need to enter anything into A1. The entries in the multiplication will occupy cells B2 through M13 (i.e. 12 rows and 12 columns). To improve the table's readability, let's use color to highlight the row and column headings. Highlight A2:A13 and B1:M1 and from the Fill Color button in the ribbon (which looks like a paint bucket), select light green from the swatch.

These numerical row and column headings we just entered are going to be useful for us, because we are going to multiply them to give us the entries in the multiplication table. Let's just take one random interior cell as an example. D6 is in the row we labeled 5 and the column we labeled 3. Thus the number in this cell must be $5 * 3 = 15$. Take a careful look at where the "5" and "3" values are. The formula for this cell would need to multiply A6 times D1. So, we can say $D6 = A6 * D1$.

To figure out what kind of mixed references we need for this formula, let's consider what the answers should be in the cells below and to the right. Let's go down one cell. D7 should contain the value $6 * 3 = 18$. So we need to multiply $A7 * D1$ to give us D7. Notice that as we "copied" the formula down the only row number in our formula that changed was the A6 changing to A7. *The "1" in D1 did not change.* Next, consider going to the right: in cell E6, we need to multiply $5 * 4$, which is $E6 = A6 * E1$. Compare this formula with that of D6: what changed? When we "copied" the formula to the right, the D1 changed to E1 but *the "A" in A6 did not change.*

The parts of the cell references that do not change get a dollar sign in front of them, so we are ready to try out a mixed reference formula. Earlier, we said that $D6 = A6 * D1$. Let's put in our dollar signs, and the formula becomes: $D6 = \$A6 * D\1 .

To complete the spreadsheet, we're going to enter a formula in cell B2 and copy it to the rest of the spreadsheet. We'll use the above discussion as our motivation for writing our formula. We found that the "A" in the row heading shouldn't change, and the "1" in the column heading shouldn't change. Thus, the formula for B2 is: $=\$a2 * b\1 . Use the fill handle to copy this formula down to B13. Then, highlight cells B2 through B13, and drag the fill handle to the right to column M to finish the table.

The placement of the dollar signs in this example is important. If we were careless and arranged the mixed references the other way, as in $B2 = a\$2 * \$b1$, the result would be completely wrong!

Save and close mult table.xlsx.

Example #4 - Loan table

Suppose a friend of yours is interested in getting an adjustable rate mortgage (ARM) for a new house. This type of mortgage often begins with a low interest rate, but over time this rate rises and fluctuates with the market. We would like to create an expanded version of our earlier loan spreadsheet, this time showing how the monthly payment varies based on the amount of the loan and especially the interest rate. This spreadsheet will practice with mixed references.

Open Excel from scratch. When the blank spreadsheet appears, save it as the name loan table.xlsx. In cell A1, let's enter a title for the spreadsheet: "Loan payment by interest rate and original balance". The structure of this spreadsheet will be similar to our earlier multiplication table. The first thing to do is to specify the length of the mortgage. In cell A3, enter "Number of years", and in cell C3 enter the number 30. Next, we need to set up row and column headings. We would like to simulate various interest rates from 3 to 10 percent, in increments of 0.5 percent. In cell A6 enter 3.0% and in cell A7 enter 3.5%. Highlight these cells and use the fill handle to copy this pattern down until you reach cell A20. Release the mouse button, and you should see an increasing sequence of rates up to 10 percent. Now, let's do the column headings. We'll consider possible loan sizes of 100,000 to 300,000 in increments of 50,000. Enter the values 100000 and 150000 in cells B5 and C5, respectively. Highlight B5 and C5, and drag the mouse to the right until you reach cell F5 at which point the value becomes 300000.

Before proceeding, we should format our cells. Notice that the percentages are expressed to 2 decimal places. We would like just 1 decimal place. Here is a quick way to change this: Highlight the percentages in column A. Then, find the ".00 → .0" button in the ribbon at the top of the Excel window. If you don't see this button, make sure you have the Home tab highlighted. You may need to widen the Excel window to see all the buttons. The ".00 → .0" button has the effect of reducing the number of decimal places by one digit. (Right next door to that button is a similar button ".0 → .00", which would increase the precision.) Next, we'll format the monetary figures: Highlight the loan amounts in row 3, and format them as Accounting style with 0 decimal places, as we have done in previous labs.

Now, we are ready to enter our loan formulas. Click in cell B6. At the top of the Excel window, select the Formulas tab, and click on "Insert function". Follow the procedure we did earlier to obtain the PMT function, and enter the appropriate values for a 30 year loan for \$100,000 at 3%. When you are finished entering this formula, it should give a result of \$421.60.

We would like to copy this loan formula in B6 to all the other cells in our loan table. But first, we need to set up the mixed and absolute references correctly. Take a look at the formula you just entered in B6. It contains the cell references A6 (the interest rate), C3 (the length), and B5 (the amount borrowed). Recall that an absolute reference is something that should not change no matter where you are in the table, any row or column. Which of these three quantities is something that is always the same? This becomes your absolute reference. Next, we need to think about mixed references. Is there a cell reference in our formula that should only change its row number or only change its column number as we copy the formula? This would become our mixed reference, and the part of the cell reference that

does not change needs to get a dollar sign placed in front of it. Manually insert dollar signs in the formula in B6 so that the absolute and mixed references make sense to you.

We can now copy our loan formula. Select B6, and drag the fill handle down the rest of column B. Then, highlight all the formulas in column B (B6:B20), and drag the fill handle to the right until you reach column F. When you are done, you should see \$2,632.71 in the lower right corner of your table at cell F20. If not, then you need to check your work: it's possible some of your mixed references are not correct and need to be changed.

Finally, let's add some color to the spreadsheet to enhance the readability of the loan table. We'll color in the row and column headings. Highlight cells A5:F5. *Hold down on the control key*, and also select cells A6:A20. Release the control key and the mouse. This technique of using the control key while selecting cells is very useful when you want to select cells that are not arranged in a rectangle, or when they are not contiguous. Now, while these cells are selected, let's choose a color. Make sure the Home tab is selected at the top of the Excel window, and look for the *Paint Bucket* button. Click the little triangle next to the paint bucket. This will bring up a color swatch. In the bottom row, click the green square. If you're not sure what colors you see in the swatch, you can hover the mouse over a square for a second, and the name of the color should appear. When you click on the green square, the cells you had selected should all turn green.

Save and close loan table.xlsx.

Some statistical functions

There are several simple statistical functions that we sometimes use in Excel. In particular: count, counta, countif, max, min. The first three have similar names, so it's important to understand how they differ. These functions are used on a range of cells, for example =count(a1:b8).

count: This function determines how many of these cells contain a number, or a formula that evaluates to a number.

counta: This function returns how many cells contain either a number or text. In other words, how many cells are not blank.

countif: This function tells you how many cells fulfill some condition you specify. For example, you may want to know how many cells in the range are greater than 50. Unlike the other two count functions, this one takes a second parameter, which is the condition inside double quotes. The following table shows you what the second parameter to countif should be based on what you are trying to count:

Counting what?	2 nd parameter should be...	Examples of 2 nd parameter
Instances of text	Cell reference or literal text	A19
		"Miami"
Instances of a number	Cell reference or literal number	A24
		250
Comparison	Inside quotes: relational operator and a number (constant)	"= 100"

For example, the expression =countif(a2:b15, "Miami") will return the number of cells in the range A2:B15 that contain exactly the word "Miami" and nothing else. *Regular expressions* can also be used. For example, =countif(a2:b15, "M*") will tell you how many cells contain text starting with the letter M. These comparisons are not case sensitive. Note that there are two ways Excel allows us to specify that we want to match a specific constant: the 2nd parameter can be either a constant by itself or it can include the equals sign (see the 4th and 5th examples in the table).

The max and min functions are straightforward: they return the largest or smallest value in the given range of cells, respectively.

To illustrate the differences among the count functions, consider this example spreadsheet:

	A	B
1	Item	Price
2	T-shirt	8.75
3	key lime pie	out of stock
4	Grapefruit juice	2.78
5	car	25000
6	moon	
7	paper clips	?

Here, =count(b2:b7) will return 3, because there are 3 numbers in the range. On the other hand, the function =counta(b2:b7) will return 5, because counta includes the cells containing text. Finally, if we have =countif(b2:b7, ">5"), this will return 2.

In addition to countif there is a similar function sumif which adds instead of counts the cells that satisfy the condition. For the above example, =sumif(b2:b7, ">5") would return 25008.75. There is also an analogous function *averageif* that does the same thing except it performs an average rather than a sum.

The “if” function

One of the most important functions in Excel is the IF function, which allows us to ask questions about our data and make immediate decisions. Like the PMT function, IF requires 3 parameters. The general format is:

=IF (question, true-value, false-value)

The first parameter is some yes-no question, for example if the value of some cell is positive. The second parameter is the result we want if the answer to the question was yes, and the third parameter is the result we want if the answer to the question was no. For example, we may give a set of job candidates an exam. To pass the exam, a candidate must obtain a score of 75 or higher. We can express this idea using an IF function as follows, assuming that B2 contains an examinee’s score:

= if (b2 >= 75, “pass”, “fail”)

The outcome of this formula is either the word “pass” or “fail” depending on the value contained in cell B2. As another example, suppose you have a commission schedule in which your commission on a sale is 2% if the sale amount is under \$1,000, but 3% if it’s \$1,000 or higher. Your commission formula could be written as follows (assuming B2 is the sale amount):

= if (b2 < 1000, .02 * b2, .03 * b2)

It is also possible to nest if-functions (or any functions) inside each other. In our commission example, if also we earn 4% for sales at or above \$5,000, our formula becomes:

= if (b2 < 1000, .02 * b2, if (b2 < 5000, .03 * b2, .04 * b2))

Note that there is more than one possible way to write if-functions. For example, if we reverse the “b2 < 1000” to “b2 >= 1000” we would also swap the true and false values that follow.

Example #5 – climate data

Let’s apply our knowledge of Excel functions to a spreadsheet containing climate data for one month. Open the file [climate.xlsx](#). We will basically follow the To-do list at the bottom of the page. The first thing we’ll do is some tidying up. This spreadsheet needs to have meaningful column headings. Click in any cell in row 3, and ask Excel to insert a new row above. Now we have a new row 3 where we’ll enter our column headings. Type the following words into cells A3 through D3: “Day”, “High”, “Low”, “Rain”. Select cells A3:D3 and center them using the “center text” button on the ribbon.

One minor problem this spreadsheet has is that it is so long that we cannot see the data all at once. We would need to constantly scroll up and down to see the column headings as well as the data at the bottom of the page. There are 2 ways we can fix this problem. In the lower right corner of the Excel window, there is a *Zoom tool*, which by default is set to 100%. You can use the – or + buttons to change

the zoom percentage. For example, hitting the – button will zoom to 90% size. In general, however, this procedure won't always be feasible. If we had a whole year's worth of data, you would not want to shrink down to the point you can't see the data. So, there is another handy approach to keeping column headings on the screen. It's called *Freeze Panes*. We want to tell Excel to keep rows 1-3 constantly on the screen whenever we scroll down the page. Click in cell A4. At the top of the Excel window, select the View menu. In the Window group of commands, select Freeze Panes → Freeze Panes. At this point, you should see a thick horizontal line underneath row 3, indicating the boundary of what stays on the screen. (This line is not visible if you print the spreadsheet.) Note that in general, when you click in some cell and choose Freeze Panes, rows above and columns to the left will stay on the screen. Thus, it's possible to keep columns on the screen if you scroll right. However, in practice we will rarely have spreadsheets so wide that we often need to scroll left and right. If you ever want to turn of the freeze panes feature, just select the Freeze Panes command again.

The second item on our To-do list is to compute the average high and low temperatures for the month, and also the extremes (max and min high and low temperatures). First, let's make some space. Click in any cell in row 36, and ask Excel to insert 6 new rows. (i.e. insert row, 6 times) In cells A36 through A38, enter the following words: "Average", "Max", "Min". Let's now look at the set of high temperatures given in the range of cells b4:b34. In cells B36 through B38, enter the appropriate formulas to find the average, max and min values in the range b4:b34, respectively. Once you have done that, copy these 3 formulas together into column C. Finally, for the rain, we want the total precipitation, not the average, max or min. Enter the word "Total" in cell A39. In cell D39, enter the appropriate formula that finds the sum of all the numbers in d4:d34.

Excel has a little shortcut to help you entering formulas that employ cell ranges. You can use the mouse to select cell ranges instead of typing them literally. For example, to enter the formula =max(b4:b34), you can type "=max(" and then use the mouse to highlight the cells b4:b34. Excel will automatically translate what you've highlighted into a range inside your formula. Then you can complete the formula with the ")" and hit enter. This mouse feature is convenient when you have a large range, and you are not sure exactly what the endpoints are yourself.

Before continuing, let's format these values we just computed. The temperature data should be whole numbers, so format them as numbers with no decimal places. In your own words, what do the numbers in cells B38 and C37 mean? _____

It turns out the amount of heating and air conditioning you need is tied directly to the average daily temperature. We'd like to measure that in this spreadsheet. First, we need to insert 3 new columns to accommodate our temperature-related data. Click in any cell in column D, and ask Excel to insert 3 new columns to the left. (i.e. new column, 3 times) Type in these column headings in cells D3, E3 and F3: "Average", "Heating", "Cooling". In order to measure how much relative heating and cooling we need, we first must find the average temperature per day, i.e. the average of the high and low temperature. In cell D4, enter a formula that finds the average of cells b4 and c4.

Engineers have found that in general you only need to heat a building when the average temperature outside is less than 65. Similarly, you only need A/C when the average temperature is above 65. Thus, when we figure how much heating and cooling we need, we will first use an IF function to ask if the temperature is in the appropriate range. Let's consider heating first. If the temperature is 50 degrees, you'll need more heat than if the temperature is 60 degrees. When we need heat the temperature is below 65, so the amount of heat we'll need is 65 minus the temperature. If the temperature is above 65, the heat amount will be 0. Therefore, the general idea for heat is:

$$\text{Heating} = \text{if}(\text{temperature} > 65, 0, 65 - \text{temperature})$$

Click in cell E4, and enter the appropriate heating formula, given that the temperature we want to compare is in D4.

The situation for air conditioning is analogous to heating. If the temperature outside is less than 65, the amount of cooling we need is zero. Otherwise, the higher the temperature, the more A/C you need. Thus our general idea for cooling is:

$$\text{Cooling} = \text{if}(\text{temperature} < 65, 0, \text{temperature} - 65)$$

Look carefully at our two formulas for heating and cooling. They differ in two ways: the comparison to 65 is either > or <, and the subtraction of 65 and the temperature goes in alternate ways. Be sure you can reason why these formulas are correct. For example, you should be able to explain why these formulas will not output a negative number. In cell F4, enter the formula that will find how much cooling we would need on the first day of the month. Then, select cells E4 and F4, and copy these formulas down to row 34. At the bottom of the spreadsheet, it would be nice for us to have totals of the heating and cooling amounts. Select cell G39, and use the fill handle to copy this formula *left* to cells E39 and F39.

In agriculture, farmers are concerned about freezing weather. We would like to count how many mornings experienced a low temperature of 32 or less. In cell A40, enter "Lows <= 32". In cell C40, we'd like to use the COUNTIF function to ask how many values in the range C4:C34 are less than or equal to 32. Enter this formula: =countif(c4:c34, "<=32"). Is the answer what you expected?

Conditional Formatting

Let's experiment with conditional formatting, an excellent way to make your spreadsheets more readable. There are two common ways to highlight values in a spreadsheet. First, we may want to use color or some other special effect on just some extreme values in a list. Another way is to use a *color spectrum* to a set of data. The color spectrum is a new feature in Excel 2007. Let's experiment with each technique.

First, we'll single out some of the warmest high temperatures of the month. Highlight cells B4:B34. From the Home menu, look for the Styles command group, and choose Conditional Formatting →

Highlight Cells Rules → Greater Than. When the “greater than” dialog box appears, enter 80 in the first box, and select the format option “Yellow Fill with Dark Yellow Text”. Click OK, and you should see that the hottest temperatures of the month, those over 80 degrees, have been highlighted. There are many possible formats you can choose for conditional formatting. Repeat the above steps, but instead of the Dark yellow text, this time choose “Custom Format...”. When the Format Cells dialog box appears, click the Fill tab, and select the dark red color at the left end of the last row of squares. Next, click on the Font tab, and in the “Color” section in the middle of the box, choose the bright yellow square along the bottom row of squares. Now you have specified a bright yellow text color on a dark red background, which will be easy to read.

The second technique we will try is the spectrum, which I think is more visually informative, and also easier to use. Select cells C4:C34. In the ribbon, click on Conditional Formatting → New Rule. A new dialog box called “New Formatting Rule” appears. In the bottom part of this box, there are two columns of controls, one for the minimum value of the range and one for the maximum value. At the bottom of each column is a control for setting the color code for each extreme. Under “Minimum”, go to the control labeled Color, and click the downward pointing arrow button to bring up a swatch. Select the color blue, which is the third square from the end of the last row. Next, under “Maximum”, click the downward pointing arrow next to the Color control, and from the swatch, choose the bright yellow square. Once you’ve selected your colors, the bottom of the dialog box shows you what your color spectrum looks like as a preview. Click OK, and now look at column C! You can instantly see the overall trend of temperatures during the month, which days were warmer or cooler. I find this color spectrum to be a very useful feature of Excel.

We’ve turned a hum-drum spreadsheet into something useful with lots of calculations and information. Since we’re done with our To-do list, you can get rid of it if you’d like. Save and close climate.xlsx.

