

Lab 3.1

Spreadsheet Basics

Spreadsheets and databases are very similar applications. They are both useful for the *organized storage of information*. However, spreadsheets are simpler than databases and are typically used for different purposes, so this lab will discuss when spreadsheets are used and the basics of how to use them. We will use the popular Microsoft Excel spreadsheet application in this lab, but these concepts and procedures will be useful for using any spreadsheet application. At the end of this lab, you should not only be able to use a spreadsheet, but also be able to recognize problems that a spreadsheet can be useful for solving. Finally, working with spreadsheets will prepare you for databases, the focus of the next labs.

Vocabulary

All key vocabulary used in this lab are listed below, with closely related words listed together:

- worksheet
- cell
- row
- column
- data type
- formula, parameter

Post-Lab Questions

Write your answers after completing the lab, but read them carefully now and keep them in mind during the lab.

1. Suppose you want to start writing some notes on a book you are reading, and you want to store them on your computer. Which would be a better suited program to use for this purpose: a word processor or a spreadsheet? (You may answer that both are equally well/badly suited, if you wish.) Explain your answer.

2. Spreadsheets allow us to present charts, graphs, and other visualizations of data. Describe one reason why the ability to visualize data can be valuable.

Discussion and Procedure

Part 1. What are spreadsheets good for?

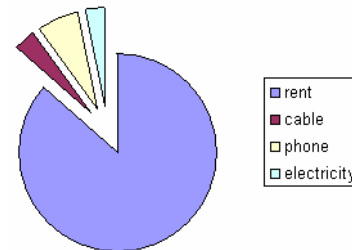
Structuring data. Spreadsheets are most commonly used to store and perform calculations on numeric data. Examples of data like this include a checkbook record, business transaction records, data from scientific experiments (e.g., daily rainfall), results of a survey. The first advantage of spreadsheets is that data can be stored in a structured way, making it easy to retrieve and present the data in different ways. Imagine, for example, keeping track of your expenses in a notebook, where you might list your expenses like this:

17 Jan 2002, phone bill, \$32.85
 electric bill, 25 Jan, \$18.47
 rent, Feb 1, \$640
 Feb 12, cable TV, \$26
 Feb 18, phone, \$52.02
 Feb 26, electric bill, \$21.88
 (etc.)

Storing the same data in a spreadsheet has several advantages over handwritten a record, many of which come from the structure that spreadsheets put on data. A spreadsheet is organized like a table, with *cells* arranged in *columns* and *rows*. This overall structure keeps the format of the entries consistent and makes the data more readable.

	A	B	C
1	date	expense	amount
2	01/17/02	phone	\$32.85
3	01/25/02	electricity	\$18.47
4	02/01/02	rent	\$640.00
5	02/12/02	cable	\$26.00
6	02/18/02	phone	\$52.02
7	02/26/02	electricity	\$21.88

Feb 2002 Expenses



In addition to the overall structure provided by rows and columns of cells, you can specify the *format* (e.g., bold, italics... just like in a word processor) and data *type* (e.g., date, text, integer, real number, monetary amount) of a cell. In the example spreadsheet shown above, the cells in the first column are formatted as dates, so they are all displayed with the same month/day/year format. Similarly, the cells in the third column are formatted as currency amounts (in dollars), so they are all displayed with two decimal places and the dollar sign.

Structuring data this way has many advantages. You can use a spreadsheet to sort your data for different views of your data. You can also easily create visualizations of your data, in the form of bar, line, pie, and other graphs. We will see both of these features in this lab.

The main feature of a spreadsheet, however, is the ability to set up calculations on data that are automatically recomputed when you modify the data. In the example above, you might be interested in the total expenses. Later in this lab, you will see how to set up a cell whose value is automatically calculated from the values in other cells. Spreadsheets are often used to store business records and scientific data, because computations like averages, standard deviations, totals, etc. are commonly useful with such data.

Part 2. Creating and Editing a Spreadsheet

In this part, you will create a spreadsheet for keeping a daily record of how many hours you spend doing various things (e.g., studying, watching television, talking on the phone). This might be useful for reexamining your time management strategy. Just as with programming, you should have a plan before starting work on your spreadsheet.

