



Computer Science

COMPSCI 111 S2C - Lecture 3/4

July 2003

The First Killer Application

Spreadsheets

Introduction

The concept of a spreadsheet was developed from the technique known as running the numbers used at the Harvard Business School. Dan Bricklin developed the idea of using a computer to “run the numbers”, and together with a programmer Bob Frankston, they created a program which they called Visicalc (standing for visible calculator). The first copy was sold in October 1979, and it was only available for the Apple II. This application was so useful that people in middle management throughout the country felt they had to have it. These people had comfortable jobs with disposable income, and so they could afford the expense of a personal computer. Apple’s sales increased massively, which generated real interest in the personal computer industry. Later, the rights to Visicalc were bought by Lotus, who produced Lotus 1-2-3 for the IBM PC, based on the Apple product. Other spreadsheets like Wings, Excel and ClarisWorks followed. For more information see <http://www.bricklin.com/>

Spreadsheets allow a user to calculate numbers, and to present the results in tabular form. They usually contain features to improve the presentation of the data. A spreadsheet application is used by people who need to manage or manipulate data organised into rows and columns. Any time you are producing a document which uses rows and columns in some form, then you probably should be using a spreadsheet to produce it.

Essential concepts

Each position in the spreadsheet is known as a cell. Each cell can contain text, numbers, or formulae. The cell can also be formatted for a nicer appearance (borders, shading, colour etc.), but the formatting does not affect the contents of the cell.

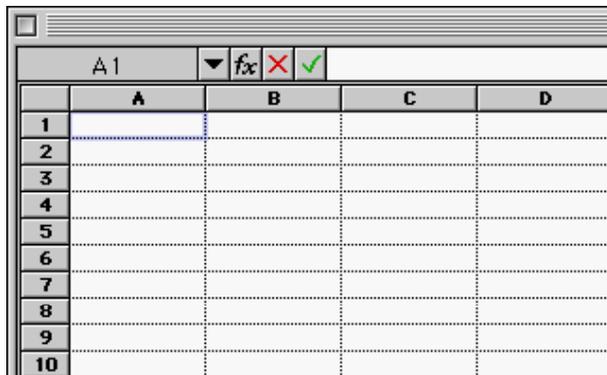


Figure 1: An empty spreadsheet

The visual image you see is the user interface. It shows the values of each cell. If you wish to enter information into any of these cells, you must select the cell by clicking on it. Any text you type will appear in an entry box at the top of the spreadsheet. This box shows the contents of the selected cell. When you hit return, the contents of the cell is recorded, any calculations made, and the results are shown in the display.

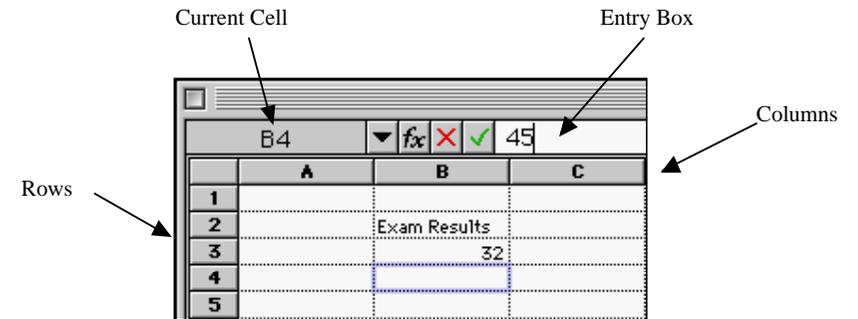


Figure 2: The number '45' is being entered into cell B4

Each cell has a row and a column which are used to refer to it. Columns are indicated by using letters, and Rows by numbers so they are never confused (just like playing Battleships). These references are needed in order to use formulae, which is where the real power of a spreadsheet lies.

Using Formula

All formulae must start with the '=' sign. The simplest type of formulae is an arithmetic expression, such as 3+4. This formula would be entered into a cell as =3+4, and the result (7) would be displayed on the page in the cell in which the formula was entered. Note that the cell *contains* the formula, but it displays the *result* of the formula.

