University of Leicester

Data Handling in Excel

University of Leicester

Overview

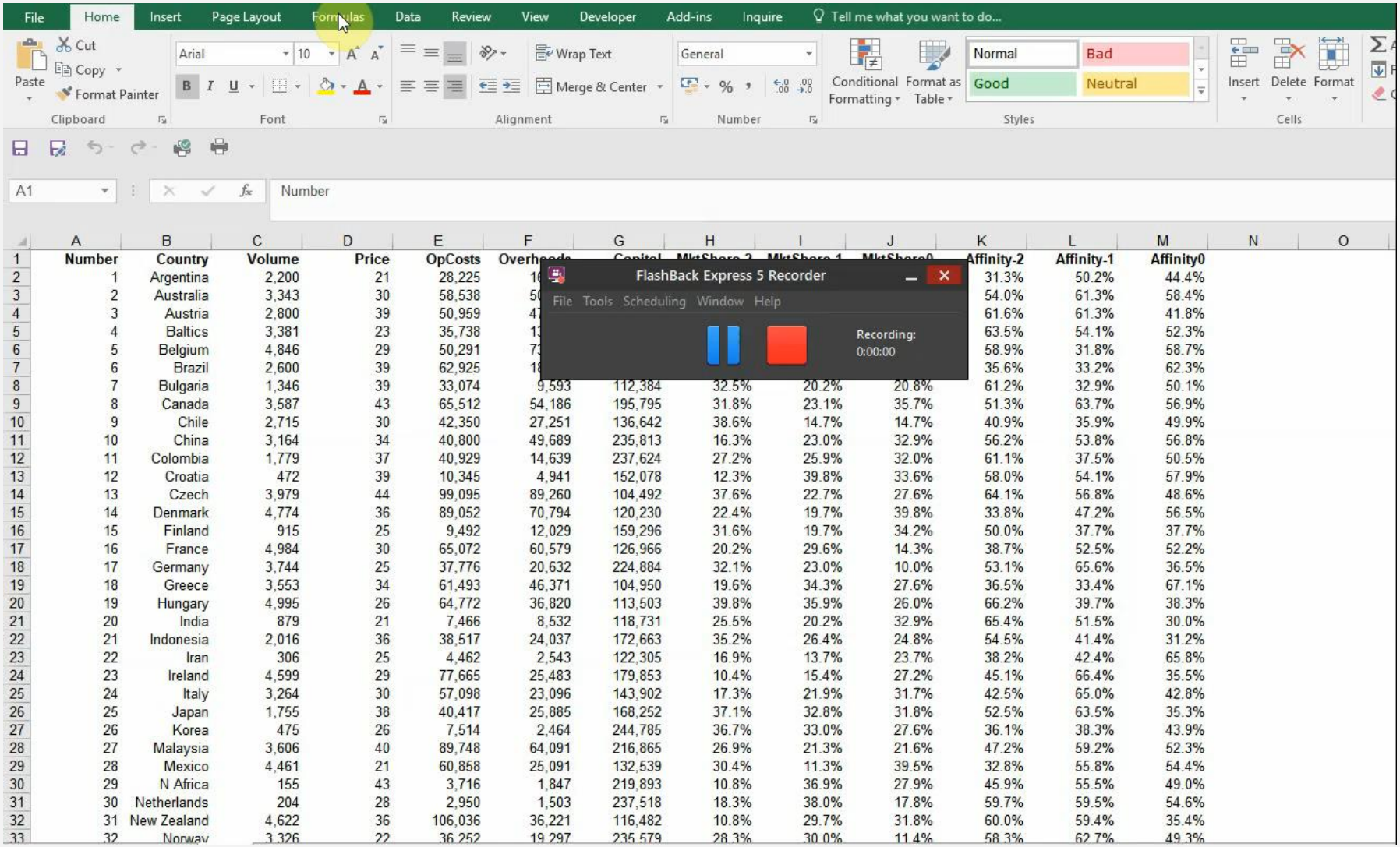
- Objectives
  - Handle large data sets efficiently and reliably
    - Extract specific information
    - Summarise data in different ways
    - Analyse data
  - Produce management reports quickly and easily
  - Analyse trends
  - Etc etc

University of Leicester

Naming data ranges

- If you intend to refer to the same array of data frequently, consider naming the range (particularly if it is large).
- To do this, you need the ‘Defined Names’ section of the ‘Formulas’ tab:
  - Name Manager
  - Define Name
  - Use in Formula
  - Create from Selection
- To define a name, select ‘Define Name’ and highlight the range you want.