Example #6 – Job candidate interviews

I’d like to show you an example of how mixed references and Excel functions can be useful in decision making. Open the file [interview.xlsx](#). This spreadsheet shows the raw data from a hypothetical round of job interviews at a company. There were 8 candidates, and 6 current employees conducted the interviews. It turns out that each candidate had an appointment with several people at the company. Each interviewer rated the candidate on a 1-10 scale encompassing traits such as technical competence and enthusiasm. In this spreadsheet, we will do some summary computations now that the interviews are over.

The raw data can be found in the first green section, cells B4:I9. In general, it’s not practical for every job candidate to be examined by every interviewer, so some of the cells are blank. Let’s compute some formulas: In row 10 we need to compute the average interview score earned by each candidate. Enter the appropriate formula in cell B10, and copy it to the right. Similarly, we are also interested in the ratings given by each of the interviewers, and these averages will go in column J. In cell J4, enter a

formula that will show Jacobs' average rating given. Copy this formula down the column through Patterson.

At this point, we can immediately determine a result. Which candidate(s) earned the highest overall interview rating? _____

We can also answer this question: which interviewer(s) tend to give the highest scores?

Because not all the interviewers have the same standards for how they rate on a 1-10 scale, and not all interviewers see every candidate, it may be a little misleading to go by simple averages. So we have a little more work to do. You should notice that we have a tie for the best candidate! Hopefully we'll be able to break this tie.

In the second part of the page, we want to re-evaluate the candidate scores, and enter the adjusted results in the 2nd green section in cells B17:I22. The comments on rows 13 and 14 give us the motivation for this calculation. For example, two candidates may have the same average, but it could be the case that one candidate was assessed by tougher interviewers. As line 14 suggests, we want to compute *how well each candidate did relative to each interviewer's average*. This may be difficult to understand until we look at an example. Consider the top left corner, Jacobs interviewing candidate #1. The score was a 9, and Jacobs' average was 8.8. Thus, candidate #1 beat Jacobs' average by 0.2. So, in cell B17, we can enter this formula: =b4-j4. We'd like to copy this formula to the other cells in the green region. However, we need to use a mixed reference somewhere. In other words, we are missing a '\$' somewhere in the formula =b4-j4. What should the formula be? _____

Copy the formula, and see if the results make sense. To check your work, the last entry, at cell I22, should be 2.5 because candidate #8's score (10) exceeded Patterson's average (7.5) by that amount.

In row 23, we can compute each candidate's average. In cell B23, enter a formula that averages cells B17 through B22. Copy this formula to the right through column I. In your own words, what does this average mean? _____

Allowing a formula result to be blank

The arithmetic is right, but we have another problem. What about cells where there was no interview? For instance, candidate #1 never saw Nijs. There should be no number in cell B19. Instead we have a number that makes no sense. We need to redo our formula earlier so that we produce a blank result if there was no interview. We'll use the IF function here. Essentially we want to say:

=IF (there was an interview, subtract score from interviewer's average, "")

Notice that the false-value in the IF function is "". This is called the *empty string*. We want to replace irrelevant numbers with blanks so that they will not get averaged.

Now, how do we know if an interview actually took place? Candidate #1 was interviewed by Jacobs but not by Nijs. Looking at the original green section, what is the difference between cells B4 and B6? B4 contains a number and B6 is blank. As we saw earlier, there is a way in Excel to see if cells contain a number: we can use the COUNT function. The expression COUNT(B4) will equal 1 because there is a number in B4. But COUNT(B6) will equal 0 because it's blank. Thus, our formula for cell B17 should look something like this:

=IF (count(b4) > 0, our earlier subtraction formula with mixed reference, "")

The second parameter to this IF function should be the formula you previously had in B17 which contains the mixed reference. Modify B17 so it now uses the IF function. Then, copy this formula to the other green cells in B17:I22. Several of the cells should now become blank, while the other cells, corresponding to bona fide interviews, should be unchanged. As a result, the averages reported on row 23 should be more accurate. Taking into account the numbers on row 10 and row 23, which is the best candidate? Explain: _____

The company management is interested in one more aspect of the interview process. Which employee gives the highest overall scores? At first, we might think the answer can be immediately determined from cells J4:J9. But that could have been a matter of luck. The person who gave the highest average score might have been lucky enough to interview the best candidates. We should do an analysis of the interviewer scores just like we did for the candidates. This is what we will do in the 3rd green section of the spreadsheet.

The idea for evaluating the interviewers is given on lines 26-27. Let's look at the entry in cell B30 as an example. Here, Jacobs awarded a score of 9 to candidate #1. But candidate #1 overall scored only 6.0. Therefore, we can say that Jacobs gave a score that was 3 points higher than #1's average. So, in this green section we want to calculate *the difference between the score given by the interviewer and the corresponding candidate's average*. In cell B30, enter the formula =b4-b10. Once again, this formula needs to be modified: we need to create a mixed reference by adding a '\$' somewhere. What is the correct formula? (Hint: consider how the formula is going to be copied) _____

Copy this formula to the other cells in B30:I35. To check your work, you should find that in the lower right corner, cell I35, Patterson gave candidate #8 a score that was 1.8 higher than #8's average.

Just as we saw when we were doing the candidate averages, we need to eliminate nonsense values in this table of interviewer averages. Go back to cell B30, and enclose it inside an IF function. If there was no interview, again we should show "". Copy the new formula to the other cells in B30:I35. Finally, in cells J30:J35, compute each interviewer's average. In your own words, what do these values mean? For example, what does it mean for Nijs to have an average of 0.4 here?

Finally, we can address management's question: who are the easiest and toughest interviewers, based on the numbers in cells J4:J9 and J30:J35? Explain:

Save and close interview.xlsx.

Example #7 – Tic-Tac-Toe game

Now, let's have some fun and play a game. With the Excel functions we have already seen, especially "if" and "countif", we can implement a game like Tic-Tac-Toe in Excel! Open the file [ttt.xlsx](#). This spreadsheet implementation of Tic-Tac-Toe is completely finished and available for you to enjoy and study. You won't need to make any changes to this file. I suggest you experiment with it and look at the formulas, so that you understand how it was constructed.

This spreadsheet consists of two parts. Rows 1-14 are where the game is actually played. Below that is where we have the nuts and bolts behind the scenes. Go ahead and play the game. The game board is highlighted in yellow in cells B3:D5. Just below, we see information about the game, such as the number of X's and O's currently displayed, whose turn it is, and the status of the game. As you place X and O characters in the board, notice how the information changes. (Excel is not case sensitive. You may enter lowercase x's and o's.) Eventually there is either a winner, or a stalemate when the board is full. All these situations are detected by the game.

Let's now examine how the spreadsheet actually works. First, we have formulas in cells C9 and C10 that count how many squares on the board are occupied by X and O. Notice that we use the countif function. For example, the formula =countif(\$b\$3:\$d\$5, b9) asks Excel to count how many squares on the board are equal to the symbol in cell b9. The second parameter is b9, which is a cell reference for the cell containing "X". We could have written this formula as =countif(\$b\$3:\$d\$5, "X"), but in general it's better to use cell references than constant values in a formula. The way we count the O's in cell C10 is analogous. The way the formula in C9 is written, using an absolute reference for the board and a regular cell reference for X, I was able to copy C9 to C10 and the new formula correctly counts the number of O's.

There are two essential things the spreadsheet must know during the game: whose turn it is, and the general status of the game. Let's first look at how we do the turns. Customarily X goes first. How do we know when it's X's turn? This is when there is an equal number of X and O characters on the board. So all we need is an IF function, and we can print an appropriate message depending on whose turn it is. I should mention that I've tried to make the design of this spreadsheet simple, and it's not really foolproof. The spreadsheet does not check to make sure the players really alternate turns, and it does not prevent players from changing the board arbitrarily, so it is possible for players to cheat.

The most interesting part of the spreadsheet is figuring out when the game is over. At any point in time, the game could be in one of four possible "states" or situations: X has just won the game, O has just won the game, the game has reached a stalemate (draw, or "Cat's game"), or we just continue playing. This is where the behind-the-scenes action comes in. Down below, in cells B20:D22, there is a duplicate version of the game board, but this time using a numerical representation. Here, the numbers 1 and -1 correspond to players X and O, respectively, and zero corresponds to blank squares. Take a moment to

verify this. Place some X's and O's in the yellow game board, and see what numbers appear in B20:B22. Do you understand the formula we use to determine which of the three numbers to display?

How do you win at Tic-Tac-Toe? You need to get 3 of your tokens aligned in any row, column or diagonal. Now we can see why a numerical representation can make this a straightforward task. If we had simply used our original board representation, the question asking if X won on the first row would look like this: =if(B3="X", if(C3="X", if(D3="X", ...))). Using a numerical representation is much easier, because this question could be condensed to: =if(sum(b20:d22)=3, ...) By using numbers, we just have to check to see if any row, column or diagonal adds up to positive or negative 3. Cells E20:E22 keep track of the row sums, B23:D23 maintain the column sums, and cells C25:D25 maintain the sums of the two diagonals. Now, we don't know in advance which of the 8 ways a player is going to win, and it doesn't really matter. We just need to detect if a player has won anywhere in the board. Therefore, we have two extra cells at the bottom that keep track of the largest and smallest values of the 8 sums being calculated. For example, if the largest sum on the board is 3, then we know that X has somehow won the game.

So far, we know that X wins the game if the value of cell C27 is 3, and O wins the game if cell C28 is -3. But how can we tell if the game ends in a draw? This is when the board is full and there is no winner. To determine if the board is full, we simply need to count all of the board squares to see how many contain any text: we can do this with the counta function. Finally, if there is no winner and the game is not over, we just want to let the players continue playing normally. All of these situations are encapsulated in the IF function in the formula in cell B14. Read this formula carefully so you understand the order in which the questions are being asked.

Implementing Tic-Tac-Toe in Excel is an interesting and perhaps challenging exercise. We made use of several functions like sum, if, countif, counta, max and min. Looking at this list of function names, do you know how each one was useful to us?

Here is a neat Excel feature: try this key sequence, CTRL-`. In other words, hit the two keys at the lower left and upper left keys on your keyboard. This keyboard command is called *toggle formulas*, and asks Excel to show all the formulas in the spreadsheet. To go back to the previous view, simply hit CTRL-` again. The toggle formulas feature is useful when a spreadsheet has many complex formulas that you want to review quickly. Close the file ttt.xlsx.

The vlookup function

The vlookup function is a shortcut of doing a lot of if functions in a row. Suppose we wanted to convert a student's percent grade into a letter grade. We could use an if-function as follows: =if(b7 >= 90, "A", if (b7 >= 80, "B", if (b7 >= 70, "C", if (b7 >= 60, "D", "F")))). What an ugly formula! We had to use 4 nested if-functions to accomplish this task, and it would be easy to mistype something in this formula. Often we want to apply some kind of grading or rating scale, and the vlookup function would be a much less tedious way to accomplish this.

Here is the approach. In one part of the spreadsheet, we will type in a grading scale, where we specify the minimum score needed to make each of the letter grades from F all the way up to A. Then, we employ the vlookup function to choose the correct letter grade.

Example #8 – Letter grades

Let's return to our earlier spreadsheet on student grades. Open the file [gradedemo.xlsx](#) that you worked on in the previous lab. Here, we had computed the average test grade earned by each student. It's now time for us to assign letter grades. Column F will contain our letter grades. Type in "Grade" in cell F1. Also, let's horizontally center column F so that its formatting matches that of the other data columns. Highlight letter column letter F and click the center alignment button.

Since we are going to use the vlookup function to compute letter grades, the first thing we need is a grading scale. We'll put this off to the side in columns H and I. In cell H1 type, "Grading scale". In cells H2:H6, enter these 5 values: 0, 60, 70, 80, 90. In cells I2:I6, enter the letter grades "F", "D", "C", "B", "A". There are a couple of features to notice about using a grading scale with vlookup in Excel. First, the numbers that we specify are the minimum criteria for the letter grades. Second, the categories have to be listed in ascending order, in this case from "F" to "A".

Now we are ready to enter our vlookup formula. Since this is the first time we've used vlookup, let's use the "Insert Function" command to guide us step by step. First, select the cell where we want to put Marianne's letter grade, which is cell F3. From the Formulas menu, click the Insert Function button. Search for the vlookup function and select it from the list of functions. Click OK, and we now have the Function Arguments dialog box. The vlookup function requires 3 parameters. First, we need to specify the value we are looking up. Second, we specify the location of our grading scale (a.k.a. vlookup table). And third, we specify which column of the grading scale contains the answer we want. Almost always this is column 2, i.e. the second column of the table. Ignore the 4th box – the 4th parameter is optional and we will not use it.

You may want to drag the dialog box so that it's not covering up the data in the spreadsheet. Click inside the first box, labeled `Lookup_value`. The value we want to look up is the student's percentage grade. Enter E3 in the box, and press the tab key to continue to the second parameter. Now we need to enter the location of the grading scale or vlookup table. This is H2:I6. One important detail is that this needs to be entered as an absolute reference. The grading criteria are essentially constants just like the exam weights. So, this parameter really needs to be `H2:I6`. Hit tab to continue to the third parameter, which is `Col_index_num`. Enter 2, and then hit the OK button.

You should see that Marianne's grade in the class is B. Next, copy the vlookup formula down to the other students in the class. Do the results make sense?

How does vlookup really work? When we give it a lookup value, such as 70.5, Excel reads the vlookup table's first column to see which category fits. Notice the values in the first column are ascending. Excel

stops when it sees the next number is too big. The number 70.5 is between 70 and 80, so vlookup returns the corresponding value in the second column. Why the second column? Because we specified “2” as the last parameter. Occasionally we have larger vlookup tables where the answer we want is not always in the second column. We’ll see this happen in the next example.

Save and close gradedemo.xlsx.

Example #9 – Tax table

Let’s look at an example of a vlookup table that has more than 2 columns. This time we’ll consider an income tax table, which traditionally looks something like this:

If the taxable income is at least	But not over	Then your tax equals	Plus this percentage	Of the income in excess of
0	45200	0	15%	0
45200	109250	6780	27.5%	45200
109250	166500	24393.75	30.5%	109250
166500	297350	41855	35.5%	166500
297350	-	88306.75	39.1%	297350

Let’s work out an example. Suppose your income is \$ 119,250. Then the third row of figures applies to you because your income is between 109250 and 166500. Your base amount of tax is \$ 24,393.75. But then you also need to add 30.5% of the amount that your income exceeds 109250. You exceed this threshold by exactly \$10,000. 30.5% of this is \$3,050. (This happens to be called the marginal tax.) Adding the two amounts of tax gives $\$24,393.75 + \$3,050 = \$27,443.75$.

We would like Excel to do this sort of calculation automatically when we enter an amount of income. Notice that the table above is not as complex as it first appears. For instance, the second and fifth columns are redundant. When we implement this table in Excel, we really only need 3 columns: the first, third and fourth ones above.

Open the file [tax table.xlsx](#). The tax table has already been entered into the spreadsheet, and we will use it as a vlookup table. Notice it has 3 columns. The spreadsheet is organized so that a user can simply enter the amount of taxable income in cell B11, and the amount of tax is figured below. It is possible to compute the total tax using a single formula, but I’ve decided to break it up into smaller steps so that it’s easier for you to see how the calculations proceed. You may find it convenient to refer back to the numerical example we worked out above so that you can determine what formulas to write in Excel.

Just for fun, enter the value 119250 in cell B11. We need to enter a formula in cell C15 that looks up what our base tax amount is. Recall that the general format of the vlookup function is

=vlookup (lookup value, location of lookup table, column number)

In this case, the value we are looking up is our income amount in B11. The location of the vlookup table is A3:C8. And the base tax amount is located in the second column (i.e. column 2) of the table. Note that it's a coincidence that column #2 in our table happens to be in column B in the worksheet. No matter where this tax table is, we should always refer to the column containing our desired result by number, not by letter. So, the formula we want is =vlookup(b11, a3:c8, 2). It's an excellent idea to make the location of the vlookup table an absolute reference, just in case you later wanted to copy the formula. As we saw earlier, the base tax should be \$ 24,393.75.

Next, in cell C16, we need to create a formula using vlookup again that gives us the tax rate from the table. What should the formula be? _____ Enter it in cell C16. The result should be 30.5%.

The value in C16 is a percentage, and it needs to be multiplied by, not our total income, but by how much of our income exceeds the bottom of the tax bracket threshold. But first, what is this threshold? If our income is \$119,250, the threshold is \$109,250, which is the number in column 1 of the vlookup table. In cell C17, write a formula using vlookup that returns the appropriate threshold based on our income. Once you have that value, then we can figure cell C18, which is the amount of the marginal tax. Write this formula in cell C18. If you are unsure of how to write out this formula, go back and review the steps of the manual calculation we did earlier.

Finally, all we need to do is add our total tax, which is the base plus marginal tax amounts. Did your total come out right? Please have me or a lab assistant check your work.

Notice that in this example, we used the vlookup function 3 times, and each time we wanted to return a value from a different column from the lookup table. Save and close tax table.xlsx.

Summary: At this point you should understand the following.