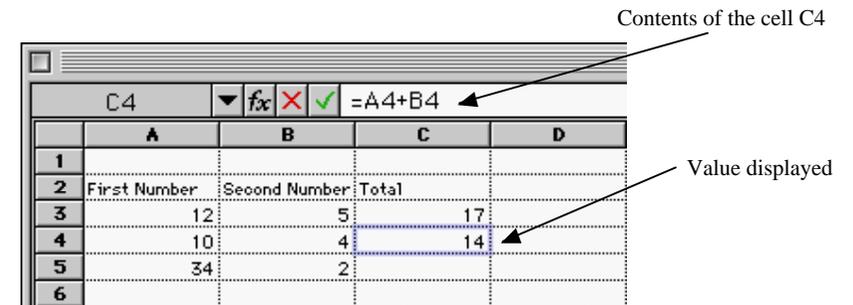


Figure 3: The values displayed by the spreadsheet are the results

Spreadsheets are designed as a tool to help us calculate numbers more efficiently, and to view the data calculated in a more pleasant manner. For this reason, we must try to construct spreadsheets so that the work required to recalculate results based upon a new set of data is kept to a minimum. In order to achieve this, the formula must refer to other cells. For example, it is useful to be able to calculate the sum of two numbers. The following example shows how this is done:

	A	B	C	D
1	415.111			
2	Test	10		
3	Exam	65		
4	Labs	25		
5				
6	Name	Test	Exam	Total Theory
7	Fred	7	46	53
8	Jane	8	60	68
9	Sally	9	57	66
10	John	3	98	
11				

Formula for calculating total

Figure 4: The sum of two numbers is calculated using a formula which refers to each cell.

In the above example, it would be more useful to be able to enter the marks for the test and exam as percentages. We then need a more complex formula to work out the total.

	A	B	C
1	415.111		
2	Test	10	
3	Exam	65	
4	Labs	25	
5			
6	Name	Test%	Test marks
7	Fred	70	7.0
8	Jane	82	8.2
9	Sally	94	9.4
10	John	36	3.6
11			

Formulae

- =B7/100*B2
- =B8/100*B2
- =B9/100*B2
- =B10/100*B2

Figure 5: Calculating the contribution of the test marks to the theoretical component.

This time the test counted towards 10% of the final grade. Perhaps in the future, this weighting may change, and next year, the test may count towards 15% of the final (and the exam only 60%). It would be better to make the spreadsheet more general, so that the user can specify how much each component contributed to the theoretical component

	A	B	C
1	415.111		
2	Test	15	
3	Exam	60	
4	Labs	25	
5			
6	Name	Test%	Test marks
7	Fred	70	10.5
8	Jane	82	12.3
9	Sally	94	14.1
10	John	36	5.4
11			

Formulae

- =B7/100*B2
- =B8/100*B2
- =B9/100*B2
- =B10/100*B2

Figure 6: We refer to the cell which contains the relevant information.

It is easier to keep the spreadsheet consistent if we store the information in a single cell, and then refer to that cell whenever we want to use that information, otherwise it is much easier to get confused. In the example shown above, we use a formula to work out the number of marks contributed to the theoretical component by referring to the percentage of final marks which comes from the test. This has been listed in cell B2, so we should refer to that cell, rather than putting the actual number in the formula.

Fill Down and Fill Right

Often, when we are designing such spreadsheets, we need to use very similar formula in each cell in a column, or in each cell in a row. It would be very tiresome to type out the formula for each new row, especially when the number of rows becomes large (e.g. recording the marks of 600 students). Since spreadsheets are designed as labour saving devices, they include a method of filling cells with information. The spreadsheet allows a user to select a group of cells, and then to Fill Right (from the Edit Menu). The leftmost cell of those selected is copied to each of the other cells. A similar option (Fill Down) can be used to fill a vertical series of cells.

Notice that sometimes we want exactly the same formula to appear in each cell, and other times we want the formula to change, depending upon its position in the spreadsheet.