University of Leicester



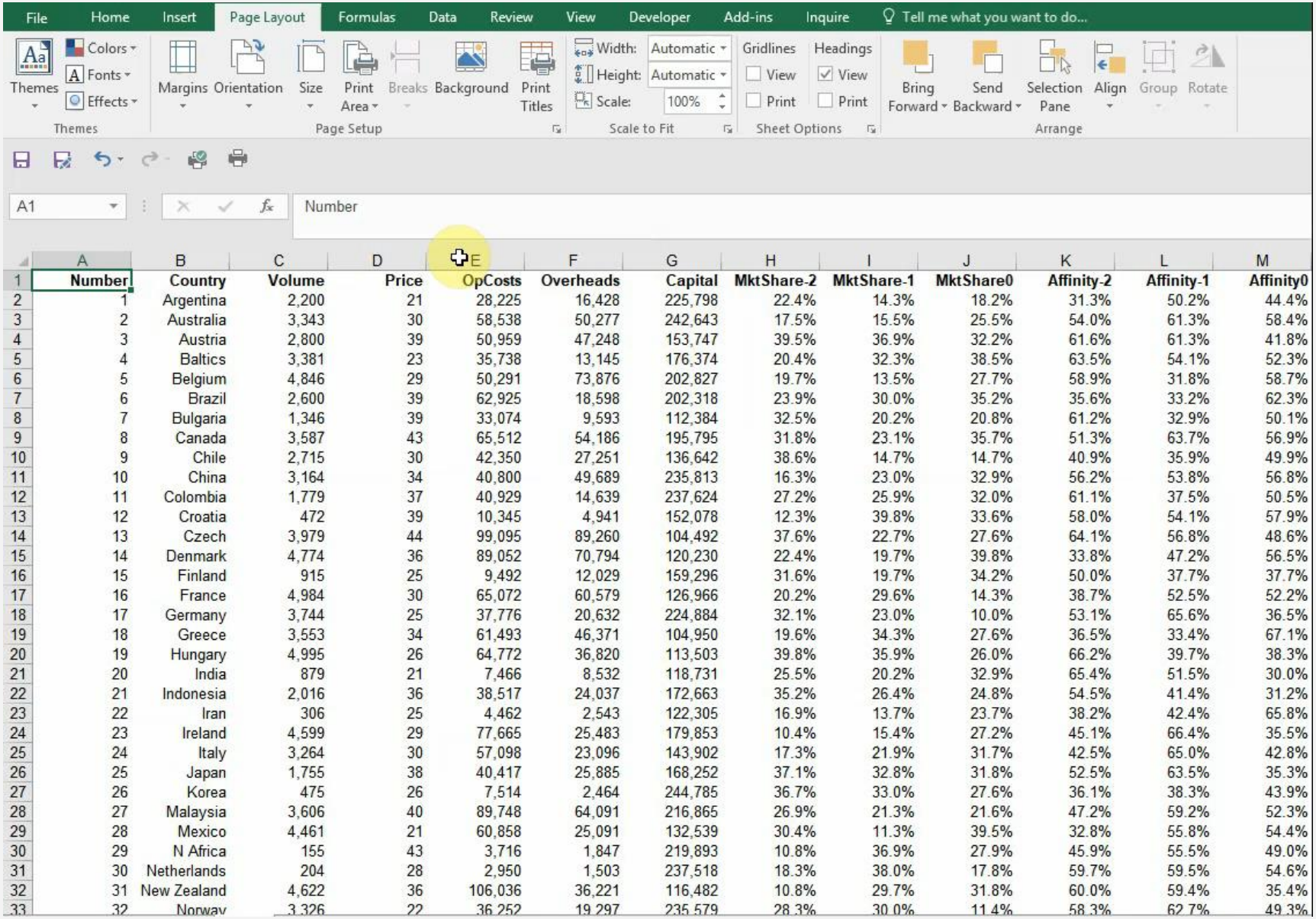
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Number	Country	Volume	Price	OpCosts	Overheads	Capital	MktShare 2	MktShare 1	MktShare0	Affinity-2	Affinity-1	Affinity0		
2	1	Argentina	2,200	21	28,225	1					31.3%	50.2%	44.4%		
3	2	Australia	3,343	30	58,538	5					54.0%	61.3%	58.4%		
4	3	Austria	2,800	39	50,959	4					61.6%	61.3%	41.8%		
5	4	Baltics	3,381	23	35,738	1					63.5%	54.1%	52.3%		
6	5	Belgium	4,846	29	50,291	7					58.9%	31.8%	58.7%		
7	6	Brazil	2,600	39	62,925	1					35.6%	33.2%	62.3%		
8	7	Bulgaria	1,346	39	33,074	9,593	112,384	32.5%	20.2%	20.8%	61.2%	32.9%	50.1%		
9	8	Canada	3,587	43	65,512	54,186	195,795	31.8%	23.1%	35.7%	51.3%	63.7%	56.9%		
10	9	Chile	2,715	30	42,350	27,251	136,642	38.6%	14.7%	14.7%	40.9%	35.9%	49.9%		
11	10	China	3,164	34	40,800	49,689	235,813	16.3%	23.0%	32.9%	56.2%	53.8%	56.8%		
12	11	Colombia	1,779	37	40,929	14,639	237,624	27.2%	25.9%	32.0%	61.1%	37.5%	50.5%		
13	12	Croatia	472	39	10,345	4,941	152,078	12.3%	39.8%	33.6%	58.0%	54.1%	57.9%		
14	13	Czech	3,979	44	99,095	89,260	104,492	37.6%	22.7%	27.6%	64.1%	56.8%	48.6%		
15	14	Denmark	4,774	36	89,052	70,794	120,230	22.4%	19.7%	39.8%	33.8%	47.2%	56.5%		
16	15	Finland	915	25	9,492	12,029	159,296	31.6%	19.7%	34.2%	50.0%	37.7%	37.7%		
17	16	France	4,984	30	65,072	60,579	126,966	20.2%	29.6%	14.3%	38.7%	52.5%	52.2%		
18	17	Germany	3,744	25	37,776	20,632	224,884	32.1%	23.0%	10.0%	53.1%	65.6%	36.5%		
19	18	Greece	3,553	34	61,493	46,371	104,950	19.6%	34.3%	27.6%	36.5%	33.4%	67.1%		
20	19	Hungary	4,995	26	64,772	36,820	113,503	39.8%	35.9%	26.0%	66.2%	39.7%	38.3%		