- Creating a spreadsheet from scratch
- Mixed references
- Using the fill handle to create a sequence of values (e.g. 1, 2, 3, ... or 5, 10, 15, ...)
- Selecting a non-contiguous range
- Using the paint bucket
- Zoom and freeze panes
- Conditional formatting and color spectrum technique
- Simple statistical functions, if and vlookup
- Using the mouse to help you specify a cell range
- Using the mouse to select a range of cells for a formula
- It's handy to be able to allow a formula's result to be blank, in case a certain calculation is not applicable. We accomplish this by using an IF function, and letting either the true-value or false-value to be "".
- Toggle formulas

Review Questions

1. If a cell contains the formula $=c3 + \$d4$, and we copy this formula to the cell that is one row above and one column to the left, what is the new formula? (You should be able to figure this out without using Excel!)
2. As you know, there are times when we want the result of a formula to be blank. Consider the following example. We want to compute the percentage of each country's Olympic medals that are gold. But if the country didn't earn any medals, then we should leave the percentage blank (so that we don't divide by zero). What formula needs to be entered in cell B6 so that when we copy the formula to the right, the result looks like the following spreadsheet?

	A	B	C	D	E	F	G	H	I
1	What percentage of a country's Olympic medals are gold?								
2									
3	country	Australia	Canada	France	Germany	Ireland	Japan	Spain	USA
4	# medals	2	24	9	29	0	1	0	25
5	# gold	1	7	3	11	0	1	0	9
6	% gold	50%	29%	33%	38%		100%		36%

3. Use the Excel "help" feature or Paste Function command to look up how to use the MOD function. How would we use MOD and IF to determine if a year number represents a leap year?
4. What do the following function calls mean? Which formula do you think would be the most useful in a typical spreadsheet?

`=countif(a2:b15, "=A19")`

`=countif(a2:b15, ">A19")`

`=countif(a2:b15, A19)`

5. What is another way to write this formula? `=if(B3 > 0, 1, 2)`
6. What is wrong with this formula? `=if(D2 > A2, A2 - D2, A2 - D2)`
7. Let's return to the spreadsheet [gradedemo.xlsx](#). I'd like to show you a convenient function called *sumproduct*. You may want to use the paste-function feature to learn about how this

function works. In this spreadsheet, column E contains formulas that compute each student's percent grade. The formula was a little tedious because we had to multiply each test score by the corresponding weight. The sumproduct function can do this automatically: it is perfect for multiplying corresponding values in 2 cell ranges, and adding the result. In mathematics this is also called a "dot product". It turns out that the formula for Marianne's grade would be `=sumproduct(b2:d2, b12:d12)`. But we need to remember that the score weights should be fixed for all students. Make the appropriate change, and copy this formula down for all the other students. You should notice that the numerical results are the same as with the earlier formula.

8. Open the spreadsheet mass ed.xlsx. This file shows the educational attainment of the population 25 and older among counties in Massachusetts. The five levels of attainment are: less than a high school diploma, a high school diploma, some college but less than a bachelor's degree, a bachelor's degree, and a graduate degree. We'd like to compute two sets of percentages. In cells C20:G33, we need to know what percent of each county's population attained the indicated level of education. Enter the appropriate formula in cell C20 and copy to the rest of those green cells. Format these cells as a percent with 0 decimal places.

For cells C38:G51, we'd like to know, for each level of education, what percent of the state's total lives in each county. Enter a formula in cell C38 and copy to the rest of those green cells. Format these percentages also to 0 decimal places.

Looking at the results in column G: which two counties have the highest percentages of their people attaining graduate degrees? _____ Which two counties account for the state's greatest share of people who attained a graduate degree? _____ Why do you think these are not the same two counties as in the first question? _____

Excel lab #3

In this lab, you will see how to create charts. Then, we will look at some common data analysis techniques: sorting and filtering data. We will also look at how to print pages from an Excel file.

A chart is a visual depiction of our data. We use charts when we want our audience to immediately see a trend or to quickly come to some conclusion. There are several types of charts we can create in Excel. The ones that we will practice with are the *column*, *pie*, *line* and *scatter* charts. In Excel, we create charts that are based on the numbers in the spreadsheet. One convenient feature is that, like formulas, charts automatically update themselves if you later change the underlying data.

When you are ready to create a chart for your spreadsheet, the first question to ask yourself is which data should be visualized. Often a spreadsheet contains many rows and columns, and it's likely you are interested in showing the results of just 1 or 2 columns. After selecting the data you want to put in the chart, you need to select the chart type. You can then customize your chart, with a title, captions, labels, a legend, etc. A chart can be placed anywhere in your spreadsheet, either alongside the numerical data, or by itself on another page. When you are finished with a spreadsheet, it's sometimes useful to copy your chart to another file such as a Word document or a Web page.

Example #1 – Furman cost revisited

Let's open one of our original spreadsheets, [furman_cost.xlsx](#). Let's begin our survey of charts with pie charts. A pie chart is used to show proportions or parts of a whole. In this case, a student's total cost of going to university can be broken down into three components in this spreadsheet. Let's create pie charts for the earliest and latest data here, in order to show a contrast over time.

Pie Chart

Our first pie chart will show the 3 costs for 1920-21. Highlight cells B3:D4. Notice that we are highlighting not only numbers, but their labels as well. Having the labels will make it convenient for us to tell the different pie slices apart. At the top of the Excel window, select the Insert tab. In the ribbon, you should see the "Charts" group of controls. Here we see the list of possible chart types. Click on Pie, and a small menu of icons representing variants of pie charts appears. Click on the first icon, which is just an ordinary pie chart. As soon as you click the type of chart you want, the chart should appear in the middle of the spreadsheet. Feel free to move it out of the way to a more convenient location in the page. Notice the pie has 3 slices because we selected 3 values for the chart. To the right is the legend, identifying each colored slice.

So far the chart is rather plain. Once a chart is created, we like to apply various options to it. For example, we can add a title to the chart. It is also convenient to label the individual slices of the pie. This is because it's usually unpleasant for the reader to constantly need to refer to the legend to determine what the slices represent. To make modifications to the chart, first notice that at the top of

the Excel window, there are some new tabs, under the heading “Chart Tools”: Design, Layout and Format. If you do not see these extra tabs, make sure the chart is selected. Select each of these tabs so that you can see the possible things you can modify on your chart.

Let’s add a title to the chart. At the top of the Excel window, select the Layout tab. From the Labels group of controls, click on Chart Title, and from the menu select “Above Chart”. A title now appears in the chart. Change the text “Chart title” to “Furman Cost in 1920”. Next, we are going to label the pie slices. In the ribbon, click on “Data Labels”. From the menu, select the last option “More Data Label Options...” A dialog box called Format Data Labels appears. We can specify exactly how much information we want displayed for each pie slice. In the “Label Contains” section, make sure the following boxes are checked: Category Name and Percentage. Below, in the “Label Position” section, select Best Fit. Click the close button, and notice the new pie slice labels in the chart. At this point, the legend is no longer necessary, so we can delete it. In the ribbon, select the Legend control, which is also located in the Labels group. Select “None” so the legend disappears.

Eventually, we’d like to have 2 pie charts side by side so that we can visually contrast the 1920 distribution of costs from 2008. Resize the chart so it’s a little narrower, and drag it to a position underneath the data and to the left, roughly overlapping with columns A through F.

Repeat the above procedure to create a second pie chart, this time for the 2008-09 data. The first thing to do is to select the data labels as well as the numbers. Since they are noncontiguous, we’ll need to use the control key. Highlight cells B3:D3. Hold down on the control key, and then select cells B12:D12. (Then you can release the control key and the mouse, making sure the proper ranges have been selected.) Now you are ready to create your second pie chart. When it has been created, resize and drag it so that it’s to the right of the first chart. Make sure the formatting of the two charts is similar. What is the most important conclusion one can draw from seeing these charts? _____

A word of caution about pie charts: when you select your data to put into a pie chart, make sure you do not include the grand total. Otherwise, Excel will treat the total just like any other value, and assign it a pie slice taking up half the pie, which is definitely not what you want.

Congratulations, you have just created your first charts! Next, we’ll create a simple column chart.

Column Chart

The column chart is the most common type of chart used in Excel. Each value is represented by a vertical bar. (In Excel, there is also the *bar chart*, which uses horizontal instead of vertical bars.) Let’s create a column chart showing the total cost of attending Furman for all the years in the spreadsheet. Select cells A3:A12, hold down on the control key, and select cells E3:E12. From the Insert menu, look in the Chart group of commands, and select Column. A menu of chart icons appears, and you can see quite a variety of column charts. If you hover the mouse on one of these icons for a second, a description will

appear. For example, read the descriptions for the first 3 chart types: “Clustered Column”, “Stacked Column” and “100% Stacked Column”. Each one has a purpose. For now, let’s just choose the first option, the 2-D stacked column chart. We can always change chart types later if necessary.

A new chart appears in the middle of the page. Right click on the title, and select “Edit Text”. Edit the title so that it says “Furman Total Cost”. Next, eliminate the legend. You might notice that the column for 1920-21 is so tiny that we can’t see it. We need to enlarge the chart. Better yet, we should move this chart onto its own page. To change the location of a chart, right click on the background of the chart to bring up a menu. Select “Move Chart...” and a Move Chart dialog box appears. Select the option for New Sheet and click OK. Your column chart now appears on a new worksheet or page of the spreadsheet file.

Multiple worksheets

When a spreadsheet contains more than 1 worksheet, it’s an excellent idea to give each one a meaningful name. Up until now, we’ve just had 1 page per file. Along the bottom of the Excel window there are tabs for each page. By default they are called Sheet1, Sheet2 and Sheet3. We can change the names of our pages and add or delete pages. (If you ever wanted to add a page to a spreadsheet, you could click the tab on the right end of the list of sheets, which shows a tiny yellow star.) Right click on the tab for Sheet2, and select “Delete” because we don’t need this page. You will get a warning message about the deletion, but delete the page anyway because it’s blank. Also delete Sheet3. Right click on the tab for Sheet1. Select “Rename” and change the name of this page to “data”. Also change the name of the Chart1 page to “column chart”.

Note that when your spreadsheet consists of multiple pages, there are two ways you can move from page to page. First, you can click on the appropriate tab. The other way is to use the key sequence CTRL-Page Up and CTRL-Page Down, which allow you to move to the next sheet forward or backward in the list.

Now that our column chart is on its own page, we can see all the columns. Overall, this is a good chart – the axes labels are easy to read, and the horizontal lines that Excel draws automatically are effective at allowing us to eyeball approximate values. Note that if you wanted to display exact dollar values on each column, you could do so in a way similar to how we labeled our pie chart. One problem we do have with this column chart is that the first 3 or 4 columns are so short that it’s difficult to visually estimate their heights. It turns out that when you have a chart like this where you are displaying both large and small values, or when you are trying to show rapid change, it may be a good idea to use a *logarithmic scale*. Here’s how you can do it: Right click on the vertical axis. From the menu, select “Format Axis...” When the Format Axis dialog box appears, you’ll see a section labeled Axis Options. Here, check the “Logarithmic scale” box, and click the Close button. How has the chart changed?

If you decide that you prefer the linear rather than the logarithmic scale for the vertical axis, feel free to undo the change. Save and close furman cost.xlsx.

A general note: with column charts, often the vertical axis is measuring money or some other large quantity. You should make sure that the precision of the axis values is appropriate. For example, since Furman costs are usually expressed in whole dollars, there should not be any cents “.00” in the axis labels. If this happens, right click on the vertical axis, choose “Format Axis” from the menu that pops up. When the Format Axis dialog box appears, choose Number, and make sure Decimal places is set to 0.

Example #2 – Summer in Canada

Line Chart

Briefly, I’d like to show you another type of chart that’s especially useful when you have several columns of values you’d like to depict. It’s a line chart. Open the file [canada.xlsx](#). This simple spreadsheet contains the daily high temperatures recorded at 4 cities during August 2007.

Highlight all the cells A1:E32. From the Insert menu, in the Charts command group, select Line, and choose the first type of line chart. When the chart appears, right click in the background, and from the menu select “Move chart”. Indicate that you want to move the chart to another worksheet. Usually we like our charts to be large for readability. Looking at this chart, what conclusions can you draw?

Save and close canada.xlsx.

Example #3 – calories revisited

Open the file [calorie.xlsx](#). In this example, we will return to our earlier calorie spreadsheet to experiment with different variants of the column chart. Column charts become more interesting when we have several rows/columns of data to put into a chart. In Excel, these are called different “series”. For example, we want to depict trends among our breakfasts, lunches and dinners in the same chart. So, each meal’s day becomes a data series. We will look at 3 types of column chart: *clustered column*, *stacked column*, and the *3-D* column chart.

As before, whenever we want to create a chart, we must first highlight the data. So, select cells A1:F4. Then, from the Insert menu, click on the menu for Column charts. Select the first option, called Clustered Column. This is one of the most popular types of charts, because it’s relatively easy to read. In fact, you may be satisfied with how Excel has created this chart that no additional changes are needed, other than a title. Shrink the size of the chart so that it only takes up about one-third of the window, and drag it to the bottom of the screen. Next, we’re going to create another chart so that we can look at the two charts side by side.

For our next chart, make sure A1:F4 is selected, and from the Insert menu, look for the Column chart menu again. This time select the second chart icon which is called Stacked Column. Position the new chart so that it's underneath our first chart. These two charts depict the same information, but in slightly different ways. In your own words, explain which one you would prefer in this case: _____

You can probably see how the “stacked” column chart got its name. It's the same as a clustered column, except that within the same day of the week the meals got stacked on top of each other. The advantage of this chart type is that it's much easier to see what the total calories are per day. Just by looking at the chart we can see that we ate the most on Wednesday. Monday was a close second. This would be almost impossible to determine by looking at the clustered column chart. On the other hand, clustered column is excellent for comparing individual meals, such as comparing the various dinners. Looking at the stacked column chart, it would be hard to tell which dinner had the most calories because they are not aligned like they are naturally with a clustered column chart.

One of the more interesting charts is a 3-dimensional column chart. The “3-D” aspect is simply the visual appearance of the chart; it does not mean our spreadsheet data is 3 dimensional. We can use a 3-d chart to depict our eating habits for the week. Highlight cells A1:F4. From the Insert menu, look for the Chart command group and select Column. From the menu of chart icons, look at the 3-d options. Select the 4th icon on the list, which is called “3-D Column, ... use it when categories and series are equally important.”

Now we have a 3-D chart! We have one dimension for the days of the week, another for the meal, and the vertical (z) axis shows the calories. Right away we can tell that a couple of breakfasts were skimpy, and the Tuesday dinner was especially indulgent. So that we don't overwhelm ourselves by having so many charts on one page, move this 3-D chart onto its own page, and call the page 3-D.

Do you see anything that can be improved with this chart? First of all, we won't need the legend, so you can delete it. It's possible that not all of your axis labels appeared. For example, the word “Lunch” may be missing from the list of meals on the depth axis. If necessary, you can make it appear by enlarging the chart, decreasing the font size, or by rotating the chart. To rotate a chart, right click on the background, and select “3-D Rotation”. You can experiment with the angles so that you can tweak the perspective of the chart, if you'd like.

Add an appropriate title to the chart. Move the chart so that it does not cover any of the data on the page. When you are satisfied with the appearance of the chart, have me or a lab assistant check your work to this point. Save and close calorie.xlsx.

Example #4 – house size by state

Open the file [bedroom.xlsx](#). This shows data from the US Census on how many dwellings in each state have a certain number of bedrooms. First, change the name of the page tab at the bottom from

“Sheet1” to “data”. We would like to create a chart that depicts the percentage of big houses in all the states. But first, we need to compute this percentage. For the purpose of this exercise, a “big” house is one that has 4 or more bedrooms. Click in cell I2. Write a formula that will divide the number of houses in Alabama that have 4+ bedrooms, by the total number of houses in Alabama. Format the result as a percentage to 1 decimal place. Copy this formula to the other rows in this column.

Note that row 53 of this spreadsheet refers to the total USA. Because we want to compare individual state data, we don’t want the national total included in our analysis. The best thing to do is to insert a new row between Wyoming and total so we don’t confuse ourselves later.

Highlight cells A1:A52 and also select cells I1:I52 (don’t forget to hold down on the control key as you do so). Ask Excel to create an ordinary 2-d column chart. After the chart appears, right click on the background, and ask Excel to move this chart to its own page. Change the name of the page to “4+ bedroom chart”. Remove the legend. There’s one more thing we should do for style: we need to format the vertical axis so that the percentages are whole number percents. The “.0” is unnecessary precision because all the axis labels end in “.0”. Right click on the vertical axis, and select Format Axis. When the dialog box appears, click on the Number tab, and change the number of decimal places to 0. Click the Close button.

At this point, we can appreciate that having this chart makes it easier than a column of numbers to quickly draw conclusions from the data. For example, which states have the highest (and lowest) percentage of big houses? _____

One thing that may improve the appearance of this chart would be if the states could be sorted in numerical order. Then there would be no doubt about how the states rank.

Sorting data

Let’s return to the “data” page. We’d like to sort the data (i.e. the rows) so that the percentages of big houses go from high to low. Click on any cell containing data. (Excel will assume that the data to sort will be in the contiguous cells that also contain data.) From the Data menu, look for the Sort & Filter command group. Select Sort. A “Sort” dialog box appears. In the drop-down box labeled “Column sort by”, which allows us to choose a column to sort on, choose “pc 4+ bedrooms”. In the order column, specify that you want to sort to be Largest to Smallest. Click the OK button to dismiss the dialog box and see the result of the sorting.

Once the data has been sorted, you can then go to the “4+ bedroom chart” page to see the chart *which has also been sorted* for you. Many times, a sorted series of columns is more attractive.

Printing

You will not have to print spreadsheets for this class, but it’s a good idea to know how to print in case the need ever arises. Printing a spreadsheet is straightforward in Excel. Usually a spreadsheet file

contains several pages, and we want to print one of them. For example, you may want to print just a chart, or just some summary table of data.