	A	B	C
1	415.111		
2	Test	15	
3	Exam	60	
4	Labs	25	
5			
6	Name	Test%	Test marks
7	Fred	70	10.5
8	Jane	82	12.3
9	Sally	94	14.1
10	John	36	5.4
11			

Formulae

- =B7/100*B2
- =B8/100*B2
- =B9/100*B2
- =B10/100*B2

Figure 7: Using a formula which refers to two different cells.

In the diagram, the formula to work out test marks is based upon a person's percentage for the test (e.g. B7, B8, B9 or B10), and the value which specifies how much the test contributes to the final (e.g. B2). We would like to be able to create only one formula, and then use the option to Fill Down, and get the correct formula appearing in each cell. This can be achieved by using both relative and absolute referencing.

Absolute and Relative References

An absolute reference should be used when you are referring to the same cell in every occurrence of the formula. You can specify a reference is absolute by using a dollar sign '\$' before the row, or column which is to remain fixed.

If you do not specify the cell is an absolute reference, the spreadsheet will use relative referencing. This means that the cell reference will change when the formula is moved to a new location. The spreadsheet stores relative references by remembering how many cells away (in terms of rows and columns) the specified cell is. For example the cell you are referring to may be 2 columns to the left, and 1 row above the current cell. Wherever the formula is moved to, it would refer to the cell which is 2 columns left of the current cell and 1 row above it.

	A	B	C
1	415.111		
2	Test	15	
3	Exam	60	
4	Labs	25	
5			
6	Name	Test%	Test marks
7	Fred	70	10.5
8	Jane	82	12.3
9	Sally	94	14.1
10	John	36	5.4
11			

Formulae

- =B7/100*\$B\$2
- =B8/100*\$B\$2
- =B9/100*\$B\$2
- =B10/100*\$B\$2

Figure 8: Formulae are able to use both relative and absolute references.

In the example above, only the first formula needs to be typed in. The formula specifies that the cell which is one cell to the left contains a value which is to be divided by 100, and then the result is multiplied by the value stored in the cell B2. The dollar signs are used to distinguish between these two forms of referencing. The cell which does not change in the formula (i.e. B2) is an absolute reference. The formula is then able to be selected and used to Fill Down. The cells referred to will be automatically altered if the reference is a relative one.

Using Functions

Most spreadsheets provide a large range of functions that help the user to create effective formulae. Some of these include *Average, Max, Min, Count, and Sum*. These functions all require a *Cell Range*. A range of cells is simply all the cells within a defined area. To define an area in a spreadsheet you need to know the top left cell, and the bottom right cell in the area. In order to specify a cell range, you should give the top left cell, then a colon ':' and then the bottom right cell (e.g. A1:B3 would define a range of 6 cells).

Average(Cell Range)

This function is used to calculate the average value of each cell in the range.

Max(Cell Range)

This function is used to calculate the maximum value of all the cells in the range.

Min(Cell Range)

This function is used to calculate the minimum value of all the cells in the range.

Count(Cell Range)

This function counts how many of the cells in the range have some value stored in them (i.e. are not blank cells).

Sum(Cell Range)

This function adds up all the values within the range given.

Example: Assume each person in the class completes 6 labs. We will use the best 5 of these labs to calculate the practical component of the course. How can we create a formula which will successfully produce the answer we want?

	A	B	C	D	E	F	G	H
1	Name	Lab 01	Lab 02	Lab 03	Lab 04	Lab 05	Lab 06	Total of Best 5
2	Fred	5	4	6	3	2	10	
3	Jane	2	9	1	8	6	2	
4	Sally	7	10	8	6	8	4	
5	John	5	4	5	4	9	5	

Figure 9: How could you calculate the best 5 labs from a set of results.

There are many ways to set up a spreadsheet, so it is possible to achieve the same result using a variety of different techniques. Here we will develop the formula slowly, so each step is obvious.