21	20	India	879	21	7,466	8,532	118,731	25.5%	20.2%	32.9%	65.4%	51.5%	30.0%		
22	21	Indonesia	2,016	36	38,517	24,037	172,663	35.2%	26.4%	24.8%	54.5%	41.4%	31.2%		
23	22	Iran	306	25	4,462	2,543	122,305	16.9%	13.7%	23.7%	38.2%	42.4%	65.8%		
24	23	Ireland	4,599	29	77,665	25,483	179,853	10.4%	15.4%	27.2%	45.1%	66.4%	35.5%		
25	24	Italy	3,264	30	57,098	23,096	143,902	17.3%	21.9%	31.7%	42.5%	65.0%	42.8%		
26	25	Japan	1,755	38	40,417	25,885	168,252	37.1%	32.8%	31.8%	52.5%	63.5%	35.3%		
27	26	Korea	475	26	7,514	2,464	244,785	36.7%	33.0%	27.6%	36.1%	38.3%	43.9%		
28	27	Malaysia	3,606	40	89,748	64,091	216,865	26.9%	21.3%	21.6%	47.2%	59.2%	52.3%		
29	28	Mexico	4,461	21	60,858	25,091	132,539	30.4%	11.3%	39.5%	32.8%	55.8%	54.4%		
30	29	N Africa	155	43	3,716	1,847	219,893	10.8%	36.9%	27.9%	45.9%	55.5%	49.0%		
31	30	Netherlands	204	28	2,950	1,503	237,518	18.3%	38.0%	17.8%	59.7%	59.5%	54.6%		
32	31	New Zealand	4,622	36	106,036	36,221	116,482	10.8%	29.7%	31.8%	60.0%	59.4%	35.4%		
33	32	Norway	3,326	22	36,252	19,297	235,579	28.3%	30.0%	11.4%	58.3%	62.7%	49.3%		

University of Leicester

Creating names from a selection

- You can also create many names quickly by using ‘Create from a selection’.
- You can use this to turn any column headings or row labels into range names.
- Select the data, click on ‘Create Names from a Selection’ and select row and/or column.
- Note that the first character of a name must be a letter, an underscore (_), or a backslash (\). Subsequent characters can be letters, numbers, periods, and underscore characters.