Make sure that the data page of bedroom.xlsx is currently open. From the top left corner of the Excel window, hover the mouse over the selection “Print”. It’s usually a good idea to look at Print Preview first before we actually commit to paper. Often the spreadsheet does not appear exactly how we want it to look, and a little formatting is needed. For example, if you have a small amount of data, you might want to center it on the page. In Print Preview, there is a Page Setup button at the top of the screen that allows you to format how the spreadsheet will appear on the page. Note that these options do *not* affect the calculations or how the spreadsheet appears inside Excel.

A common problem when printing a spreadsheet is that not all the rows or columns will fit on one page. Sometimes this problem can be fixed by selecting landscape instead of portrait mode. But usually it’s necessary to shrink the font size a little. Click the Page Setup button, and the Page Setup dialog box should appear. We need to shrink the font size: in the Page tab, in the “Scaling” section, make sure that the radio button labeled “Adjust to” is selected. Next to this button, we can adjust the percentage down to 90%. Now our data fits comfortably on one page.

Don’t confuse the page output percentage with the Excel window zoom percentage that you see in the lower right corner of the Excel window. When you change the zoom percentage, you immediately see the result on the screen, but this has nothing to do with the output on paper.

More about Page Setup

The page setup options are organized into 4 tabs: Page, Margin, Header/Footer and Sheet. Each tab has several useful options that you can experiment with. For example, on the Sheet tab, there is an option “Gridlines” that allows you to display the cell gridlines on paper. By default they are not shown. Although it’s not necessary for our current example, sometimes a spreadsheet needs more than one page because it has so many rows. In this case, it’s a good idea to repeat the column headings at the top of each page. To do this, click on the Sheet tab in Page Setup. Under the section “Print titles”, next to “Rows to repeat at the top”, we would need to select row 1.

Saving worksheet as Web page

In addition to printing pages of a spreadsheet, we can also save them in HTML format to upload them to a Website. To do this, we “Save As”, and choose .html as the requested file type. It’s also possible to save just a portion of a worksheet as its own Web page. Simply highlight the data you want to save before choosing Save As. Experiment with this feature: save each page as a separate HTML file, and open them with a browser to see how they come out.

Save and close bedroom.xlsx.

Example #5 – A chart showing a correlation

Scatter chart

Another useful chart type is the *scatter chart*. It's also called an "XY" chart. We use this when we have 2 columns of data that we think have a close correlation. Open the spreadsheet [gradedemo.xlsx](#). We'd like to depict the relationship between the scores on the first test and the final exam. The instructor would like to know to what extent doing well on the first test may foreshadow a good performance on the final.

Before drawing a chart, let me show you an Excel function you can use to measure to what degree 2 columns of data may be related. It's called *correl*, short for "correlation coefficient". This is sometimes called the "R" value. It's a statistical function that returns a value between -1 and $+1$. A number near $+1$ means the two sets of quantities are highly correlated. An answer near -1 means they are strongly negatively correlated (as one increases, the other decreases). A correlation near zero means there is basically no relationship.

In cell h12 enter "Correl:", and in cell i12 enter this formula: `=correl(b2:b10, d2:d10)`. What is the result, and what do you think it shows? _____

You have probably heard the phrase "correlation does not prove causation." In other words, just because two sets of values coincidentally increase together, we can't assume that one quantity is directly causing the other to increase. Be careful not to conclude too much about correlations.

Now let's make the chart. Highlight these cells: b1:b10 and d1:10 (holding down on the control key as you select since this is a noncontiguous range). From the Insert menu, select "Scatter" from the list of chart types, and choose the first scatter chart type called "Scatter with only Markers". When the chart appears, ask Excel to move the chart onto its own page. Change the title of the chart to "Final score by first test score". Here, the horizontal axis plots the first test score, and the vertical axis measures the score on the final. Hover the cursor at one of the points in the chart to see an example ordered pair.

When using scatter charts, it's usually a good idea to include a *trendline*. This is a "best fit" line that shows the linear relationship among the data. Right click on one of the data points. From the menu, select "Add trendline". A dialog box appears. By default, the type of trendline is Linear. But you should be aware that more complex relationships are possible. At the bottom of the dialog box, check the last 2 boxes "Display Equation on chart" and "Display R-squared value on chart". Click on the Close button. At this point you should see the trendline. The equation and R^2 value can be dragged to the upper left corner of the chart so they are not in the way. R^2 is simply the square of the correlation value "R" that you found earlier. For technical reasons, statisticians sometimes prefer to use R^2 instead of R.

What is the correlation equation? What does it mean for the slope to be less than 1? (Hint: consider the difference between plugging in 70 versus 71 in for x in the equation, and comparing the two results.)

Save and close gradedemo.xlsx.

Calculating a surplus

There is a useful quantitative concept I'd like to share with you. It has to do with finding the "highest concentration", or in other words a surplus or excess of something above an average. For instance, this sort of computation is done by social scientists do when they want to measure racial segregation in a city or income inequality among populations. Businesses can also use this quantitative technique to make decisions on where to direct their commercial plans such as opening new offices and stores or targeting their advertising.

Example #6 – city education

Open the file city education.xlsx. This is a list of all the metropolitan areas in the United States, with statistics on the educational attainment of the population. Suppose you work for a company that would like to open a new office. The management has decided they want to locate in a city that has the "greatest concentration" of college-educated adults. Unfortunately, this term is vague and has several possible meanings. It might mean the city with the largest raw number of college grads. Or it could mean the city with the highest percentage of people with college degrees. These would be easy to calculate, and we wouldn't learn much from those conclusions. New York is the largest metro area in the country, so it would have the highest number of practically anything. And a city with a very high percentage of college degrees would most likely be a small college town.

By "highest concentration", we'd like to look at the best of both worlds. Another way to put it is that we want the greatest numerical "excess" or "surplus" of people with college degrees. A place that has many more college-educated people than we would have expected given the size of the city. Such a place would have a high percentage of college grads, and likely be a large city as well.

Column E shows the number of people in each city who have a college degree, while column F has the total number of people aged 25 or over. In a nutshell, our goal is to sort the list of cities based on this highest concentration figure, and make the management's job easier by presenting a list of about 5 top cities (a "short list") to recommend as a location for the new office. At present, how are the cities sorted? _____

A sequence number

First, let's number the rows in column A. In other words, we want to provide a reference or sequence number for each metro area. The reason why is that at some point we are going to sort the cities, and we would like some way of reverting to the original order. It's true that the "undo" operation would accomplish the same thing, but often we save and close the file, or we have done so many operations, that doing undo is not practical or possible. Enter the value 1 in cell A2 and 2 in cell A3. Highlight cells A2:A3, and double-click on the fill handle so that Excel can continue the sequence of numbers. Scroll

down to the bottom of the spreadsheet to verify that all the cities have been given numbers in this ascending sequence.

Another benefit to the sequence number is that it shows you how the cities are ranked according to another criterion without having to resort the data.

The apostrophe

This spreadsheet contains data that I downloaded from the Census Bureau. The raw data begins in column H, while the interesting data for our example is in columns A-F. We basically want to ignore all the columns after F because we are not interested in so much detail. So, I decided to insert a blank column G to separate the relevant and irrelevant information. However, when we sort the rows later, we have to be careful that entire rows are selected. One convenient feature in Excel is that if you simply click on a cell that contains data, Excel will automatically assume that the data that you want to sort is all the contiguous rows and columns that are *nonempty*. We may want to take advantage of this shortcut. But I just introduced blank column G. When we sort the rows, we don't want to leave columns H-AO left out of the cold. Therefore, I've put an apostrophe in cell G1. Go ahead and take a look.

If you click in cell G1, the cell itself looks empty, but notice that the formula bar does show an apostrophe. When you start a cell with an apostrophe, this is called an "escape". In this case, we are telling Excel that there is no data in this cell, but not to treat it like an empty cell for the purpose of selecting contiguous rows/columns. This may seem complicated at first, but the idea is simple: Putting an apostrophe in G1 allows us to consider columns A-AO part of the same table of data, rather than 2 distinct tables of data separated by a blank column. This is important because if I had forgotten to put the apostrophe in G1, and I didn't manually select entire rows of data, I run the awful risk of scrambling my data when I sort. In other words, the data in columns H-AO would not correspond to the data in columns A-F, thus making the worksheet ruined and useless.

The apostrophe is also useful whenever you want to enter text that begins with a special character, such as an equals sign, a slash or another apostrophe. An example is where you want to start a cell of text with an equals sign, and you don't want confuse Excel into thinking that what you are entering is a formula.

Now, let's continue with the calculations. In cell D2, enter a formula that computes what percentage of the 25+ population having a college degree. Copy this formula down column D. Sort the cities based on the percent of college degrees from highest to lowest. What are the top 5 cities on the list? Why do you think these cities have such high percentages? _____

Scroll to the bottom of the page. In row 283, we have grand totals for all cities. Copy the formula from D281 to D283 so we obtain the overall percentage of the 25+ population who have a college degree.

Next, we compute each city's *surplus* of college educated people. Click in cell C2. Enter this formula: $= (d2 - d\$283) * f2$. The number in parentheses is the difference in percentages, and we multiply by the total 25+ population. Copy the formula in C2 down to all cities. Notice that I've already set the formatting of this column so that it will automatically display negative numbers in red and in parentheses. What are the top 5 cities in terms of their surplus? How would you describe these 5 cities, such as compared to the top 5 cities in total population? _____

Save and close city education.xlsx.

Large Spreadsheets

We start to appreciate the real power of Excel when it can handle large spreadsheets, which may contain many thousands of rows of data. As a practical note, if you do work with large amounts of data, Excel begins to slow down once the file size is well above 20 MB. Of course, this depends on your system and how much memory (RAM) you have. Based on my experience, if you have more than 50 MB of data, you are better off splitting your data into more than 1 spreadsheet file, or finding ways to reduce the file size. For example, you can delete calculations (formulas) you no longer need, or convert them into raw data values that take up less space than formulas.

Example #7 – list of K-12 schools

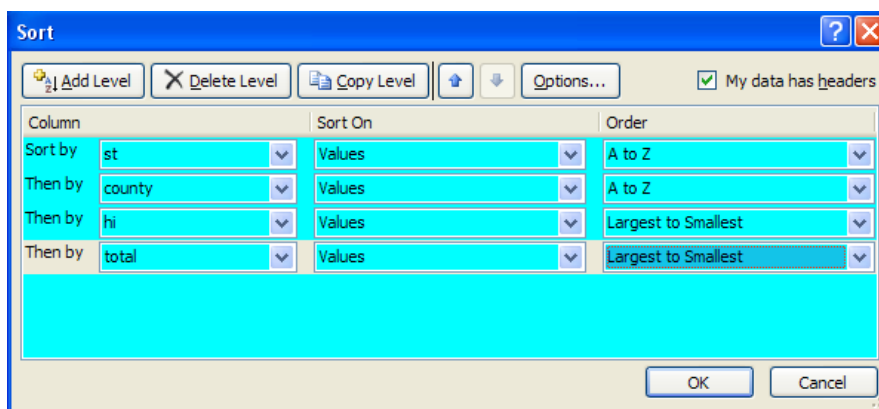
When you have a very large spreadsheet, it's a good idea for the data to be sorted so that it's easy for you to find your data. An example of a large spreadsheet is [schools.xlsx](#). This is based on data downloaded from the US Department of Education, containing basic information on nearly 100,000 public schools in the United States in 2007. For sake of brevity, additional data such as address, telephone number, enrollment breakdown by sex and grade level, etc. has been omitted.

How should we sort the schools? We can do a purely geographical sort according to ZIP code. But, we don't know the ZIP codes for many cities, so may would be hard to find schools (such as having to constantly use CTRL-F). It is probably more interesting to sort the schools according to geographical as well as other factors that people can easily relate to.

With large spreadsheets, often we want to sort by several columns. For example, if we only sorted schools by state, it may be difficult to locate a school in California. So, we need to sort by other fields as well. Let's sort the schools according to these 4 fields in turn:

- State (in alphabetical order)
- County (in alphabetical order)
- Level (the highest grade level the school offers, from high to low)
- Enrollment size (from high to low)

Select any cell containing data. From the Data menu, choose Sort, and the Sort dialog box appears. Make sure that the box beside “My data has headers” is checked. Next to “Sort by” use the drop-down menu to choose “st”. In the “Order” column, make sure it says “A to Z” since we want to sort in alphabetical order. Now we are ready to add more levels to our sort. Click the Add Level button. You should see a new row of text boxes corresponding to a second level. Click the Add Level button two more times so that we have a total of 4 levels specified. Select the appropriate column names and sort orders so that your dialog box matches this picture:



One rule of thumb with sorting is that we usually want to sort text “A to Z”, and that numbers are usually sorted “Largest to Smallest”. Can you think of an example of a numerical field that would make sense sorted from low to high instead? _____

Click the OK button to perform the sort. Take a moment to verify that the sort appears to make sense. One common mistake with sorting is that we accidentally specify one of our fields to be sorted in reverse order. In our case, we should see that within the same county and state, the schools are listed with the high schools first, and within each type of school, the largest schools are listed first. Before we continue, I want to mention something about sorting states alphabetically. In this spreadsheet, we are using 2-letter state abbreviations, so AK (Alaska) comes before AL (Alabama). Had the state names been spelled out, the sort outcome would have been different.

Another possible mistake with sorting is to specify the fields out of order. For example, what would happen to our result if we were to reverse the “st” and “county” in the Sort dialog box?

As with other things that can go wrong with Excel, if you realize you made a mistake, you can simply undo the sort and try again.

Filtering

Sometimes when working with a large spreadsheet, we don’t want to see all the data, but rather a sample of it. For example, with our spreadsheet of schools, we may be interested in seeing just the

schools in one area or just the high schools. To focus our attention on just certain rows we use Excel's *filter* feature.

To begin filtering, go to the Data menu and select "Filter" from the Sort and Filter command group. At this point, you should notice small triangle buttons at the top of each column of the spreadsheet. Each of these buttons is a drop-down menu, allowing us to make a selection. Let's obtain a listing of the public schools just in Alabama. Click the triangle button at the top of column C. A menu appears, allowing us to specify our filter. You should see a list of all the possible states, preceded by check boxes. A check means that this state should be displayed. In contrast, an unchecked box means that rows containing this state should be hidden. Notice that the first check box is "(Select All)". To filter on Alabama, uncheck the Select All box. This will remove the checks from every box. Now, check the box next to AL, and click OK.

Your spreadsheet now shows only the schools in Alabama. In the status bar at the bottom of the window, you'll see a message: "1587 of 99156 records found" indicating how many schools made it through your filter. Notice that along the left side of the spreadsheet, the row numbers are blue. This tells you that a filter is in effect.

This list of schools is still pretty long, so we can be more specific. Let's say we only want schools near Mobile, Alabama. In other words, we'll now filter on county and select just Baldwin and Mobile counties. We are still in filter mode, so click the triangle button atop column D. Uncheck the Select All box and select the boxes for Baldwin and Mobile. Then click OK. Now, the status bar informs us that we have 165 results. As you can see, we can filter on multiple criteria, in this case state as well as county. Finally, let's add a third criterion to our filter: high schools only. Select the filter button for column O, uncheck the Select All box, and check the box for 12. (Because a high school goes up to grade 12) How many results do you have? _____

Filtering is usually a temporary operation. Don't do it now, but once you are ready to undo a filter and revert to the entire list, we simply go back to the main Data menu and click on Filter once again. The columns' triangle buttons will disappear and all data will be displayed again. However, it's often desirable to keep the results of a filter. We do this by copying them to another page of your spreadsheet before undoing the filter. Highlight all the cells containing data. Hit CTRL-C to copy. At the bottom of the Excel window, click on the tab for a new page, and when the new blank page appears, click on cell A1. Hit CTRL-V to paste. Rename this new sheet "Mobile hs". Return to the main 06-07 data page. Click on the Data → Filter button to remove the filter. Verify that the 2nd page still has your original filtered data intact.

Filtering is very useful when you want to do analysis on some rather than all your data. For example, after copying the results of a filter, you can sort, draw a chart based on the data, copy to a Word document, etc. Save and close schools.xlsx.

Summary

- With pie chart, don't include a slice for total, or else it's 50% by itself, and not helpful.
- Use appropriate precision for your axis label values. For example, if your axis values contain large numbers, it's a good idea if they don't contain unnecessary precision, such as "\$ 5,000.00".
- When you have a chart with a large number of values, it's often convenient if the values are sorted. This usually makes the chart easier to understand.
- When sorting data,
 - Usually, numbers are sorted Highest to Lowest (descending), while text is usually sorted A to Z (ascending).
 - If you see columns listed by letter instead of your column headings, usually you'd want to check the box labeled "My data has headers".
 - If you have a row for totals, make sure this does not get sorted along with the rest of your data. A good place for the total is 2 rows below the last row of data, so it's not contiguous to the rows being sorted.
 - It's sometimes convenient to have a sequence number column, consisting of the numbers 1, 2, 3, ..., so that you can easily revert a sorted list to its original order. This is especially true if the original order did not have an obvious sorting scheme such as alphabetical order.
- If you begin a cell with an apostrophe, this special symbol tells Excel that this cell contains ordinary text and should not be evaluated or treated as blank. In fact, the counta function will consider a blank-looking cell containing just a single apostrophe as a nonblank cell.
- Before you need to print a page of your spreadsheet, use Print Preview to make sure data appears as you want. Sometimes adjustments are needed, so we use Page Setup, for example to adjust the font size. But as a general rule, don't shrink the font size too much.