	A	B	C	D	E	F	G	H	I
1	Name	Lab 01	Lab 02	Lab 03	Lab 04	Lab 05	Lab 06	Total	Total of Best 5
2	Fred	5	4	6	3	2	10	30	
3	Jane	2	9	1	8	6	2		
4	Sally	7	10	8	6	8	4		
5	John	5	4	5	4	9	5		

First, we add another column to the spreadsheet. This column will store the overall total. We simply use the Sum function to add the numbers up (an alternative would be =B2 + C2 + D2+ E2 + F2 + G2). We can use this formula to Fill Down, and obtain results for the other students.

	A	B	C	D	E	F	G	H	I	J
1	Name	Lab 01	Lab 02	Lab 03	Lab 04	Lab 05	Lab 06	Total	Minimum	Total of Best 5
2	Fred	5	4	6	3	2	10	30	2	
3	Jane	2	9	1	8	6	2	28		
4	Sally	7	10	8	6	8	4	43		
5	John	5	4	5	4	9	5	32		

Now we create another new column, and get the minimum value from the labs. This formula can be used to Fill Down also.

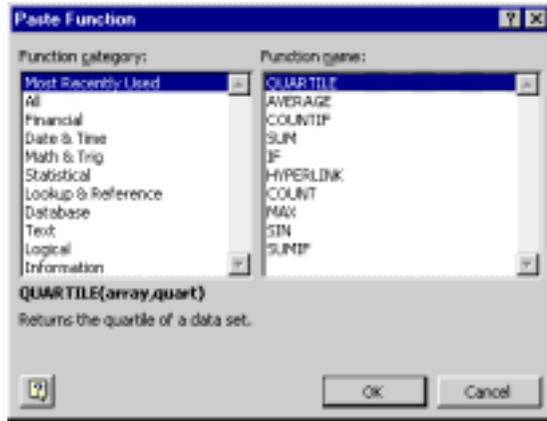
	A	B	C	D	E	F	G	H	I	J
1	Name	Lab 01	Lab 02	Lab 03	Lab 04	Lab 05	Lab 06	Total	Minimum	Total of Best 5
2	Fred	5	4	6	3	2	10	30	2	28
3	Jane	2	9	1	8	6	2	28	1	
4	Sally	7	10	8	6	8	4	43	4	
5	John	5	4	5	4	9	5	32	4	

Finally, we can subtract the lowest lab mark from the total, leaving use with the best 5 marks in the labs. In fact, the whole formula could have been done in a single line. It would be

=SUM(B2:G2)-MIN(B2:G2)

Functions add the ability to complete more complex calculations to the spreadsheet software. Some functions are designed simply to save the user time by providing the appropriate formulae in a predefined template. For example; you could work out the average by adding each cell and dividing by the number of cells. To save time, you could use two functions, `SUM(Cell_Range) / COUNT(Cell_Range)`, or even a single function `AVERAGE(Cell_Range)`. Other functions add the ability to produce results which would otherwise be impossible (a good example of this would be the IF function).

All functions in Excel can be found by choosing to **Insert a Function**. This will provide you with a dialog box containing all the functions, and a brief description of each one. The functions are divided into different categories to make them easier to find. You are encouraged to make use of this feature when writing calculations. If you think there might be a function which does the job, have a look through the most likely category.



Statistical Functions

Statistical calculations are commonly required by the users of spreadsheet software, so it is no surprise that a large range of formulae have been provided as functions. The statistical category contains many of the common functions you will need (including minimum, maximum, quartiles, averages, and standard deviations).

Boolean functions

These functions are used to return a boolean value (i.e. a value which is either *true* or *false*). They are often combined with other functions to perform more complex tasks.

And(logical_1, logical_2)

This function returns true if both *logical_1*, and *logical_2* are true.

	A	B	C	D	E
1					
2	Name	Location	Hours	Working Hard?	
3	Joe	Work	12	FALSE	
4	Julie	Holiday	68	FALSE	
5	Jenny	Work	45	TRUE	
6	Jack	Home	42	FALSE	

Or(logical_1, logical_2)

This function returns true if either *logical_1* is true or *logical_2* is true.