1. *Decide what data your spreadsheet will store.* In this step, you should decide how your data will be stored in rows and columns. In this case, it seems sensible to store each day's data in a row. (Technically, you could also have one column per day, but scrolling from left to right as you add or read data is less common than scrolling from top to bottom.) With one row per day, you should next decide what data you should have for each day, which determines what your columns will be. Pick two activities you regularly engage in. With each row, you should store, in addition to the date, the number of hours spent that day on each of these activities, for a total of three columns.
2. *Start Excel.* Microsoft Excel is usually found on the Start menu under **Start \ Programs \ Microsoft Office \ Excel**. Ask your instructor if you cannot find Excel in the **Start** menu.

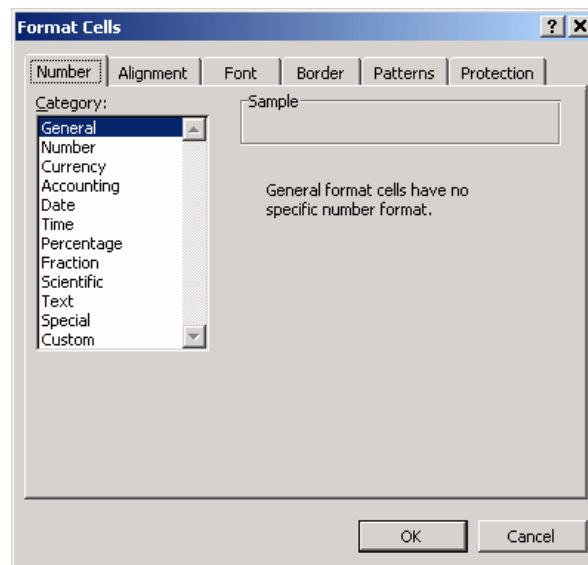
Most of the Excel window is occupied by your “worksheet,” the grid of cells where you will record your data. Just like many Microsoft applications, there is a menu bar at the top, with a toolbar of buttons just below. Below the toolbar is an Excel-specific bar called the **Formula bar**, which we will use later. We will only use a few of the most essential features of Excel in this lab, although it does offer many more.

Note that each column can be identified by a letter, and each row by a number. This allows each cell in the grid to be identified by a letter-number combination. This will be useful later when defining formulas to perform automatic computations on your data. However, since column letters alone are not useful for reminding us what data is in each column, we begin by using Row 1 of the worksheet to name our columns.

3. *Enter column names.* Editing a cell is as easy as clicking the cell or using arrow keys to move to the cell you wish to edit, then typing the data value that you want in the cell. As shown below, set up column names in the first cells of Row 1 as shown below, using the names of the activities you chose in Step 1 above.

	A	B	C
1	date	TV	studying
2			
3			

4. *Format the columns according to data type.* Select the “date” column by clicking the gray “A” column label. From the **Format** menu, select **Cells...** This should bring up the **Format Cells** dialog, as shown below. Under **Category**, select **Date**, and in the **Type** list that appears on the right side, select one that includes the year, in addition to the month and day. (Any one of these types will do.) Click **OK** when you are done and format your two activities columns to use **Category** “Number” with 0 decimal places, making the assumption that your timings will just be rounded to the nearest hour.



5. *Enter a row of timing data.* Begin by entering today’s date in Row 2’s “date” column. Note that because this column is formatted to store dates, you can enter the date in a variety of different ways and the spreadsheet will automatically display it in the format you selected. For instance, try entering “1/4/02”, “1-4-02”, “jan 4, 02”, or “4 jan 2002”. Excel will interpret what you enter as a date and convert it to the format you selected. For the other columns, for the purposes of this lab, just make up numbers for the number of hours for each of your activities.

Assuming you intend to record data on a daily basis, you will have to fill in each day’s date in the “date” column. Typing these in by hand is not only tedious, but error-prone,

so we will use a feature called “fill” to automatically enter a series of dates in this column.

6. *Select about seven cells in the “date” column, starting from the cell containing the date you just entered in the last step.* To select a group of cells, use the mouse and drag over the cells you want to select. You can select any rectangular region of cells this way.