Review Questions

1. Name one advantage and one disadvantage to using a stacked column chart as opposed to a clustered column chart.
2. What is a pie chart particularly useful for? Why would we use a line chart over a column chart?
3. In gradedemo.xlsx, create a bar chart that shows each student's average course grade. Display this chart underneath the data.
4. What do you need to type into a cell so that the value appears as 'A' with the single quotes displayed?
5. Use Excel's filter feature with the spreadsheet schools.xlsx to find out how many high schools have an enrollment of 1000 or more. To handle a range of values in a filter, choose "Number

filters” and then choose the appropriate relationship such as “Greater Than Or Equal To”. What is the total enrollment in these schools? Roughly what percentage of the total nationwide high school enrollment does this represent?

Excel lab #4

Today, we will look at some powerful ways to summarize data with Excel: subtotals and pivot tables. We will also experiment with three-dimensional spreadsheets.

Subtotals

A spreadsheet usually has a lot of data, and we are not always interested in looking at all of it. We want to be able to summarize the data. Subtotals are a way to do this. Computing subtotals will allow you to group your data into meaningful categories, and compare those categories. For example, from a long list of sales, you can determine offices have made the most money for your company.

To compute subtotals, you essentially follow the following steps:

1. Sort your data according to the appropriate categories you wish, such as by state, athletic conference, breed of animal, etc.
2. Select what data is to be subtotaed, which is often the entire page of data
3. Call up the Subtotals command, and select a summary function. Usually the summary function is “sum” because we want to add up numerical values. But sometimes we want to find counts or averages instead of sums.

Let’s begin with a typical example.

Example #1 – Employee salary

Suppose you run a small investment company that has offices in various cities. Working with a spreadsheet showing salary data, it would be interesting if we could summarize the data by job title or by position or both to see who tends to make the most money.

Calculating time

Open the file employee.xlsx. Before jumping into our subtotals, let's complete the basic data in the spreadsheet. We have an empty column for the years of experience. We can determine this by subtracting the hire date from today's date. In Excel, if you want to ask for today's date, you can use the function =TODAY(). Subtracting dates gives you a difference in days. In cell F2, enter the formula =today()-e2. At first the result looks strange. It needs to be formatted as a *number*, so ask Excel to format it to be a number with 1 decimal place. This is the number of days since the person was hired. To convert to years, we need to modify the formula we entered: divide the result by 365. Then, copy the formula down to the other cells in column F.

Do you know why the spreadsheet stores the hire date as a constant and then uses a formula to compute the years of experience? What would have happened if instead we had just the years of experience as a constant? _____

Our company has offices in 6 cities. We'd now like to compute average salaries for each city. We can do this by using Excel's subtotal feature. Even though it's called "subtotals", it's also good for finding other summary data besides just a sum: we can find the average, count, max and min as well. First, click on any cell that contains data. From the Data menu at the top of the window, look in the Outline group of commands and select Subtotal.

A Subtotal dialog box appears. Where it says "At each change in", make sure that the field "location" is selected. If not, select it from the first drop-down menu. Next, where it says "Use function", we need to change this from Sum (the default) to Average. Finally, we need to specify which column(s) we want to compute. In this case, we only want to find the average salary, so make sure that just the "salary" field is checked. The other questions at the bottom of the dialog box can be left alone; we'll just accept the defaults. Click the OK button, and your subtotals, or rather "sub-averages", appear!

Notice the left side of the spreadsheet looks a little different. At the top-left corner of the spreadsheet we have some new buttons that pertain to how we want to display our subtotals. These buttons, labeled 1, 2 and 3, are called the *level* buttons. They specify how much detail you want to show. We'll check out each one. Click on level button 1. This tells Excel to only show the grand total (or in this case, the grand average). Not very useful. Next, click on level button 2. Here, we see just our subtotals plus the grand total. Finally, click on level button 3. Now we see all the data again, along with their subtotals and grand total.

Copying subtotals

When computing subtotals, it is very convenient to copy them onto another page for further analysis. For instance, we'd like to be able to sort subtotals. In our example, we're interested in knowing which cities have the highest pay. There is a specific technique for copying subtotals that we need to practice, because it may come in handy for you someday, and not many people know this trick.

Click on level button 2 so that we display just our subtotals and grand total. Highlight the cells A1:G26, and notice that many rows are skipped because they contained raw data. From the Home menu, on the right side of the ribbon in the Editing group of commands, click on “Find & Select”. From the menu that pops up, select Go To Special. A dialog box appears: select the button labeled “Visible Cells only”, which is near the end. Click the OK button. What you just did was tell Excel that you want to select only the rows that contain subtotals, not all the rows. Now hit CTRL-C to “copy” the subtotals. Select Sheet2 from the set of page tabs at the bottom of the window. When you arrive on the Sheet2 page, click in cell A1 and hit CTRL-V to “paste” the subtotals. Voila! We have copied the subtotals and only the subtotals to a new page. If you had not done the steps with “visible cells only”, you would have merely copied the entire page.

Let’s rename our spreadsheet pages: change “Sheet1” to “data” and “Sheet2” to “subtotals”. On the subtotals page, delete columns A and B since we don’t need them. It’s up to you if you want to also delete the columns in between the city names and the salaries. If necessary, widen the column containing the city names so they are easily visible. The last thing we need to do with our subtotals is to sort them. Right click on any cell in row 8 containing the grand total. From the menu, select Insert → Entire Row. We need to have a blank line above our grand total because we don’t want it sorted with the subtotals! Otherwise, we might conclude that “Grand Average” is one of the cities by mistake. Now, click on any cell in rows 2-7 that contain data. From the Data menu, select Sort, and ask Excel to sort the data by salary from largest to smallest. Which city pays the most on average? _____

Return to the data page. Now that we don’t need to have these subtotals anymore, we can get rid of them. Click on any cell containing data. From the Data menu, select Subtotal. When the Subtotal dialog box appears, click the Remove All button. Now the data page is back to normal like it did before we created the subtotals. It’s good to remove subtotals in case we want to continue entering further data or resorting or doing other modifications: we would not want the subtotals to get in the way.

Subtotal is a nice feature in Excel, but I want to show you something even better: the pivot table! This is a fancier subtotal that allows you to see more aggregate information at once.

Pivot tables

The pivot table is one of the most important features of Excel. Essentially, it’s a two-dimensional (or multi-dimensional) subtotal. For example, with a pivot table, you can find: the average salary of employees by job title and location; the total population of the USA by state and age, how many colleges there are in each state and type. You probably noticed that all of these problems had two dimensions to them.

Pivot tables allow you to discover subtleties. For example, we can ask probing questions such as: what is the relationship between college tuition and size of school? Is there a relationship between foods containing sugar and caffeine? Pivot tables are based on an underlying page of spreadsheet data. One detail to note is that before you can create a pivot table, Excel requires you to have a column heading

for each column of data. Usually this means that you should not have blank cells in row 1 of your spreadsheet above your data. Excel will use these column headings as “field names” as you build your pivot table.

We’ll create a simple pivot table with our employee spreadsheet. We’d like to know what the average salary is by city and by job title. Make sure you are looking at the “data” page of employee.xlsx. Click on any cell that contains data. From the Insert menu, look in the Tables group of controls on the left side of the ribbon, and click PivotTable. A dialog box “Create PivotTable” should appear. Here, you can verify that all the data on the spreadsheet is selected, A1:G19. Click OK to begin creating the pivot table.

You are now taken to a new page of the spreadsheet. On the right side, you should see a heading “PivotTable Field List”, which shows a list of all the columns from the original data page. We need to check the columns that interest us for our pivot table: location, job title and salary. Underneath the list of fields is a section labeled “Drag fields between areas below”. This section allows us to specify the precise format of the pivot table. We want the Row Labels to be location, the Column Labels to be job title, and the Values to be salary. By default, Excel will sum the salaries, but we want averages. Click on “Sum of salary” and when the menu appears, choose the last option “Value Field Settings”. A dialog box appears, and you can change the “Summarize value field by” from Sum to Average. Click OK. Now your pivot table displays the appropriate information, but let’s make it more readable: select all the monetary figures, and format them using the Accounting style with 0 decimal places. Also, change the name of the page to “pivot table”. Your pivot table should look like this:

The screenshot shows the following PivotTable data:

Average of salary	Column Labels					
Row Labels	accountant	broker	economist	manager	Grand Total	
Brussels	\$ 42,500	\$ 60,850			\$ 54,733	
Chicago	\$ 84,125		\$ 51,000	\$ 68,650	\$ 67,925	
Cincinnati	\$ 48,000	\$ 51,900			\$ 50,600	
Dublin	\$ 46,850	\$ 63,500			\$ 57,950	
Prague	\$ 48,000	\$ 70,000	\$ 41,300		\$ 53,100	
San Diego	\$ 60,250	\$ 106,000			\$ 75,500	
Grand Total	\$ 55,369	\$ 57,889	\$ 75,667	\$ 54,975	\$ 59,968	

Take some time to examine the pivot table. For example, why do you think some of the entries such as C6 are blank? _____ This pivot table gives us so much information: we can see subtotals by city, by job title and by a combination of both. You can address more interesting questions such as finding which city pays the highest average salary depending on job title.

It's not always the same place! Can you give an example? _____

Save and close employee.xlsx.

Example #2 – Car mileage

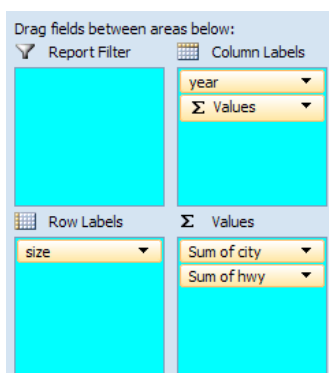
How much more fuel efficient are today's cars compared to those of the 1970s? This question is the subject of the spreadsheet file mileage data.xlsx. Open this file, and let's create a couple of pivot tables that compare typical automobile mileage for model years 1978 and 2006. Initially, this spreadsheet just contains one page of data. Take a moment to examine what raw data has been collected. The columns that we will focus on are A (year), B (car size), I (number of cylinders), K (city mileage) and L (highway mileage).

As you know, a pivot table is a two-dimensional subtotal. We would like to find the average city and highway mileage for cars based on two columns. Naturally one of these columns should be the year. The other column can be the car size (large, medium or compact), or the size of the engine in cylinders (4, 6 or 8). Thus, we'll create two different pivot tables comparing 1978 and 2006 data: one for car size and one for engine size.

Click on any cell that contains data. From the Insert menu, select PivotTable. A dialog box "Create PivotTable" should appear. Here, you can verify that A1:M96 has been selected. Click OK to begin creating the pivot table.

A new spreadsheet page appears that will house the pivot table. On the right side, you should see a heading "PivotTable Field List", which shows a list of all the columns from the original data page. We need to check these fields: year, size, city, hwy.

Under the list of fields is the heading "Drag fields between areas below". Now, we specify the format of the pivot table. This means we decide what data is used for the rows, for the columns and for the interior of the table. Usually, we want to have one field for each of these areas. But in this case we actually selected 4 fields earlier. Make sure that "size" is listed under Row Labels, and "year" is listed under Column Labels. If not, find the appropriate control (which may be listed instead under Values), and drag them to the appropriate position. The Values area should list two fields: "Sum of city" and "Sum of hwy". Check to see that your pivot table design matches the figure below.



There's one more change we need to make. We don't want the sum of the gas mileages – we want the average. Click on the control "Sum of city". When the menu appears, select Value Field Settings. When the dialog box appears, change the summary function from Sum to Average, and click OK. Repeat this with the "Sum of hwy" field as well, changing it to an average also.

Next, we should format the values in the pivot table so that they are expressed to 1 decimal place. Also, let's rename this page from "Sheet1" to "size pivot table".

If the Pivot Table Field Values section of the spreadsheet is in your way, feel free to click the X to dismiss it. If you want to bring it back, go to the (Pivot Table Tools) Options menu, and from the Show/Hide group of controls, select Field List.

Alphabetizing pivot table rows and using Find/Replace

Now, we have a little problem. If you look at the rows of the pivot table, you see that they are labeled compact, large and mid. Is this really the order we want? Excel arranges the rows of the pivot table in alphabetical order, which is not the size ordering we want. The way to fix this is to change the names of these row values so that alphabetical order will coincide with the order we intend. My suggestion is that we change large, mid and compact to: "A-large", "B-mid" and "C-compact", thereby giving an A-B-C alphabetical grading scale to our car sizes. Go back to the mileage page. Hit control-F to bring up the "Find and Replace" dialog box. Click on the Replace tab. In the text box labeled "Find what", type "large", and in the box labeled "Replace with" enter "A-large". Click on the Options button, and check the box labeled "Match entire cell contents". Click on the Replace All button. Similarly, replace "mid" with "B-mid", and then "compact" with "C-compact".

Once we have corrected the names of our sizes so they are alphabetical, we need to update our pivot table. This part is easy: go back to the Size pivot table page. From the (Pivot Table Tools) Options menu: in the Data group of commands, click on Refresh. Voila! Your rows are now in the right order!

Let's go back to the mileage page, and get ready to make a second pivot table. This time we want to do the same thing, except that we want our cars arranged by how many engine cylinders rather than size of

vehicle. The steps are practically identical to the first pivot table, so I will streamline the instructions for you.

For our 2nd pivot table: click on any datum in the mileage page. Select Insert → Pivot Table. When the “Create Pivot Table” dialog box appears, verify that it has selected A1:M96, and click OK. A new page appears. Change the name “Sheet2” to “engine pivot table”. In the Pivot Table Field List, check the following fields: year, cyl, city, hwy. (The only difference here is that our 2nd selection is cyl rather than size.) Drag the field controls so that cyl is a row label, year is the first column label. Change the “Sum of city” and “Sum of hwy” so they are averages: click on the control, select Value Field Settings, and change the summary function from Sum to Average. Format the numbers in the pivot table to 1 decimal place. Please have me or the lab assistant check your work and assist you if you have any trouble making the pivot table.

Let’s do one more useful thing with our 2nd pivot table. You may notice that our first table arranged our cars from large to small, but our engine sizes go from small to large. To make our two tables, consistent, we should arrange the engine sizes so that the cylinders are listed in descending order, 8, 6, 4. In cell A5, click on the downward pointing arrow to bring up a menu. Select the option to sort from largest to smallest. (Alternatively, you could have decided to change the 1st pivot table’s sort order instead if you wanted both lists to go from small to large sizes.)

With pivot tables, we can answer interesting questions. For example: Which is a more important determinant of car mileage, its size class or its engine? How do large cars in 2006 compare to compact cars in 1978? _____

Copying a pivot table

Sometimes when a spreadsheet has multiple pivot tables, it may be helpful to be able to see both of them on the same page, especially if they are small and use the same row and column categories. Unfortunately, Excel does not allow you to see more than 1 pivot table on the same page. So, to get around this problem, we can simply copy the resulting pivot tables to a new page. To copy a pivot table to a new sheet, highlight all the values inside the pivot table, and hit CTRL-C to copy. Then, go to the new page, right click and choose *Paste Special*, and select “Values and number formats”.

Save and close mileage data.xlsx.

Paste Special

Most of the time when you want to copy some information, you do “copy-paste”, such as with the CTRL-C and CTRL-V key commands. But occasionally when we paste, we should use “paste special”. You can get to paste special by right clicking on the cell of your destination. We will not practice this extensively, but here are some examples of when it is useful to use Paste Special:

- When you want to transpose data. For example, turning the data sideways so that the rows and columns are swapped.
- Copying the results of formulas or a pivot table onto a new page, so that you can save space. (Doing a simple “copy-paste” without paste special would cause formulas to be invalid.)

Example #3 – Colleges

Open the file [college2006.xlsx](#). This spreadsheet shows selected data on many American colleges. This example is a little more sophisticated. Let’s investigate the relationship between how much tuition a school charges and its enrollment. Do larger schools tend to charge higher tuition? To answer this question, we’ll do the following steps:

- Filter the list so we only analyze the private colleges.
- Establish grading scales for enrollment and tuition.
- Assign “grades” or categories to each college using the vlookup function.
- Create the pivot table.

First, we need to make sure that we make a fair comparison between colleges. Tuition charges are very different at public versus private schools, so let’s focus just on private colleges. This means the first thing to do is to filter our list.

From the Data menu, select Filter from the Sort & Filter group of commands. Triangle buttons appear at the top of each column of the spreadsheet. Click the button in column H. Uncheck the “Select All” box, and check the two boxes for private-not-for-profit colleges. In the status bar at the bottom of the window, Excel should tell you that there are 1959 such colleges found. Copy these rows onto a new page of the spreadsheet, and call this new page “private colleges”.

The columns we’d like to compare are column C (for enrollment) and column Y (for tuition). Eventually, we’d like to create a pivot table that shows how many students attend various colleges by size of school and level of tuition charge. Note that we should not create the pivot table right away. If we did, the pivot table would be enormous, and it would not be useful to us. Let me show you a brief illustration so you can see what I mean. From the Insert menu, choose PivotTable. If you let “enroll” be the row labels and the “out-state” be the column labels, the pivot table winds up having over 1000 rows and over 1000 columns! And most of the cells in the table are blank, of course, since we don’t have anywhere near 1 million colleges in the country. What’s wrong with this pivot table? The reason why we have so many rows and columns is that Excel has created a row and a column for each unique value of the enrollment field and each unique value of tuition. We are not really interested in exact values of tuition and enrollment. Instead what we need is some kind of “grading scale” that can group colleges into meaningful categories. This is a very important step if our pivot table is to be successful.