	A	B	C	D	E
1					
2	Name	Location	Hours	Happy	
3	Joe	Work	12	TRUE	
4	Julie	Holiday	68	TRUE	
5	Jenny	Work	45	FALSE	
6	Jack	Home	42	TRUE	

Not(logical_1)

This function returns true if *logical_1* is false, and it returns false if *logical_1* is true.

	A	B	C	D
1				
2	Name	Location	Hours	Not Working?
3	Joe	Work	12	FALSE
4	Julie	Holiday	68	TRUE
5	Jenny	Work	45	FALSE
6	Jack	Home	42	TRUE

IsBlank(Cell)

This function returns true if the cell specified does not contain any data.

Conditional Functions

A conditional function is used when you want to put different values into a cell depending upon a particular condition.

If(logical, true_value, false_value)

If *logical* is true, then *true_value* will be displayed in the cell, otherwise *false_value* will be displayed in the cell. Note that any test which produces a boolean value can be used in the *logical* test.

	A	B	C	D	E
1	Name	Location	Hours	Status	
2	Joe	Work	11	-	
3	Julie	Holiday	63	Happy	
4	Jenny	Work	47	-	
5	Jack	Home	49	Happy	

Note that an IF function can be nested. This means that the value displayed by the *true_value* or the *false_value* could actually be the result of another IF function.

D3				
=IF(NOT(B3="Work"),"N				
	A	B	C	D
1				
2	Name	Location	Hours	Current status
3	Joe	Work	12	Just working
4	Julie	Holiday	68	Not working
5	Jenny	Work	45	Working too hard
6	Jack	Home	42	Not working

Cell D3 contains: =IF(NOT(B3="Work"),"Not working",IF(C3>40,"Working too hard","Just working"))

Counting Functions

These functions are used to count cells. You must use different functions depending upon what you need to count.

Count(Cell Range)

This function is used to count the number of cells within the range which contain a numeric value.

Countif(Cell Range, Criteria)

This function counts the number of cells in the range which match the criteria. There are many ways of specifying a criteria, but we will use only the simple match. In other words, the function will count all the cells which contain the same text as that given in the criteria. To count all the cells which contain "Waiter", you could use:

COUNTIF(Cell Range, "Waiter").

If the text "Waiter" were located in another cell, say A15 for example, then you could use:

COUNTIF(Cell Range, A15).

B12		
=COUNT(\$B\$3:\$B\$10)		
	A	B
1		
2	Name	Hours Worked
3	Joe	56
4	Julie	13
5	Jenny	80
6	Jack	
7	Jerry	4
8	John	16
9	Jim	
10	James	27
11		
12	People Working	6

Cell B12 contains: =COUNT(\$B\$3:\$B\$10)

B15			
=COUNTIF(\$B\$3:\$B\$10, \$A15)			
	A	B	C
1			
2	Name	Job	Hours
3	Joe	Waiter	4
4	Julie	Chef	19
5	Jenny	Cleaner	15
6	Jack	Waiter	4
7	Jerry	Cleaner	27
8	John	Chef	15
9	Jim	Cleaner	5
10	James	Cleaner	8
11			
12	Total Hours		97
13			
14	Job	Staff Number	
15	Waiter	2	
16	Chef	2	
17	Cleaner	4	

Cell B15 contains: =COUNTIF(\$B\$3:\$B\$10, \$A15)

Lookup functions

These functions are designed to look up a value in a table and return a result from a particular row or column within the table. These lookup functions are extremely useful, and are supported by all modern spreadsheets.

Vlookup(lookup_value, Cell Range, index)

This function is designed to look up a value in a vertical table. It tries to find the *lookup_value* in the table defined by the *Cell Range*. When the *lookup_value* is located, the function returns a value from the table which is a number of columns equal to the *index* value to the right of the position where the *lookup_value* is found. Note: Excel uses the index = 1 to refer to the cell where the value was found, whereas some other spreadsheets (such as ClarisWorks) count that cell as index = 0.