ALTERNATIVE: First, decide which rectangular region of cells you want to select. Click the cell in the upper left corner of the region you want to select, then while holding down **Shift**, click the lower right corner of the region. (You can actually do this with any two opposite corners of the region you want to select.)

	A	B	C
1	date	TV	studying
2	1/1/2002	0	3
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			

Before you go on to the next step, make sure that you have the cell with the date you just entered and several cells directly below it selected.

7. *Fill the selected region.* From the **Edit** menu, select **Fill \ Series...**, and make sure **Series in Columns**, **Type Date**, **Date unit Day**, and **Step value 1** are selected. These options specify that (1) the series you want runs down a column, as opposed to across a row, (2) that the data type in these cells is Date, and that (3) the dates in this series should go up by one day per cell.

	A	B	C
1	date	TV	studying
2	1/1/2002	0	3
3	1/2/2002		
4	1/3/2002		
5	1/4/2002		
6	1/5/2002		
7	1/6/2002		
8	1/7/2002		
9	1/8/2002		
10	1/9/2002		
11	1/10/2002		
12			

8. *Enter timing data for four more days.* Again, for the sake of this lab, just fill in estimates.
9. *Delete the date values for the remaining rows.* For now, we’ll just work with the rows of data that are completely filled, so clear the date values for the remaining rows, if any. Just select the cells you want to clear and press **Delete**.

Part 3. Functions for Automatic Calculations on Your Data

For data like this, it might be useful to calculate values like the total hours and the average hours spent daily on each activity. These values are different from the data in that they are not values that you enter directly but are calculated from the data you enter. Using a *function*, you can set up a spreadsheet cell to display the result of such a calculation. A spreadsheet function is just like a formula or mathematical function in that it is used to compute a value based on the values of one or more variables. In the case of a spreadsheet function, you can set up a cell to perform some computation (like finding the sum or average) on the values in other cells in your spreadsheet. The advantage of a spreadsheet function is that this cell's value is automatically updated whenever the data in the other cells is changed.



10. Set up text labels for the total and daily average number of hours spent on each activity in the "date" column, as shown at the right.

	A	B	C
1	date	TV	studying
2	1/1/2002	0	3
3	1/2/2002	2	4
4	1/3/2002	1	3
5	1/4/2002	3	2
6	1/5/2002	2	1
7	total		
8	average		


11. Select the cell for the total time for your first activity (B7 in the example at the right).
12. Click the **AutoSum** button on the Toolbar, which is marked by a capital sigma (Σ), then press **Enter** or click the green check button. Instead of entering a value in this cell, you have now entered a formula called **SUM**, which adds the values in a range of cells. In this case, the **AutoSum** button has guessed that you want the sum of the values in the cells directly above the currently selected cell. You should now see the total number of hours for this activity as the value for this cell. Click the cell again to see its contents in the formula bar, and note the syntax of the formula in the cell's contents, which should look like, "**=SUM(B2:B6)**".

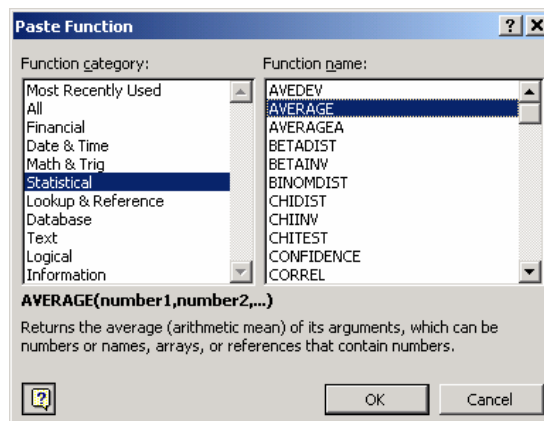
Let's break down this cell's contents to see how it computes the total value shown. The equals sign in "**=SUM(B2:B6)**" means that this cell contains a function, rather than a fixed value. The function (in this case, **SUM**) is indicated after the equals sign, and the "**B2:B6**" part in the function's parentheses is the function's *parameter*. (Parameter is actually a general computing term, and we'll see it again when we work with programming.)

13. Test the automatic recalculation of a formula cell. Change one of the timings for the activity for which you just created a total cell. The total hours cell should automatically update as soon as you enter the new value. If not, clear the contents of the total time cell by deleting it and try again from Step 11.