Please delete the page containing the awful pivot table. Pivot tables are best when they are concise. We need to create 2 extra columns of data: we should give each college a “grade” or category for what

level of tuition they charge and how big of a school it is. In cell AD1 type “enroll-class” and in cell AE1 type “tuition-class”. These columns will contain our categories for the colleges, which will be determined by using the vlookup function.

Before we can use vlookup, we need to type in our grading scales. Type in “Enrollment grading scale” in cell AG1 and “Tuition grading scale in cell AJ1. Let’s create a grading system of 4 grades for each of these fields. To enhance readability, let’s highlight the areas where our grading scales will appear. Select cells AG2:AH5 and cells AJ2:AK5 (hold down on control to select the noncontiguous range). From the paint bucket, select a soft color such as light green along the bottom row of the swatch.

Column AG will contain the threshold values and column AH will contain the names of the grades. Let’s start with the grade names. We want 4 grades of enrollment, so enter these values: “D-small” in cell AH2, “C-medium” in cell AH3, “B-large” in AH4, and “A-huge” in AH5. Enter the value 0 in cell AG2. Now, the next thing we need to figure out is the threshold values for how we want to define a medium, large and huge college. What should they be? Using our intuition, we could start with some simple numbers like 5000, 10000 and 20000. But I want to show you a clever way to determine these thresholds. I’d like the 4 grades to be approximately equally distributed among the college student population. It’s not interesting if we have one grade category that is very large or very small, as it would skew the results.

We’ll let Excel help us determine our enrollment level thresholds. Click on any cell in rows A-AE that contains some data. From the Data menu, select Sort. Ask Excel to sort the colleges according to enrollment from highest to lowest (i.e. in descending order). If you highlight column C, you see that the total enrollment in all private colleges is 3,688,674. What is one-quarter of this figure? It’s a little over 900,000. We’d like to know how far down the list of colleges we need to go until we have a *cumulative total* of about 900,000. Click in cell C2, and slowly drag the mouse down to highlight more cells in this column. As you do so, pay attention to the status bar, where it gives the quick sum. Stop when the quick sum reaches 900,000. What is the enrollment of the last college you just highlighted? It doesn’t have to be exact, so we can just round this to the nearest thousand to keep things simple. This is the lower threshold for our “huge” colleges, so enter this number in cell AG5.

We will repeat the quicksum technique, this time to find the cutoff college size for the “large” class of schools. Click in cell C2. Drag the mouse down the column until the quicksum slightly exceeds 1,800,000, which is half of the total private college enrollment. Take note of the enrollment figure, and round it to the nearest thousand or five hundred, and enter it in cell AG4. And finally we’ll determine the cutoff for the “medium” schools. It may be more efficient to work from the bottom this time. Click in the last cell in column C. Drag the mouse up until you reach about 900,000 in the quicksum. Take note of the approximate enrollment level, and enter it in cell AG3. Now that we have specified the thresholds for the various categories of college size, our grading scale is complete.

Next, we need to do the same thing with tuition values. We’d like to have 4 grades just like we did with school size. (There is nothing special about the number 4, but it’s convenient for pivot tables to be concise.) First, we need to sort the colleges by tuition in descending order, i.e. from highest to lowest

tuition rate. The tuition grading scale will go in the area you previously highlighted AJ2:AK5. In cells AK2 through AK5 enter the grades, “A-expensive”, “B-average”, “C-low”, “D-cheap”. Notice that we’re making the grades of our two vlookup tables distinct. It would be confusing to simply see the word “high” and not know if this refers to a high enrollment or high tuition. Widen column AJ as necessary so it fits the text comfortably (but not so wide that AJ1 can accommodate “Tuition grading scale”).

From the Data menu, choose Sort. Ask Excel to sort the colleges by tuition (the field called “out-state” in column Y) from highest to lowest. We need to find our thresholds for what defines an expensive, average and low cost school. Enter 0 in cell AJ2. As we did before, click in cell C2, and slowly drag the mouse down until the quick sum reaches about 900,000. Take note of the tuition value of the last school highlighted. Round to the nearest thousand so the thresholds don’t seem too arbitrary, and enter this value in cell AJ5. Click in cell C2 again, and drag the mouse down until you see about 1.8 million students. The approximate tuition value of the last school highlighted should be entered in cell AJ4. Finally, click the last cell of data in column C. Drag the mouse up until the number of students reaches 900,000. This level of tuition will be the threshold to enter in cell AJ3.

Now our grading scales are finished. The next step is to complete columns AD and AE, which assign the grades to each college. We need to use the vlookup function to accomplish this. Click in cell AD2. We need to look up the enrollment of this college given in C2, and compare it against the grading scale for enrollments given in AG2:AH5. Thus, the formula should be `=vlookup(C2, AG2:AH5, 2)`. Why are the dollar signs needed in this formula? _____

Feel free to use the insert-function feature if you are more comfortable entering the formula that way.

Similarly, we’ll assign marks for tuition in column AE. Click in cell AE2. We need to compare the tuition value given in Y2 against the grading scale given in AJ2:AK5. Enter this formula: `=vlookup(Y2, AJ2:AK5, 2)`. Highlight cells AD2 and AE2, and copy these formulas down the two columns for the other colleges, down to row 1960, so that every college gets graded.

We are ready to create the pivot table. Recall that we want to know how many students attend certain types of schools, defined by the enrollment and tuition grades. From the Insert menu, select Pivot Table. Excel should verify that the set of data the pivot table is based on is A1:AE1960. When the new pivot table page appears, look along the right side of the page: under Pivot Table Field List, check the boxes for “enroll”, “enroll-class” and “tuition-class”. Underneath, in the area labeled “Drag fields between areas below”, drag enroll to the Values area and drag tuition-class to the Column labels section. Now your pivot table has taken shape. But we need to change “Count of enroll” to “Sum of enroll”. In the Values section, click on Count of enroll, and choose “Value field settings” from the menu. Change the function from Count to Sum.

The pivot table is essentially finished. At this point, we can format for readability, and draw our conclusions. Let’s format the numbers so that they show commas. Highlight cells B5:F9. Right click and choose Format Cells. In the Number tab, from the list of format categories, choose Number. Set the number of decimal places to 0 and check the box “Use 1000 Separator”. Click OK. You may need to widen some columns so they comfortably fit the values.

Take a look at the values in the “A-Expensive” column. What can we say about students attending expensive colleges? _____

Similarly, we could look across the row labeled “A-huge”, and we may draw a similar conclusion. On the flip side, what do you notice about the trend of values in the “D-cheap” column, and how does it compare with the A-Expensive column? _____

As was mentioned earlier, a correlation does not prove that one thing causes another.

Example #4 – College grade distribution

Conditional Sum

Open the file [ag_college.xlsx](#), which contains statistics on grades from a hypothetical agricultural college. The dean has some questions about the grade data. Are 1st year courses tougher than the others? Do students tend to do better if their classes are held at certain times? To help address these questions, we’d like to subtotal the number of A’s, B’s, C’s, D’s and F’s based on the level of each course, and also separately by time of day. Let’s first address the issue of finding the grade distribution by level of course (1-4).

There are at least four ways to approach this general problem of finding subtotals. First, we could use the ordinary subtotals feature. However, this would require us to sort all the data by course level, and doing so would ruin the formatting of this attractive spreadsheet. We would need to copy the data to another sheet, and this seems like too much trouble for a simple computation. Second, we could try a pivot table. However, when creating the pivot table, we would have to select 6 columns: the “level” column as well as columns for the 5 different letter grades. Selecting more than 4 columns makes a pivot table very complex. We’ll opt for a third approach: we’ll use the **sumif** function. A fourth option, database functions, will be discussed later.

As you saw in a previous lab, the sumif function is similar to the sum function. But rather than summing all the cells in the range, Excel checks to see if a condition is satisfied. For example, =sum(d7:d163) would find the total of all A’s awarded in all classes during the year. We need to use sumif to specify constraints. The general form of the sumif function takes *three parameters* as follows:

=sumif(cell range to check, value to look for, cell range to sum)

Note that the first and third parameters are cell ranges. They need not be the same range of cells: they are usually two different columns of data. If it so happens that the range of cells that you want to query is the same range you are going to be adding up, then the third parameter can be omitted. In our case, if we want to add up all the “A” grades by course level, we need to sum values in column D, but we need to check column B to see what level each course is. Thus, the formula to add up the A’s can be written: =sumif(b7:b163, 1, d7:d163). In English this is saying: look at the course level numbers to see which ones are “1”. For each “1” level course, add up the corresponding number in column D. It turns out that

it is very convenient if we can express the *second parameter as a cell reference* rather than a constant. Otherwise we would have to write the formula 4 different times (once for each value 1,2,3,4)!

In cell A165 enter “Grades by course level”. In cells B166:B169 enter the numbers 1, 2, 3, and 4, respectively. These 4 values will be used as the second parameter in sumif as we mentioned at the end of the previous paragraph. In cell D166, we’ll enter a formula that will be used to find the number of A’s in level 1 classes. It should be: =sumif(b7:b163, b166, d7:b163). The result should be 403. Before we can copy this formula to the other cells in D166:H169, we need to insert dollar signs (\$) in the formula as appropriate. Some of the row/column references should not change. Which ones are they? _____

When you have fixed the formula in B166, copy it to the other cells in D166:H169. For example, you should find that there are 19 total F’s at level 4.

To answer the dean’s question about comparing various class levels, we need to compute an average grade. We’ll put this in column I beside the grades. In cell I166, enter a formula that finds the average grade among the distribution on row 166. To figure the quality points, count each A as 4 points, each B as 3, each C as 2 and each D as 1 point. Divide the quality points by the total number of grades. Then, copy your formula from I166 to I167:I169 as well. Format the results to 2 decimal places. What significant difference (if any) do you find among the data?

Next, let’s also compute grades based on class time. The times are given in column C. This step is analogous to what you did before when you were finding subtotals by level. In cell A171 enter: “Grades by class time”. In cell B172 enter “8 a” and then immediately hit enter. Excel understands that you are typing in a time and displays this as “8:00 AM”. This is a nice shortcut for entering times. Continue down the column entering the hours all the way up to C178 which should be 2:00 PM. Now we can enter formulas as before. In cell D172, we need to enter a sumif formula that asks how many A’s were awarded in 8:00 classes. This is almost exactly the same formula you entered earlier in D166. After you type in your formula, copy it to the other cells in D172:H178, and format to 2 decimal places. In column I, copy the average grade formula to cells 172 through 178. Do you find any significant patterns in grades by class time?

One thing to note is that if we wanted to perform subtotal calculations on the various terms, we would need to add a new column for term. A good place for this column is to the left of the current column A.

Save and close ag college.xlsx.

Three-dimensional spreadsheet

Up until now, all of our cell references have been two dimensional: we have specified a column and a row for each cell. But cells also reside in pages or worksheets. Occasionally, we want to write a formula that refers to a cell that is on a different page. To do this, we need to specify the page, as well as the column and row. This is called a three-dimensional cell reference. An example is Data!B6. This refers to the cell that is in a page called “Data”, in column B and row 6. Three-D references are the same as 2-D references except we prepend the name of the page, and then type “!” before the column and row.

Using 3-D references is also useful when you have a page of data that has a lot of rows and columns to start with, and you need to perform a separate computation on each cell. For example, converting Celsius temperature data to Fahrenheit, or expressing raw numbers in terms of percentages. Sometimes it’s convenient if the results are on a separate page so that you don’t clutter your spreadsheet or make it unwieldy.

Example #5 – Summing sales data

A typical example of where we may want to use a 3-D reference is when we have a spreadsheet that has several pages that are organized identically, and we want to create a final summary page. Open the file [sales.xlsx](#). This spreadsheet has four pages. The first 3 pages show sales figures for the first three months of the year. The fourth page is intended to summarize the figures into quarterly sums.

First, we’ll find the sales total for men’s clothes at the Miami store. Then, we’ll copy this formula to the other cells in the table. In the “total” page, click in cell B5. Type “=” to begin the formula. Click the page tab for January, and select the B5 cell on that page. The formula bar should now read “=January!B5”. In the formula bar, type “+” because we want to continue adding more values. Next, click on the February page tab. Select cell B5. Now, the formula should read “=January!B5 + February!B5”. Edit the formula to put a + at the end, and then click on the March page tab. Select B5 on the March page. Now that the formula is finished, hit enter. You will be taken back to the total page, and you should see that the Miami men’s clothing total is \$71,454. Copy the formula in B5 to the other green cells in the total page. What is the total sales figure for the entire chain? _____
Make sure the columns are wide enough to display all the formula results. Save and close sales.xlsx.

Summary

- Subtracting dates
- If you need to copy a subtotal to another area of your spreadsheet, such as onto another page, you need to remember the technique of using “Go To Special” and selecting “Visible cells only” that we practiced in the example earlier.
- Concerning pivot tables:

- Most pivot tables rely on 3 columns (or “fields”) from your page of raw data. One is used for row categories, one is used for column categories, and one is used for the values to be aggregated (e.g. summed, averaged, counted) inside the table. In this way, the pivot table is used as a 2-dimensional subtotal.
- When you are ready to prepare a pivot table, think about which 2 columns you want to group your data by. Make sure these columns only have a few distinct values. Otherwise your pivot table will be extremely large and have mostly blank entries. To solve this problem, create grading scale(s) like we did in the College example.
- When Excel creates a pivot table, it sorts the categories in alphabetical order. So, if you have created your own set of categories (e.g. “small”, “medium”, “large”), then you need to alphabetize them like we did in the college spreadsheet earlier. For example, you can prepend your category names with letters (A, B, C, ...) as needed.
- If you want to keep around the results of a pivot table, here is a technique to save some space: copy and paste special the contents of the pivot table onto a new page. Then, you can delete the original pivot table. I mention this tip because some pivot tables use up a large amount of memory.
- Note that it’s possible to convert a pivot table into a chart. The result is called a Pivot Chart.
- Using sumif with 3 parameters is an effective way to perform a conditional sum.
- Three-dimensional cell references

Review Questions

1. Looking back at college2006.xlsx, what is another question that can be answered using a pivot table?
2. Suppose you print out a spreadsheet today, and six months from now you print out the same spreadsheet again. If you never made any changes to the formulas, explain (using a realistic example) how the values in the spreadsheet could change.
3. A gambler goes to Las Vegas and records his progress in the following spreadsheet. What formula should be entered in cell C6 to compute the total net winnings? Use sumif functions.

	A	B	C	D	E	F	G	H
1	Las Vegas results							
2								
3	won	lost	lost	won	won	lost	won	lost
4	\$ 500	\$ 750	\$ 200	\$1,000	\$ 50	\$ 300	\$ 200	\$ 700
5								
6	Net result =							

4. Use the Excel help feature to learn about the SUMIFS function. How does it differ from SUMIF? How does it compare to using a pivot table? Explain how could we make use of SUMIFS in our

ag college.xlsx spreadsheet: think of a question we could ask about the data where SUMIFS would come in handy.

5. The file [temp.xlsx](#) shows average high and low temperatures for many US cities. The data is arranged by month. The first page shows all of the average high temperatures. The second page shows low temperatures. Notice that the arrangement of data on these two pages is the same. We'd like to create a third page of data that will average the corresponding high and low temperatures to produce the average temperature by city and month. First, copy the text data in row 1 and column A onto the "average" page. Next, go to the average page and click in cell B2. Enter a formula that will average the corresponding values on the other two pages. Copy this formula to all the other relevant cells in the page. You can make the data more attractive by adding some conditional formatting

Excel lab #5

In today's lab, we will look at various alternative formats for storing data, and how we can convert them to Excel's native format. We will also investigate text functions, scenarios, and the solver tool.

Electronic data can be stored in a number of formats. In this lab, we will look at 3 alternative formats and how to convert them into an ordinary Excel spreadsheet. Specifically, these types are:

- Plain text
- Comma separated values (CSV format)
- HTML table (on a Web page)

Example #1 – using Import Text Wizard

If data is contained in a plain text file, we can open it in Excel. It does not matter what the file name extension is. Once you open the file inside Excel, we can immediately convert it into a proper spreadsheet by telling Excel where the columns are. We'll see this in our first example.

Before we begin, take a quick look at the file [gradedemo2.txt](#). This is our original data file, which is in plain text format. Close this file so we can prepare it in Excel. Open Excel from scratch. You can do this from the Start menu and selecting All Programs → Microsoft Office → Microsoft Office Excel 2007. Once Excel opens, click on the Windows icon button in the top left corner. Choose Open, and then you will see an "Open" dialog box appear. At the top, where it says "Look in", you will need to navigate to the folder containing your spreadsheet files. Then, in the box labeled File name, type "gradedemo2.txt".

When you open a plain text file in Excel, you will automatically see the Text Import Wizard. This “wizard” will lead you through the process of converting your text file into a regular spreadsheet format which you can then save as a new .xlsx file. The wizard tells us that there are 3 steps.

Step 1 – Here, the wizard is asking if the data on each row is separated from each other by a “delimiter”, which could be some special punctuation symbol like a space or a comma, or if the data is merely arranged visually into columns. In our example, our data is in fact arranged into columns, so make sure the “Fixed Width” option is selected. Click the Next button.

Step 2 – This is the most important step. Here is where we have to tell Excel where the columns are. In the “data preview” section, you should see some black vertical lines. At the top of these lines are arrows pointing to numbers along a ruler indicating the character position on the line, in case you are interested in this much detail. These black lines are the boundaries between different columns. If you see 4 lines, this means there will be 5 total columns of data. Essentially, all we need to do in this step is make sure that we have enough lines, and that we don’t have a stray line in the middle of some data. Often, the default placement of the vertical lines is correct. In this case, we want separate columns for the person’s first and last names, plus the 3 individual test scores. The directions at the top indicate what to do if any line happens to be drawn in the wrong place. Click “Next” to continue.