C2						
=VLOOKUP(A2, \$A\$9:\$C\$13, 3, FALSE)						
	A	B	C	D	E	F
1	Employee	Hours Worked	Hourly Wage	Job	Total Earned	
2	Andrew	14	\$12.00	Labourer	\$26.00	
3	Aldus	3	\$18.00	Warden	\$21.00	
4						
5						
6						
7						
8	Name	Job	Pay Rate			
9	Andrew	Labourer	\$12.00			
10	Anne	Cleaner	\$34.00			
11	Abel	Management	\$31.00			
12	Aardvark	Pest Control	\$8.00			
13	Aldus	Warden	\$18.00			

Figure 1: Diagram showing how Vlookup can be used to find a value in a table.

Table ordering:

You can add an optional parameter into the lookup function to specify how the table is sorted. By default, spreadsheets assume that the table is sorted in alphabetical order. If this is not the case, then you should specify how the lookup function operates by specifying a value for the optional parameter. In Excel, this is as follows:

D2						
=VLOOKUP(A2, \$A\$9:\$C\$13, 2, FALSE)						
	A	B	C	D	E	F
1	Employee	Hours Worked	Hourly Wage	Job	Total Earned	
2	Andrew	14	\$12.00	Labourer	\$26.00	
3	Aldus	3	\$18.00	Warden	\$21.00	
4						
5						
6						
7						
8	Name	Job	Pay Rate			
9	Andrew	Labourer	\$12.00			
10	Anne	Cleaner	\$34.00			
11	Abel	Management	\$31.00			
12	Aardvark	Pest Control	\$8.00			
13	Aldus	Warden	\$18.00			

The table in the example below is ordered, so the value which is found by the lookup is an approximate value.

	A	B	C	D	E
1	Name	Salary	Tax Rate	Tax Paid	
2	Bert	\$4384.00	0	\$0.00	
3	Brenda	\$28763.00	0.15	\$4314.45	
4	Barry	\$17652.00	0.09	\$1588.68	
5	Bernie	\$35298.00	0.24	\$8471.52	
6	Brooklyn	\$5672.00	0	\$0.00	
7	Bozo	\$309847.00	0.33	\$102249.51	
11	Tax rate Table				
12	Income	Tax Rate			
13	0	0			
14	15000	0.09			
15	24000	0.15			
16	35000	0.24			
17	45000	0.33			
18	1000000	0.5			

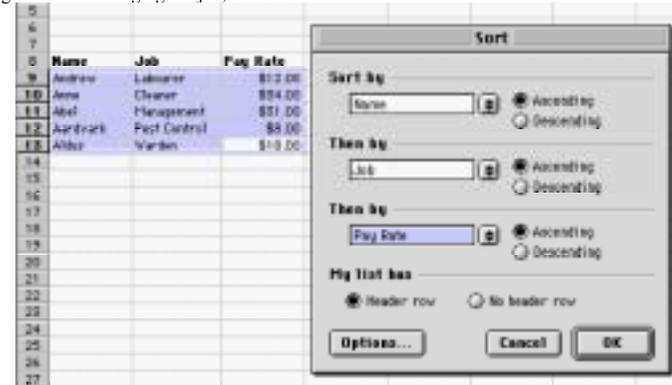
Hlookup(search_value, Cell Range, index)

This function is very similar to Vlookup, but is used for looking up values in a horizontal table. The index number indicates that we want to return a value from the table in a different row from the place we found the value we based the search on. If the index is one, then the value returned is the same as the value we searched for. If the index is 2, then we return the cell located in the row below the value we searched for. This is just the same as the vertical lookup (VLOOKUP), but should be used when the data is stored in horizontal rows instead of vertical columns.