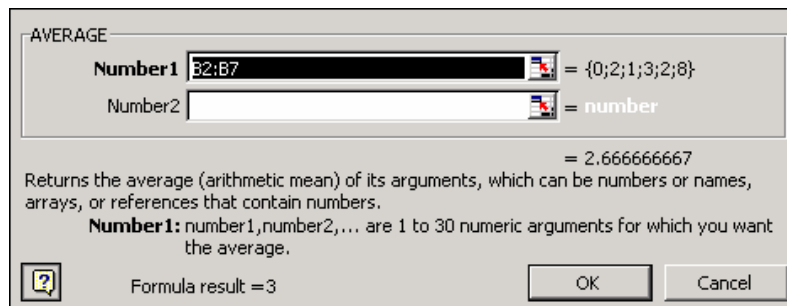
14. *Set up a total time cell for your second activity by copying and pasting the formula.* Although you could follow the same instructions as above to set up your second activity's total time cell, there is an easier way. Select the first activity's total time cell and copy it by clicking the toolbar **Copy** button (), selecting **Edit \ Copy** or pressing **Ctrl-C**. Select the cell where you want the second activity's total time and paste by clicking the toolbar **Paste** button (), selecting **Edit \ Paste** or pressing **Ctrl-V**.

Note that the contents of the new total time cell has a formula that is similar but not identical to the original. Excel guesses that you want the new total time cell's formula to work on the column of cells above it, rather than be an exact copy of the first total time cell.

15. *Set up an average daily time cell for your first activity.* Select the cell where you want the average daily time for your first activity to be displayed. This time, instead of clicking the **AutoSum** button, click the **Paste function** button (). This should bring up the **Paste Function** dialog box, in which you should select **Function category Statistical** and **Function name AVERAGE**. Click **OK** when you are done.



16. *Select the range of cells the average should be calculated over.* When you click **OK**, you should see the dialog box below, which you should use to specify which cells' values you want to average. (You can drag this dialog box out of the way so that you can see your data and the dialog box at the same time.)



Excel has guessed for you that cells B2 through B7 are the ones you wish to average, but this probably includes one more cell than you really want, i.e., the total time cell you created earlier. You can correct this by typing in the **Number1** entry box and replacing B7 with B6, or you can select the contents of the **Number1** entry box and select the cells you want (B2 through B6) on the worksheet.

If you want to go back and change the range of cells for this formula later, you can select the cell and click the = button in the formula bar.

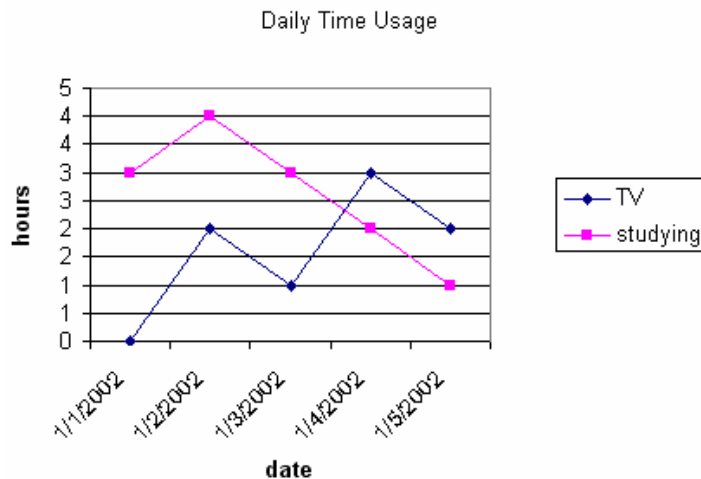
	A	B	C	D
1	date	TV	studying	
2	1/1/2002	0	3	
3	1/2/2002	2	4	
4	1/3/2002	1	3	
5	1/4/2002	3	2	
6	1/5/2002	2	1	
7	total	8	13	
8	average	2		
9				

Formula bar: B8 =AVERAGE(B2:B6)

17. *Reformat the average time cell to show two decimal places.* Recall that the rest of this column is formatted to show no decimal places. Since an average is likely to have a decimal part, you should format this cell to show at least two decimal places. Select the cell and select **Format \ Cells...** or press **Ctrl-1 (Ctrl-one)** to open the **Format Cells** dialog box.
18. *Create a properly formatted average time cell for your second activity.* Just as in Step 14, the copy-and-paste method is fastest.