University of Leicester

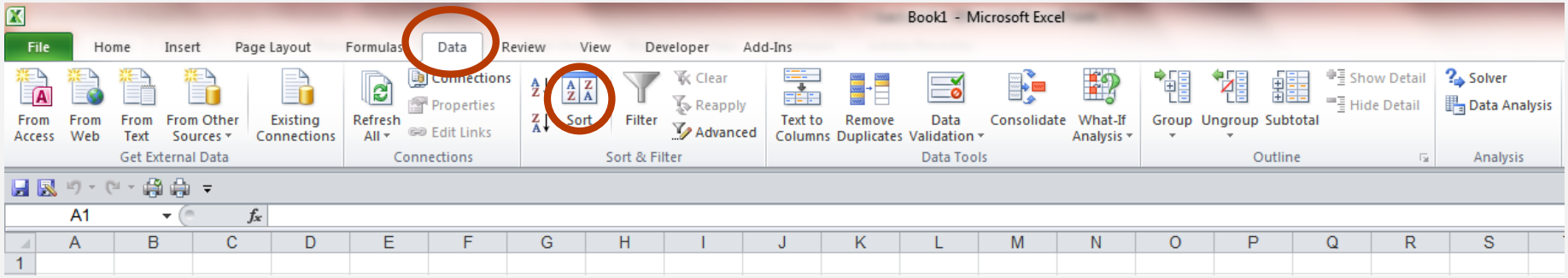


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Number	Country	Volume	Price	OpCosts	Overheads	Capital	MktShare-2	MktShare-1	MktShare0	Affinity-2	Affinity-1	Affinity0
2	1	Argentina	2,200	21	28,225	16,428	225,798	22.4%	14.3%	18.2%	31.3%	50.2%	44.4%
3	2	Australia	3,343	30	58,538	50,277	242,643	17.5%	15.5%	25.5%	54.0%	61.3%	58.4%
4	3	Austria	2,800	39	50,959	47,248	153,747	39.5%	36.9%	32.2%	61.6%	61.3%	41.8%
5	4	Baltics	3,381	23	35,738	13,145	176,374	20.4%	32.3%	38.5%	63.5%	54.1%	52.3%
6	5	Belgium	4,846	29	50,291	73,876	202,827	19.7%	13.5%	27.7%	58.9%	31.8%	58.7%
7	6	Brazil	2,600	39	62,925	18,598	202,318	23.9%	30.0%	35.2%	35.6%	33.2%	62.3%
8	7	Bulgaria	1,346	39	33,074	9,593	112,384	32.5%	20.2%	20.8%	61.2%	32.9%	50.1%
9	8	Canada	3,587	43	65,512	54,186	195,795	31.8%	23.1%	35.7%	51.3%	63.7%	56.9%
10	9	Chile	2,715	30	42,350	27,251	136,642	38.6%	14.7%	14.7%	40.9%	35.9%	49.9%
11	10	China	3,164	34	40,800	49,689	235,813	16.3%	23.0%	32.9%	56.2%	53.8%	56.8%
12	11	Colombia	1,779	37	40,929	14,639	237,624	27.2%	25.9%	32.0%	61.1%	37.5%	50.5%
13	12	Croatia	472	39	10,345	4,941	152,078	12.3%	39.8%	33.6%	58.0%	54.1%	57.9%
14	13	Czech	3,979	44	99,095	89,260	104,492	37.6%	22.7%	27.6%	64.1%	56.8%	48.6%
15	14	Denmark	4,774	36	89,052	70,794	120,230	22.4%	19.7%	39.8%	33.8%	47.2%	56.5%
16	15	Finland	915	25	9,492	12,029	159,296	31.6%	19.7%	34.2%	50.0%	37.7%	37.7%
17	16	France	4,984	30	65,072	60,579	126,966	20.2%	29.6%	14.3%	38.7%	52.5%	52.2%
18	17	Germany	3,744	25	37,776	20,632	224,884	32.1%	23.0%	10.0%	53.1%	65.6%	36.5%
19	18	Greece	3,553	34	61,493	46,371	104,950	19.6%	34.3%	27.6%	36.5%	33.4%	67.1%
20	19	Hungary	4,995	26	64,772	36,820	113,503	39.8%	35.9%	26.0%	66.2%	39.7%	38.3%
21	20	India	879	21	7,466	8,532	118,731	25.5%	20.2%	32.9%	65.4%	51.5%	30.0%
22	21	Indonesia	2,016	36	38,517	24,037	172,663	35.2%	26.4%	24.8%	54.5%	41.4%	31.2%
23	22	Iran	306	25	4,462	2,543	122,305	16.9%	13.7%	23.7%	38.2%	42.4%	65.8%
24	23	Ireland	4,599	29	77,665	25,483	179,853	10.4%	15.4%	27.2%	45.1%	66.4%	35.5%
25	24	Italy	3,264	30	57,098	23,096	143,902	17.3%	21.9%	31.7%	42.5%	65.0%	42.8%
26	25	Japan	1,755	38	40,417	25,885	168,252	37.1%	32.8%	31.8%	52.5%	63.5%	35.3%
27	26	Korea	475	26	7,514	2,464	244,785	36.7%	33.0%	27.6%	36.1%	38.3%	43.9%
28	27	Malaysia	3,606	40	89,748	64,091	216,865	26.9%	21.3%	21.6%	47.2%	59.2%	52.3%
29	28	Mexico	4,461	21	60,858	25,091	132,539	30.4%	11.3%	39.5%	32.8%	55.8%	54.4%
30	29	N Africa	155	43	3,716	1,847	219,893	10.8%	36.9%	27.9%	45.9%	55.5%	49.0%
31	30	Netherlands	204	28	2,950	1,503	237,518	18.3%	38.0%	17.8%	59.7%	59.5%	54.6%
32	31	New Zealand	4,622	36	106,036	36,221	116,482	10.8%	29.7%	31.8%	60.0%	59.4%	35.4%
33	32	Norway	3,326	22	36,252	19,297	235,579	28.3%	30.0%	11.4%	58.3%	62.7%	49.3%