	A	B	C	D	E	F
1		Order				
2	Code	Number	Cost			
3	CP30	1	\$5632.00			
4	R2	3	\$17676.00			
5	T2	5	\$5000000.00			
7	Table of Goods					
8	CODE	K9	CP30	BISHOP	T2	R2
9	PRICE	\$32.56	\$5632.00	\$39.00	\$1000000.00	\$5892.00
10	COLOUR	Grey	Gold	Pink	Silver	White

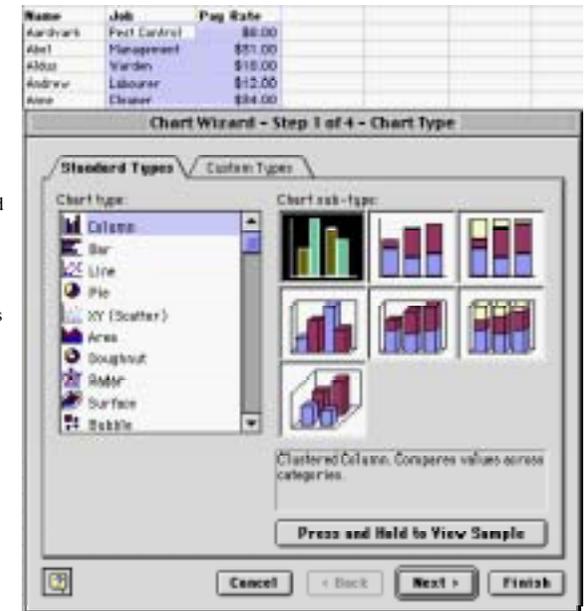
Sorting

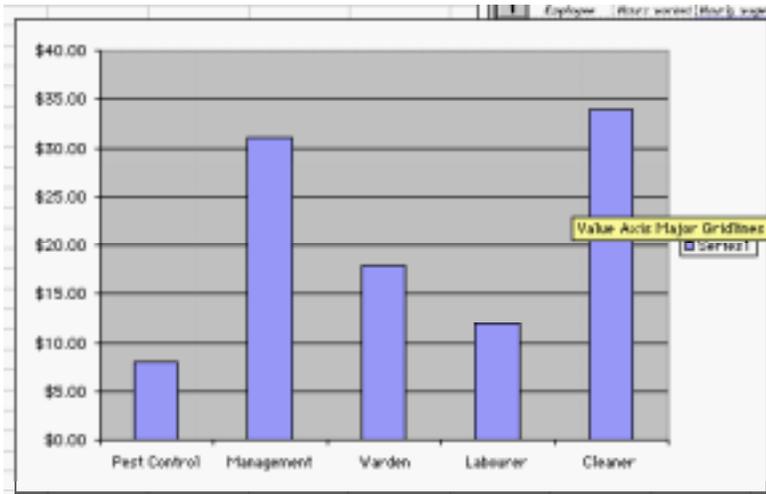
Sorting data is a common operation when using spreadsheets. In order to sort data, select all the data which is to be sorted and choose **Sort** from the **Data** menu. You will be asked which data should be used as the key for the sort. The key is the element which the ordering is based upon. Notice that all the data which is related has been selected when sorting. If you only selected the names, then they would be rearranged without changing the job, or hours.



Making Charts

To make a chart based upon the data in the spreadsheet, first you must select the data (including the headings for each row or column). You should choose **Chart** from the **Insert** menu, and you will be given a choice of which sort of chart you wish. During the creation process, you are able to alter the labels for each axis and the overall title of the chart. A chart will then be created based upon the data selected. If the data is changed, the chart will be altered to reflect the current data. The size of the chart can be altered by selecting it, and moving one of the handles in the corners (just like an object in a drawing application).





The final chart

Printing

When you want to print a spreadsheet, you must first select the cells that you wish to print. You are then able to go to the **File** Menu, select the **Print Area** options and **Set the Print Area**. When you print, the area printed will only be the highlighted cells. Remember that you can alter the **Page Setup** to ensure that the data is scaled so that it fits onto a single page (or a single width of a page). You are also able to put a header and footer onto a document using the **Page Setup**.

If you want to print a graph, you are able to select the graph, and simply choose print. The graph will then be printed by itself.

Author: Andrew Luxton