Part 4. Charts for Visualizing Data

Visualizing data in charts can often make it easier to interpret. In this part, you will use Excel to create a chart showing daily time spent on your activities, like the one shown below.



19. *Select the data that you wish to chart.* Select all of the non-empty cells in your worksheet except for the those in the total time and average time rows. You should include the column headers for “date” and your two activities.

	A	B	C
1	date	TV	studying
2	1/1/2002	0	3
3	1/2/2002	2	4
4	1/3/2002	1	3
5	1/4/2002	3	2
6	1/5/2002	2	1
7	total	8	13
8	average	1.60	2.60
9			

20. *Start the **Chart Wizard**.* Click the Chart Wizard toolbar button () or select **Insert \ Chart....** Under **Chart Type**, select **Line** and choose one of the line **Chart sub-types** that appeals to you. Click **Next >** when you are done.
21. *Verify the data ranges set for charting.* In **Step 2** of the **Chart Wizard**, if you wish, you can modify the section of the worksheet you selected to display in the chart. On the **Data Range** tab, **Series in** should be set to **Columns**, since each column holds a series of data across time. In this case, the default settings should be fine, so you may just click **Next >**.
22. *Title the chart and each of the axes.* You should give the whole chart and each of the X and Y axes titles so that someone reading the chart can immediately understand the data being shown. Consider titling the chart “Daily Time Usage” and titling the X and Y axes “date” and “hours”, respectively. Click **Next >** when you are done.

Although we have only worked with one worksheet, a single Excel spreadsheet file can contain multiple sheets, including sheets devoted to charts. In the next step, you can choose whether to put your newly created chart in the worksheet you are working on or to put it in its own worksheet. You can switch among the multiple sheets in a file by clicking the tabs in the bottom left corner of the Excel window. Your main worksheet is named “Sheet1” by default. (You can rename any sheet by right-clicking its tab and selecting **Rename**.)

23. *Choose a chart location.* For now, use the default choice (**As object in “Sheet1”**) and click **Finish**. The result will be a chart appearing over your worksheet. You can drag the chart to move it or select and drag its corners to resize it.

You can adjust an existing chart’s properties by right-clicking the chart and selecting the options from the pop-up menu.

Submitting your work. Check with your instructor for instructions on how to submit your spreadsheet. You might be asked to save it to floppy disk, e-mail it as an attachment, or print it on paper.

Part 5. Sorting Data

The way you have entered your timing data, the rows are ordered by date. You might, however, be interested in seeing which days you spent the most time on a certain activity. More generally, you might be interested in sorting your rows by a column other than date.

24. *Select the region of cells whose rows you wish to sort.* Select the same region of cells you selected for the chart in Step 19, i.e., all non-empty cells except those in the total and average rows. Make sure you include the first row of column names in your selected region.

25. *Specify the sorting order you want.* Select **Data \ Sort...** and use the drop-down lists to select the column(s) whose values you wish to sort by. In this case, choose your first activity under **Sort by**, and, if you want ties to be broken by sorting by a value in another column, select it in the first **Then by** list. Click **OK** when you are ready to sort.

Note that the rows in the selected region are in the specified order, but your chart does not change. This is because your sorting did not change data in each row; it just reordered the rows. The chart itself is not affected by the row order.

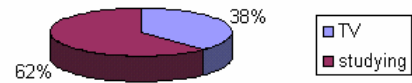


	A	B	C
1	date	TV	studying
2	1/1/2002	0	3
3	1/3/2002	1	3
4	1/5/2002	2	1
5	1/2/2002	2	4
6	1/4/2002	3	2
7	total	8	13
8	average	1.60	2.60
9			

Optional Additional Activities

- Create a pie chart showing how your total time splits across the activities you chose.

Total Time Across Activities



Further Reading

- Do a web search for “spreadsheet history” to find more information about spreadsheets, including the history of the application, which was first introduced in 1978 by Robert Frankston and Dan Bricklin as the application Visicalc.