Step 3 – Here, we have an opportunity to exclude any columns we wish. Often, there is nothing to do in this step (as we can always delete columns later), so click the Finish button.

Your data is now in proper spreadsheet format. Save it as an XLSX file, and close. In your folder, you should see the two files, gradedemo2.txt and gradedemo2.xlsx. What are the sizes of these two files?

Comma Separated Values

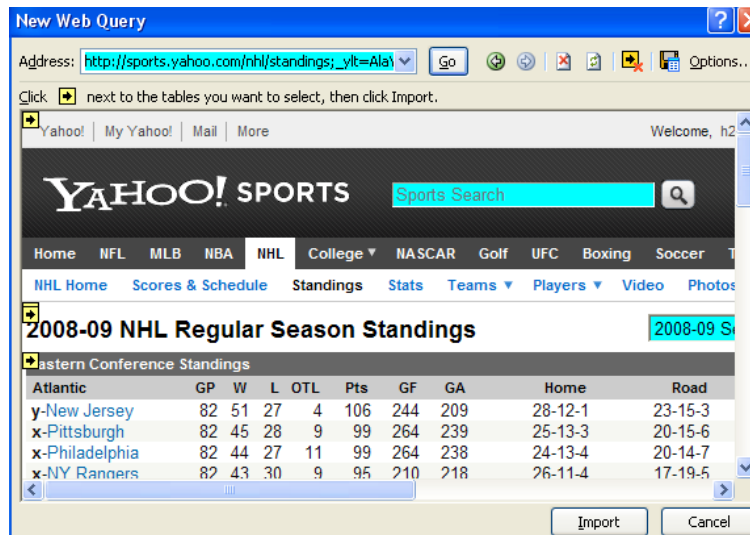
Another format is CSV, which stands for “comma separated values”. This is a plain text document that Excel can immediately interpret. As the name implies, each line of data has commas separating the data. For example, if the first line of the file contains “5,6,7”, then these three numbers will go into cells A1, B1 and C1, respectively.

Open the file [gradedemo3.csv](#). Inside Excel, it looks like a regular spreadsheet! The difference is that CSV can only store values: it cannot store any formatting information or formulas. The advantage of CSV is that it can store data more efficiently than the usual XLSX format. Usually when working with CSV files we want to convert them to XLSX format so that we can modify and add features to it. Close gradedemo3.csv.

Example #2 – a Web query

Sometimes a Web page has data we want, and it's organized as a table. In many cases, we can download just the table of data and convert it to a spreadsheet. This is a very convenient feature! As an example, let's download the NHL standings.

Use a Web browser to go to this page. <http://sports.yahoo.com/nhl>. Click on the Standings link. This page has a large table of data that we want. Highlight the address bar, and hit CTRL-C: we want to copy this Web address soon. Now, open Excel from scratch. Inside Excel, go to the Data menu, and click on Get External Data. From the menu, select "From Web". Now, a new window pops up called "New Web Query". In the address bar at the top, paste the address that you had copied from the browser by hitting CTRL-V. Then, click the Go button. The NHL standings should now appear inside this small window. It should look something like this:



Under the address bar, there is an instruction telling us to click on a little yellow square containing a black arrow. These yellow squares are located at the top left corner of anything that Excel thinks is a table of data. Click on the yellow square before the heading "Eastern Conference Standings". Then, click the Import button. Finally, Excel will ask you where to put the data on the spreadsheet. The default location, A1, is fine. Click OK, wait a couple of seconds, and your data arrives!

Save the file as NHL.xlsx and close.

Text functions

It may surprise you to learn that Excel can do more than just numbers: it also has functions that can work on text data. For example, looking at the NHL standings above, you may notice that a team's record is often expressed as a string, such as "28-12-1". If we wanted to do arithmetic on the numbers in these strings, we would need to extract the numbers from them first. Excel has specific functions *left*,

right and *mid*, that allow us to extract part of a string of text. And there is a function *search* that allows us to find where some text is located inside a string.

To illustrate this general idea, suppose that cell D7 contains this phone number: “(305)555-8050”. Here is how we can use some text functions:

Function call	Return value	Meaning
=search(“)”, d7)	5	Find position of the first ‘)’ character in d7
=left(d7, 5)	“(305)”	Give the first 5 characters of d7
=right(d7, 5)	“-8050”	Give the last 5 characters of d7
=mid(d7, 6, 3)	“555”	Give the part of d7 that begins with the 6 th character, and give a total of 3 characters.

If the text we are extracting is actually a number, then we should also use Excel’s *value* function. Otherwise, Excel will treat the number as text and you will not be able to perform any arithmetic on it. In the above example, if we wanted to treat the result of =mid(d7, 6, 3) as a number, we would say: =value(mid(d7, 6, 3)), and the result would be the number 555, as opposed to the text string “555” (without the quotes). You can tell the difference between a number treated as text versus treated as a number: Excel by default will left-justify text, and right-justify numbers. If you see “555” left justified, it’s likely to be in text mode.

Example #3 – win/loss record

Let’s practice text functions with a simple spreadsheet example. Open the file [win-loss.xlsx](#). It shows a team’s win/loss record for 6 consecutive seasons. The coach’s job is on the line: does the team have an overall winning record? We’d like to add up all the wins and all the losses to produce a 6-year total win/loss record.

In cell C1 enter “Wins” and in cell D1 type “Losses”. Our data starts on row 2. When we look at our first win-loss record in cell B2, “15-3”, we should consider how we can extract the numbers 15 and 3. These two numbers are separated by the hyphen. The first thing that probably comes to mind is that we should use the *left* function to get the 15, and the *right* function to get the 3. But we need to know exactly how many characters to extract. Will the numbers of wins and losses always be 2 digit numbers? Try the following: enter the formula =left(b2,2) in cell C2 and the formula =right(b2,2) in cell D2. Copy these formulas down to row 7. Can you explain what is wrong with the result? _____

It turns out we need to do an intermediate step. We need to find the exact position of the hyphen inside the win/loss string. Insert a new column between B and C. In cell C1 type “where’s hyphen”. Our wins and losses will now be figured in columns D and E. In cell C2, we need to enter a formula that will give us the position of hyphen in cell B2. Use the *search* function to accomplish this. You may want to refer to the phone number example to see how the search function works. After you’ve entered the

formula for C2, the result should be 3. Just for looks, horizontally center the number, and if necessary, format this number to have 0 decimal places. Copy the formula down to C7. Do the results make sense? Why are the results not all the same? _____

Next, click on cell D2. This cell still has the incorrect formula `=left(b2,2)`. Change the second parameter of this function from 2 to C2. In other words, we don't always want the leftmost 2 characters: we want the number of characters to depend on where the hyphen is. Copy your formula from D2 down through cell D7. Uh-oh! We made another little mistake. What is wrong? _____

You need to make a small modification. The correct formula is `=left(b2,c2-1)`. Finally, we need to use the value function to convert what we are extracting into a true number. This means the formula should be `=value(left(b2,c2-1))`. Copy this formula down to D7. Notice that your numbers have been automatically right justified, because Excel now understands they should be treated as numbers instead of text.

We are now ready to handle the number of losses. This is the number after the hyphen. It turns out that the *mid* function is easier to use than right for this purpose. Recall that the *mid* function takes 3 parameters: the cell we extract from, the beginning character position, and the number of characters we want. The nice thing about *mid* is that we can always ask for "too many" characters. For example, if you say `=mid(d7,10,50)` this means that you want 50 characters from d7, starting at the tenth character. But what if d7 doesn't have that many characters? No problem – Excel just starts at character 10 and goes to the end. The result will contain not necessary 50 characters, but *up to* 50 characters.

Click in cell E2. We want to enter a formula using the *mid* function. The text we are extracting from is B2. But where is the starting position? Is it the location of the hyphen? No – we want to start extracting at the character just after the hyphen. So, we will need to add 1. Try this formula: `=mid(b2,c2+1,2)`. The "2" as the last parameter means that the number is "up to" 2 characters. We could have used any number larger than 2 if we wanted. Next, we need to enclose this formula in a call to the value function just like we did for the number of wins. Copy this new formula down column E to cell E7.

Now that we know the number of wins and losses for each season, we can easily add them up. In cells D9 and E9, enter appropriate sum formulas. What is the coach's overall record? _____

Concatenation

Now that we know how to extract parts out of strings, we can experiment with putting them together. This is easy: we use the *concatenate* function, and specify the strings we want to combine. And this function can take as many parameters as we need. For example, if A1 = "dog" and A2 = "house", then `=concatenate(a1,a2)` would return "doghouse". In our win-loss example, we can put the overall record as a string in cell B9 based on the numbers we determined in cells D9 and E9. We just need to put a hyphen between the two numbers. In cell B9, enter this formula: `=concatenate(d9, "-", e9)`.

Save and close win-loss.xlsx.

Example #4 – grade scenarios

Scenario Manager

Let's return to gradedemo.xlsx. In this spreadsheet, we had calculated percentage grades for students, and converted them to letter grades using the vlookup function. Before doing anything, let's rename Sheet1 as Grades, and delete Sheet2 and Sheet3. The instructor thinks the grades are a little low, and would like to experiment with a more lenient grading scale. We could edit the threshold values in cells H3:H6 and see what effect this has on the resulting grades in F2:F10.

To expedite this process, we are going to use Excel's *Scenarios* feature. With it, you can change the values of several cells automatically with the click of a button, and immediately see the result. This saves you the trouble of typing several values every time you want to do the experiment.

Specifically, we want to see what happens if the grading thresholds were changed to 55, 67, 77 and 87. From the Data menu, look for the Data Tools group of tools. In it, you should see the What-If Analysis drop-down menu on the right. (You may need to widen the Excel window in order to see it.) Select Scenario Manager. A dialog box appears, and it informs you that you don't have any scenarios. Naturally we'd like to add one, so click on the Add button.

We are going to create 2 scenarios, one for the "regular" grading system, and one called "curve". Each scenario should have a name, and we specify which cells change values, and what those values happen to be. In the text box labeled Scenario name, type "regular" (without the quotes). Next, we need to specify which cells to change. Click the spreadsheet icon on the right side of the "Changing cells" text box. This will allow you to use the mouse to highlight cells H3:H6. Alternatively you could simply type H3:H6 in the text box. Click the OK button. Next, you should see a Scenario Values dialog box. Make sure the values here for cells H3:H6 are 60, 70, 80 and 90, respectively. Click OK.

You are now back to the main Scenario Manager dialog box. It shows that you have 1 scenario defined. Let's define the other one in much the same way. Click on the Add button. Enter "curve" as the scenario name. Notice that the range of cells given in "Changing cells" is already H3:H6. Click OK. When prompted for specific values of H3 through H6, enter: 55, 67, 77 and 87 in the 4 text boxes, respectively, and click OK.

The scenarios have been created! Now we can test them out. To apply the "curve" scenario, select it from the list of scenarios, and click on the Show button at the bottom of the dialog box. Did anything happen in the spreadsheet? You may need to move the Scenario Manager dialog box so that it does not cover up your data. Let's try it again. To go back to the "regular" scenario, select it from the list of scenarios, then click Show. Choose the curve scenario, and click Show. Which letter grade(s) changed?

It may be a little inconvenient to have to go back and forth between scenarios all the time, so Excel has one additional shortcut for us. It's called the Scenario Summary. It allows you to see the results of all

scenarios simultaneously. In the Scenario Manager, click on the Summary button. Excel asks you to confirm which cells contain the final results, which should be cells F2 through F10. Click OK. A new spreadsheet page called "Scenario Summary" appears. (Note that your spreadsheet now has 2 pages.) At first it may be a little difficult to read, so let's center the data. Select cells D3:F15, and center them horizontally. Now it should be easier to tell the difference between the outcomes of the two scenarios. Save and close gradedemo.xlsx.

Example #5 – Using goal seek

You may have noticed that the Scenario Manager was listed under a set of tools called "What If Analysis". Let's look at another simple application, called Goal Seek.

The purpose of Goal Seek is to ask Excel to determine the value of a cell that's being used in a formula, if we know what the result of the formula should be. As a simple example, let's suppose you have two numbers, A1 and A2, and the cell A3 contains the formula $=A1+A2$. If you provide numbers in A1 and A2, Excel can easily determine the result for A3. That's what every formula does. Goal Seek turns this idea around: if we know what the value of A3 should be (the goal), we can have Excel solve for either A1 or A2.

Open [loan.xls](#), a simple spreadsheet we worked on before that uses a loan payment function PMT. Often, we want to set a budget according to the monthly payment we can afford. For example, if we know we can afford a monthly payment of \$800, how much can we afford to borrow? Rather than changing the formula in B5, we can keep using the PMT function, but this time *ask Excel to fill in one of the needed parameters*. PMT requires 3 parameters, and we can let Excel determine the value of one of them, if we have a particular "goal" for the result of the formula. Specifically, we want to know what value of B1 (loan amount) will cause B5 (the payment) to be 800.

Here is how Goal Seek works in action. Click in cell B5. From the Data menu, look in the Data tools command group for the drop down menu called What-If Analysis. From the list, select Goal Seek, and a small dialog box appears. For "Set cell", enter B5 if it's not shown already. For "To value", enter 800. And for "By changing cell" type in B1. Click OK, and Excel determines the solution for B1. Click OK to accept the solution. What just happened is that Excel has entered a specific value in B1, and the PMT formula in B5 automatically updates to reflect this new loan amount. Of course, it's \$800 because that was the goal we desired, and a solution was possible.

Save and close loan.xlsx.

Example #6 – solver

Goal seek is nice in that it can "fill in the blank" and tell you what input value is needed to produce an end result. Excel has another feature called Solver that's even more powerful. It can determine values

for *several variables*, and you can even express limitations on what those input values can be. Let's look at a couple of examples. Open the file [solver.xlsx](#).

This file has 2 pages, each with a different example of using Solver. First, let's look at the "math formula" page. We'd like Solver to help us with this equation: $10x + 15y = 80$. What values of x and y will work? As you probably know, there's an infinite set of solutions, but we just want one. On this spreadsheet page, the coefficients 10 and 15 are in A6:A7, and the values for variables x and y are in cells B6:B7. The sum is in cell B9, which we want to equal 80 by changing values of B6 and B7.

To use Solver, go to the Data menu, and from the Analysis command group, select Solver. If your Data menu does not have an Analysis command group or you do not see Solver, follow these steps: Click on the Windows icon button in the top left corner of Excel. At the bottom of the menu, click on Excel Options. On the left side of the "Excel Options" dialog box, select Add-Ins. From the list of Add-ins, select "Analysis ToolPak", and at the bottom click on the Go button. A new dialog box appears called Add-Ins. Check the box for "Solver Add-in", and click OK. Now, you should be able to see Solver under the Data menu.

When you invoke Solver, a Solver Parameters dialog box appears. We want the "Set Target Cell" to be B9. Next is the "Equal To" section: click on the "Value of" button, and enter the number 80 in the small text box. Finally, we specify which cells are the variables of the problem. In the "By Changing Cells" section, we need to specify B6:B7. Click the Solve button. Solver should have found a solution to make B9 equal to 80. Click OK to accept the solution.

Let's try Solver again, and this time we'll provide some constraints on the variables. We'd like integer values. In the section called "Subject to the Constraints", click the Add button. For cell reference, select the cells B6:B7. In the middle you see a " \leq " operator, which we'll change to "int" which stands for integer. Click OK. We've now returned to Solver's setup of the problem. Click the Solve button, and then OK to accept the solution. This time, the solution should have whole number answers.

Using Solver doesn't prevent us from tinkering with the data ourselves. For example, we can manually enter 11 for B6 and -2 for B7. In fact, this could have been one of Solver's solutions. Now, for practice, let's enter a second constraint to our problem. Let's set an upper limit of 5 for our variables. Invoke Solver again, and under the "Subject to the Constraints" section, click on Add. For Cell Reference, select B6:B7. The operator " \leq " is actually what we want. Under "Constraint", enter the number 5. Click OK, and when we return to the main Solver box, notice we now have 2 constraints. Ask Solver to finish the problem.

Next, we'll look at a more interesting example: the "cookie factory" page. This is a classic case of the need to use something like Solver to find values of several variables. A bakery makes 3 kinds of cookies. All cookie recipes use the same essential ingredients, but in differing amounts. And each type of cookie has a different profit margin. We'd like to compute and even maximize our profit, but we are subject to the constraints of how much of the raw ingredients we have on hand.

The inputs to this problem are cells B12:D12, where we specify how much of each kind of cookie to make. The output is in cell B14, which is our total profit. The sumproduct function turns out to be convenient to use for our computation. Just for fun, enter the numbers 1, 1 and 2 in cells D12, C12 and B12, respectively. Besides computing our profit, you should also notice that the cells F3:F8 have also changed: we must keep track of how much of our basic ingredients get consumed when we prepare the cookies.

Here's the problem: We'd like to maximize our total profit (b14) by changing our mix of cookies that we make (b12:d12), subject to the constraints that the values in f3:f8 are less than or equal to the corresponding values in e3:e8. In other words, we can't run out of any ingredient. We'll use the solver feature to enter this problem. As before, under the Data menu, select Solver. Specify B14 as the target cell. In the "Equal To" section, make sure the Max button is selected. Why do you think we are not choosing "Value of" as we did for the previous example?