University of Leicester

Sort

- The most basic form of data handling is ‘Sort’. This can be found on the Data ribbon and gives several depths of sorting, high to low or low to high, by numerical or alphabetical order.

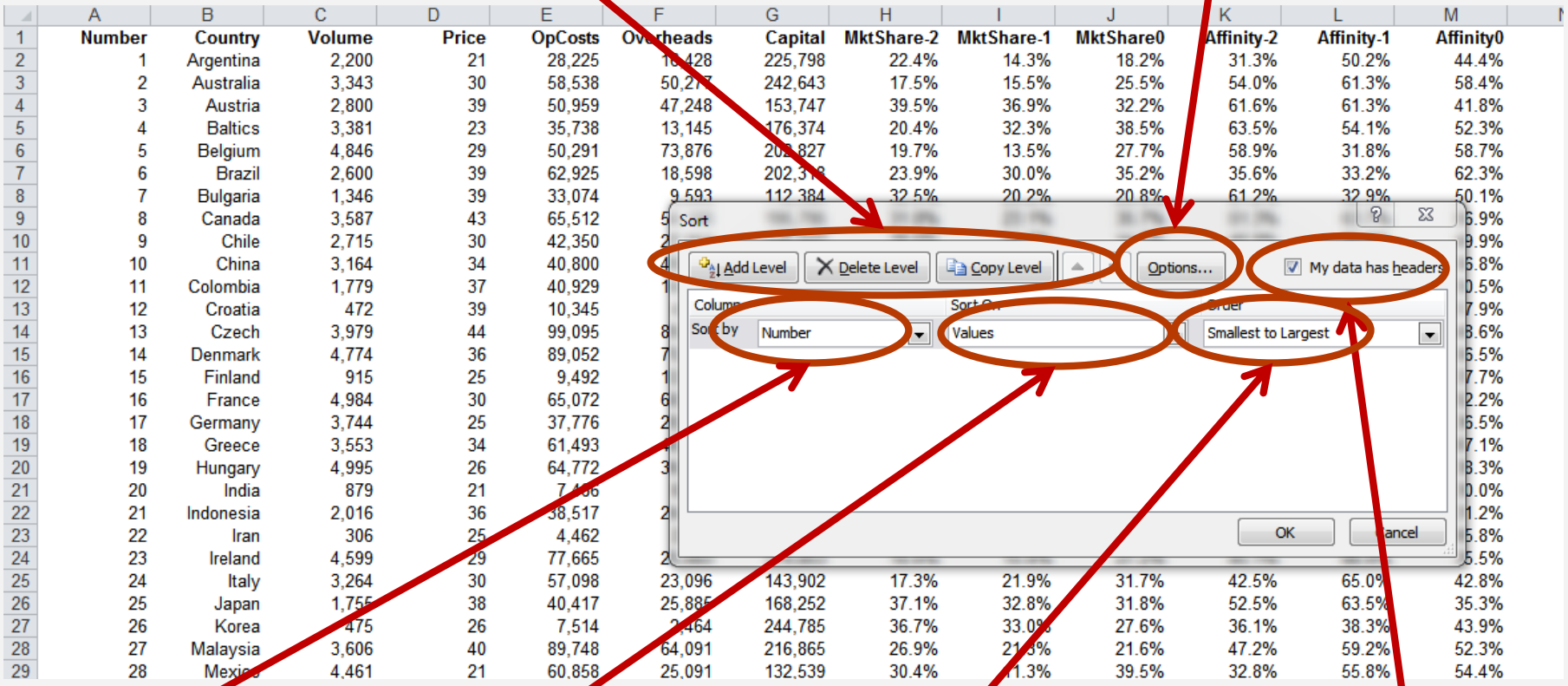


- Selecting ‘Sort’ brings up a new dialogue box.

University of Leicester

Add, delete, copy or reorder levels of sort

Use this to change to sorting by rows rather than columns



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Number	Country	Volume	Price	OpCosts	Overheads	Capital	MktShare-2	MktShare-1	MktShare0	Affinity-2	Affinity-1	Affinity0
2	1	Argentina	2,200	21	28,225	16,428	225,798	22.4%	14.3%	18.2%	31.3%	50.2%	44.4%
3	2	Australia	3,343	30	58,538	50,217	242,643	17.5%	15.5%	25.5%	54.0%	61.3%	58.4%
4	3	Austria	2,800	39	50,959	47,248	153,747	39.5%	36.9%	32.2%	61.6%	61.3%	41.8%
5	4	Baltics	3,381	23	35,738	13,145	176,374	20.4%	32.3%	38.5%	63.5%	54.1%	52.3%
6	5	Belgium	4,846	29	50,291	73,876	202,827	19.7%	13.5%	27.7%	58.9%	31.8%	58.7%
7	6	Brazil	2,600	39	62,925	18,598	202,318	23.9%	30.0%	35.2%	35.6%	33.2%	62.3%
8	7	Bulgaria	1,346	39	33,074	9,593	112,384	32.5%	20.2%	20.8%	61.2%	32.9%	50.1%
9	8	Canada	3,587	43	65,512	5							6.9%
10	9	Chile	2,715	30	42,350	2							9.9%
11	10	China	3,164	34	40,800	4							6.8%
12	11	Colombia	1,779	37	40,929	1							0.5%
13	12	Croatia	472	39	10,345								7.9%
14	13	Czech	3,979	44	99,095	8							8.6%
15	14	Denmark	4,774	36	89,052	7							6.5%
16	15	Finland	915	25	9,492	1							7.7%
17	16	France	4,984	30	65,072	6							2.2%
18	17	Germany	3,744	25	37,776	2							6.5%
19	18	Greece	3,553	34	61,493								7.1%
20	19	Hungary	4,995	26	64,772	3							8.3%
21	20	India	879	21	7,486								0.0%
22	21	Indonesia	2,016	36	38,517	2							1.2%
23	22	Iran	306	25	4,462								5.8%
24	23	Ireland	4,599	29	77,665	2							5.5%
25	24	Italy	3,264	30	57,098	23,096	143,902	17.3%	21.9%	31.7%	42.5%	65.0%	42.8%
26	25	Japan	1,755	38	40,417	25,885	168,252	37.1%	32.8%	31.8%	52.5%	63.5%	35.3%
27	26	Korea	475	26	7,514	2,464	244,785	36.7%	33.0%	27.6%	36.1%	38.3%	43.9%
28	27	Malaysia	3,606	40	89,748	64,091	216,865	26.9%	21.3%	21.6%	47.2%	59.2%	52.3%
29	28	Mexico	4,461	21	60,858	25,091	132,539	30.4%	21.3%	39.5%	32.8%	55.8%	54.4%