The "By Changing Cells" should be set to B12:D12. And for our constraints, click on the Add button and specify: F3:F8 <= E3:E8. What does this constraint mean? It says that each value we specify in column F (the raw ingredients we consume) should be less than or equal to the corresponding value in column E (the amounts we have available).

What is Solver's solution? B12 = _____, C12 = _____, D12 = _____, B14 = _____

If the answer is not realistic, then we need to redo Solver to enter some constraints. As before, we should specify that the numbers should be integers. Also, they all have to be at least zero. In fact, some of your customers may be upset if you discontinue a flavor, so I would say each of the input values should be at least 1. Enter these constraints, and ask Solver to solve again. Enter the new results below:

B12 = _____, C12 = _____, D12 = _____, B14 = _____

Do you make as much money when you have the constraints? Why do you think there is a difference?

Save and close solver.xlsx.

Summary

- Text functions: left, right, mid, search, value, concatenate.
- Don't forget to use the value function if the result should be treated as a number.
- When a spreadsheet is storing names of people, such as students in a gradebook, or employees in a company, it's a good idea to have the first and last names as separate columns. Sometimes you may be downloading a list of names, and it might be necessary to extract the first and last names using text functions.
- The purpose of Scenarios in Excel is for you to be able to automatically change the input values in several cells to see what effect the changes have on output calculations. A Scenario Summary

is even more convenient, because you can see the results of several scenarios all on the same page, without having to manually select and run each scenario.

- While goal seek can solve an equation with one variable, Solver attempts to find a solution when there are multiple variables.

Review Questions

1. The spreadsheet [texas.xlsx](#) has a list of Texas counties along with their 2000 population and 1999 median household income, according to the US Census Bureau. It seems like column D is unnecessary because it simply repeats the county name. We'd like to delete this column. But first, we need to make sure that the corresponding entries in columns A and D are the same. To compare strings for equality, you can use '=' or the EXACT function. Write a formula in cell C2 that tests to see if A2 and D2 are the same string. Copy this formula all the way down the list of counties. Do you see any counties that don't match up? How can the spreadsheet be fixed?

How would we use or modify the formula you wrote in C2 if we wanted to know how many county names in columns A and D did not match? How can we easily locate the row containing the first pair of counties that do not match?

2. Let's continue working with [texas.xlsx](#). We'd like to isolate just the name of the county from the string that contains both the county and state name. This would be especially helpful if we had a list of counties from many states, and we wanted to sort all the counties by state first. Insert 2 new columns to the left of column A. In the new column A, enter a formula that will compute the location of the hyphen in the name of the county/state now in column C. In column B, enter a formula that will give you just the name of the county, without "Texas" and without the comma or any trailing space character. Copy your formulas in columns A and B down for all counties in the spreadsheet to make sure they work correctly.
3. Open the file [tax table.xlsx](#), a spreadsheet we worked on during an earlier lab. We'll use the Scenario Manager and to investigate what would happen if the five income tax percentage rates were each reduced by 1 percentage point. But before you do that, let's make the spreadsheet a little more interesting by allowing it to compute income taxes on three example incomes rather than just one. Copy the formulas in column B to columns C and D. Then, create two scenarios, called "old tax" and "new tax". The new tax rates are 1 point lower than the old rates.
4. Recall the cookie example we did with Solver. Can you think of another type of problem that has a similar structure where we would need to use Solver? In other words, there are several values that need to be determined, and they are all being used in a formula whose value needs to be maximized or minimized.

5. At the beginning of the lab, we saw how text functions could be used to split a string into two parts. But in the general case of having to split text into several parts, we should probably use a feature called the Text to Columns wizard. For example, NHL records consist of 3 values: wins, losses and overtime losses. Use the Excel help feature to learn more about this feature. What is a “delimiter”?

Cumulative Review Exercises

Spreadsheet #1 – California energy: Practice with subtotals, pivot tables and especially charts.

The spreadsheet [calif energy.xls](#) shows the distribution of energy consumption over a 6-year period for a sample of California households. The 24 columns of data correspond to the 24 hours in a day. For example, the 1am column refers to how much energy was used between midnight and 1am. There is a row for each day from 2000 through 2005.

This is a mountain of data, and it would be sensible to encapsulate the information into some useful charts. Create three [line charts](#) that depict the relationships described below in #1-#3. Place each chart on its own worksheet. Before creating these charts, you will need to write formulas that determine summary data, such as an *average*.

1. Show how much energy is used, in general, according to the hour of day and month of year. In other words, in a single chart, we should be able to see how the power demand changes during the day, and how the demand varies according to the time of year. Hint: the month() function allows you to extract the month number from a date, as in the “12” in 12/15/2002.
2. Show how much energy is used according to hour of day and day of the week.
3. Show how much energy is used according to day of the week and month of the year.
4. Finally, create a new worksheet called Documentation. In this worksheet, write down the conclusions you can draw from the charts. Hopefully from doing this exercise you can see how powerful Excel is at helping people understand their data.

Spreadsheet #2 – Furman class schedule: Practice with text functions, subtotals and pivot tables. The file [schedule.csv](#) displays the schedule of all classes taught in fall 2008. We’d like to compute some summary statistics on the data using Excel. First, save this file as .xlsx format.

1. On the left side of the spreadsheet, create a new column called “course”. It turns out that many courses have multiple sections. For example, the section called “ACC-111-01” is a section of the

course “ACC-111”. In the course column, write a formula that isolates just the course code. In other words, dropping the section number and the second hyphen in the string.

2. Again on the left side of the spreadsheet, create a new column, this one called “subject”. Consider the section called “ACC-111-01”. The letters before the first hyphen, in this case “ACC” are the subject of the course. In the subject column, write a formula that will isolate just the subject code for the course. At this point, each row in the spreadsheet will identify a section by subject, course, and then the name of the section.
3. We would like to know what the total enrollment is for each subject at Furman. Use Excel’s subtotal feature. Copy the subtotals onto a new worksheet, and make sure that you are only copying the subtotals. Sort these subtotals in descending order so that we can easily see which subjects are most popular.
4. Repeat step 3, but this time find subtotals for each course at Furman, rather than each subject.
5. We’d like to examine the distribution of enrollments according to the time that classes are offered. For example, we’d like to know which days and times are most often used. The first thing to do is to clean up the “meeting times” column. We are mainly interested in the start time of a class. For example, in the string “MWF 09:30AM – 10:20AM (more)...”, we only want the part that says “MWF 09:30AM”. Create a new column next to the “meeting times” column, and call it “start time”. Write a formula that extracts only the start time of the class. Once you have this column completed, you are ready to create a pivot table. Create a pivot table that is based on the start time and enrolled columns. Copy the pivot table to a new worksheet. Use conditional formatting to highlight the total enrollment figures that are over 500.
6. Finally, let’s create a pivot table that shows how the various rooms in Riley Hall are used throughout the day and week. Create a pivot table that is based on the following columns: “enrolled”, “location” and “start time”. In the pivot table, make sure it doesn’t list every single room on campus: only check the rooms whose names begin with “RLY” and leave the others unchecked.

Spreadsheet #3 – Hearts scores – practice with Excel functions, absolute references, charts

Who’s the best player at Hearts? Suppose last summer a friend of yours named Sam played Hearts against the computer 150 times. The scores are tabulated in [hearts.xlsx](#). The computer opponents are called Pauline, Michele and Ben. Sam is interested in how well he did overall against the computer. Modify the spreadsheet hearts.xlsx as follows:

1. The shaded areas of the spreadsheet contain blanks we need to fill. Use formulas to calculate the average score for each player over the 150 games, and enter these in the range b9:e9.
2. Use formulas to calculate the winning (i.e. lowest) score for each game, and store these in column F next to the scores.
3. In columns G-J, enter formulas that determine the rank of each player in each game. For example, we want to say that in game 1, that Sam came in first (1), Michele came in second (2), Ben came in third (3) and Pauline came in fourth (4). Use Excel's =rank() function. Note that there can be ties. So it is possible for there to be two players coming in first, and then the other two players come in 3rd and 4th. The rank function determines this automatically. To improve readability, you may insert blank columns – in this case the information in columns G-J will be located farther to the right.
4. In the range g4:j7, use formulas to count how many times each player came in first, second, third and fourth.
5. Give "Sheet1" a more descriptive name. Delete sheets 2 and 3. Create a new sheet called Documentation. On this sheet, enter your name and the date, and write some comments about the results of the formulas that you observe. In particular, who wins the most games, who comes in last most often, etc.
6. Create a chart that shows the average score of each player. Sam will be especially interested in seeing this. Put this chart on its own page.
7. Judging from the overall result, it looks like Sam had an easy time against the computer. But was this always the case? Sam recalls that during the summer, Pauline was giving him problems.
 - a. To the right of each game, for example in columns L through O, enter formulas that determine the *running averages*. For example, we're interested in the average score that each player has attained since the start of the summer – after each game.
 - b. Pauline was Sam's closest rival. Create a new column that keeps track of the difference between Sam's and Pauline's running average.
 - c. Create a chart that depicts how this rivalry changed over time. In other words, show their running average scores as a line chart. Put this chart on its own sheet.
 - d. At what point did Pauline have her greatest overall lead over Sam? When did Sam finally pull ahead of Pauline for good? Did Sam pull ahead of Pauline because Sam's scores improved, or because Pauline faltered? Enter these answers in your Documentation sheet. You may want to highlight the relevant cells in your spreadsheet to draw the reader to the most interesting numbers.

Sam is interested in other statistics. Use your knowledge of Excel functions and features to answer these questions. In your Documentation sheet, give answers to these questions, and explain how you used Excel to determine them.

8. What is the average score when each player wins? (In other words, you need to find Sam's average score in games that he wins; Pauline's average score in games where she wins, etc.) Also determine the average score when each player loses. These statistics may give some insight into how aggressive the players are.
9. For each game, calculate the difference between the best (lowest) and 2nd best (second lowest) score. If there is a tie for the win, this difference should be considered to be zero.
10. Who tends to win close games? Who tends to win blowouts? You may define a "close" game as one where the best & second best score are below some threshold. Similarly you could define a "blowout" as when the difference between the highest and lowest is quite high. Can your answers support the conclusion that Sam is an aggressive player at Hearts, or that he plays conservatively?

Spreadsheet #4 – Looking for a house

The spreadsheet [house.xlsx](#) is a list of all the houses in the Greenville area that were on the market in early 2005.

1. Column C shows the square footage for each house. But it does so as a range, as in "2800-2999". This range is a text string and not a number, so it was necessary to convert this range to a number by using a formula in column D. This formula uses the functions `if`, `left`, `mid`, `search` and `value` as well as the numbers 7, 1, 5 and 99. Examine this formula carefully. Describe in step-by-step terms what this formula is saying, and why it is correct.
2. Between columns B and C, create two new columns in the spreadsheet. The new columns will store the number of bedrooms and number of full bathrooms in each house. Ignore half baths. You will need to use a text function to extract a number from a string. Also, don't forget to use the `value` function to convert the final answer to a number.
3. Create a pivot table that determines the average price of a house based on the number of bedrooms and full bathrooms inside. Create a second pivot table that determines *how many* of each type of house we have. Copy these two pivot tables onto a new worksheet so we can easily see them side by side. (Hint: When you do the copy-paste, tell Excel that you want to "paste special", and paste only the values from the pivot table.) The page with the two copied pivot tables should be called "average by bed and bath".
4. Go back to the original list of houses, and let's do some filtering. We want to show only those houses that are between \$200,000 and \$350,000 where the approximate square footage is at least 2500, and the lot size is at least 1 acre. Once you have filtered the data, copy the resulting cells onto a new page called "house hunting". On the original data page, be sure to remove the filters. On the house hunting page, sort the houses descending by price.

Spreadsheet #5 – Law school grades.

Suppose that you work for the dean’s office in a law school. A semester has just ended, and professors have turned in their grades. Your job is to compute summary statistics about the college’s grade distribution. Open the file [law school grades.xlsx](#).

1. Each row in the spreadsheet shows a particular course that was taught. Starting in column J are the individual student grades. The first thing to do is to generate a raw grade distribution for each course. In columns D-H, enter formulas that will compute the total number of A’s, B’s, C’s, D’s and F’s in each class. In column I, enter a formula that will give the total number of A-F letter grades. Note that other letters such as S are not going to be used for this exercise.
2. In column C, enter a formula that will determine the average grade (GPA) for each class, using A=4, B=3, C=2, D=1 and F=0 quality points. Use conditional formatting to highlight the distribution of course GPA. Define a color spectrum that uses a blue background for the lowest GPA and a yellow background for the highest GPA.
3. In row 74, enter formulas that will show the totals for the entire law school. The cell C74 would then give the average grade given by all instructors.
4. The law school dean is especially interested to know which courses are the easiest and which are the most difficult. In column B, enter a formula that determines the quality point surplus that each class has over the average GPA. According to your analysis, which are the easiest and hardest law school classes? Are they necessarily the ones with the highest or lowest GPA? Explain: _____

Spreadsheet #6 – Analyzing census data

The Website factfinder.census.gov is a vast repository of data from recent censuses. In this example, we will examine the question of how long it takes people to commute to work. Here, you will download data from the Census Bureau, and then make some summary calculations based on the raw data, and hopefully draw some meaningful conclusions. Note that the Factfinder site may change its interface from time to time, so you may want to take the downloading instructions with a little grain of salt.

1. From the main factfinder.census.gov page, click on Data Sets. When the new page appears, click the button for “2000 Summary File 3”. On the menu of links to the right, choose “Detailed Tables”.
2. We want to select all the counties in 3 neighboring states (your choice). Select the tab called “geo within geo”. The geographic type will be: county within state. In other words, it should read “Show me all counties within state ...”) Choose whichever states you wish, and

choose *all* the counties in that state. Once you've selected all your counties, click the "next" button.

3. Select table P31 called "Travel Time to Work for Workers 16+ Years...". Click the "Add" button, and then "show result". From the Print/Download menu, click on Download. When the dialog box appears, choose "Comma delimited (Transpose Rows/Columns)". If necessary, uncheck the box that says only the table on the screen, because we want *all* the data. When you save the file, make sure it goes in the proper folder (e.g. the same folder you've used for other Excel files) and give the file an appropriate name such as commute.csv.
4. Open the CSV file in Excel, and save it as type XLSX, in other words a regular spreadsheet. Now you are ready to do the fun part!
5. Between columns A and B, insert ten new columns, and use Excel formulas to compute the following for each county: The number of people who ...
 - Work at home (zero commuting time)
 - Commute less than 15 minutes to work
 - Commute 15-30 minutes to work
 - Commute 30-45 minutes to work
 - Commute 45 minutes or more to work

The next 5 columns should be the corresponding percents of the total number of workers. Now, sort the counties by: state alphabetically, and within each state by the percentage of long (45+ minute) commuters in descending order.

6. Use conditional formatting on the column showing the percentage of long commutes. Use a color threshold with blue for the lowest percentage and yellow for the highest percentage.
7. Have Excel create subtotals for each state. You need to compute a subtotal for all the numbers, but be careful with the percentages. You'll need to re-enter or copy the appropriate percent formulas. Copy the three subtotal rows onto a new worksheet (e.g. Sheet2, and give the sheet a new name.) Remove the subtotals from the original data page.
8. For each of the 3 states, make a pie chart that shows the overall distribution of commute times. Note that the pies should have 5 slices, to also include the people who work at home. Put these charts on the same page as the new worksheet that has only the subtotals. Make sure the pie charts are big enough to see, but small enough so we can see all 3 pies at the same time. Be sure the charts have appropriate titles, so we know which state is which.
9. So that we can better see the pies, it may be a good idea not to label the slices of the pie or have a legend. Instead, type in the spreadsheet itself a guide for which color represents each level of commute time. Use your best judgment as to how best to convey the information in the charts. (For example, have a legend on just one of the pies, not all 3 which may be redundant.) In your own words, what is the most interesting or surprising facts about the data? _____

How do the states compare to each other? _____

Congratulations! You have completed your introduction to Excel, and by now you are quite proficient at doing routine tasks with spreadsheets. Please keep this handout for future reference.

Where can you go from here? If you are interested in learning more about Excel beyond what we do in this course, a good reference book is the *Excel 2007 Bible* by John Walkenbach. You may also find the Excel “help” feature useful from time to time. Here are some concepts we did not do in the labs that you may find useful in your future Excel work.

1. Database functions, such as DSUM, DAVERAGE, DCOUNT, etc. These are similar to functions you have seen like countif and sumif. The advantage of “D” functions is that they are good at capturing multiple criteria joined by “or” relationships. For example, if you wanted to find the total number of children who are enrolled either in a Greenville county school, or in any *high* school in South Carolina. Another example would be to find the average GPA of students who are either business majors or classified as seniors.
2. The OFFSET function. This function is used when you want to return a value located in a cell located near a value returned by another function. For example, you might have a spreadsheet depicting baseball statistics. In column A, you’d have the player names, and in column B you’d have their batting averages. You can use the MAX function to find the value of the highest batting average. But how would you get the corresponding player’s name? You would use the offset function to look one cell to the left of the batting average to get the correct name.
3. Additional formatting and printing options. There are many things you can do in Excel to make your data more eye-catching. For example:
 - a. Formatting features that we also see in Word and PowerPoint, including text boxes and geometric shapes.
 - b. We can also create custom headers and footers on printed pages.
 - c. Displaying custom borders around or beside your cells. It’s sometimes useful to have a thick line between rows or columns to divide your data into logical sections to make the spreadsheet easier to read.
 - d. Merging and centering cells. This is useful at the top of a worksheet when you want to center a title across several columns.