Select which column to sort on

Basis of sort (usually values)

Sort order

Check this box if you have column headers

University of Leicester

Excel data handling functions

- Excel includes a large number of functions for handling data.
- These are used, in particular, for extracting specific information from a large array.
- We will look briefly at some of the most important ones here. You can find more in the ‘Lookup and Reference’ section of the function library on the ‘Formulas’ tab.
- Use Excel help to find more details and examples.

University of Leicester

COUNT functions

- Excel includes several functions for counting the number of cells in the ranges specified that meet certain conditions:
  - COUNT - cells with numbers or number-like entries
  - COUNTA - cells that are not blank
  - COUNTBLANK - cells that are blank
  - COUNTIF - that satisfy the given condition
  - COUNTIFS instead of COUNTIF if there are several conditions.
- For example, =COUNTIF(C2:C53,">4000") counts countries with sales volume greater than 4,000

University of Leicester

INDEX

- What it does: (in its most straightforward use) picks out a particular value from an array
- Syntax: =INDEX(Array, Row nº, Column nº)
  - This picks out the value in the cell at the intersection of the row and column specified.
  - You can also use INDEX to extract a whole column or row by entering it in array form
- Uses: extracting partial data from a larger array

University of Leicester

INDEX

- Example: =INDEX(A4:E9,3,2) returns ‘Belgium’ – the entry in the 2nd column of the 3rd row of the data array

	A	B	C	D	E
1	**Number**	**Country**	**Volume**	**Price**	**OpCosts**
2	1	Argentina	2,200	21	28,225
3	2	Australia	3,343	30	58,538
4	3	Austria	2,800	39	50,959
5	4	Baltics	3,381	23	35,738
6	5	Belgium	4,846	29	50,291
7	6	Brazil	2,600	39	62,925
8	7	Bulgaria	1,346	39	33,074
9	8	Canada	3,587	43	65,512

University of Leicester

MATCH

- What it does: picks out the position of a given value in an array
- Syntax: =MATCH(lookup_value, lookup_array,0) (use 1 instead of 0 for approximate match).
- Uses: when you want to know the position of an entry in an array rather than the value.
  Example: =MATCH("Belgium",B1:B9,0) returns 6, the position of ‘Belgium’ in the range B1-B9.

University of Leicester

VLOOKUP and HLOOKUP

- What it does: looks up other information in the same row as a specified value
- Syntax: VLOOKUP (lookup_value, table_array, col_index_num, 0). This looks for the lookup value in the first column of the table array and returns the value in the column specified by its index number.
- Example: =VLOOKUP("Belgium",B1:E9,2) returns 4,846 – the sales volume for Belgium
- HLOOKUP works the same way, but searches horizontally rather than vertically

University of Leicester

OFFSET

- What it does: returns a cell or range of cells that is a specified number of rows and columns from a cell or range of cells
- Syntax: In its simplest form, OFFSET(reference, rows, cols) returns the contents of the cell ‘rows’ down and ‘cols’ across from the reference cell
- Example: =OFFSET(A1,5,1) returns ‘Belgium’ – the value in the cell 5 rows below and 1 cell across from A1

University of Leicester

INDIRECT

- What it does: allows you to change the reference cell in a formula without changing the formula.
- Syntax: =INDIRECT(ref_text) where ref_text is a string of characters that Excel can interpret as a cell reference. It returns the value in the cell specified by ref_text.
- Example: =INDIRECT("B"&MATCH(5,A1:A9)) returns ‘Belgium’, the value in B6.
- We can make the ‘5’ a cell reference, and by changing the value in this cell make INDIRECT point to different parts of the array.

University of Leicester

More uses of INDIRECT

- Suppose we use ‘Create from a Selection’ to name just the first piece of information after the country name – the volume. Then the name refers to a single cell. We can now use this in INDIRECT. For example, =INDIRECT(“Belgium”) returns 4,846, the sales volume for Belgium.
- We can select all the other pieces of information for Belgium by using an OFFSET function.
- And if INDIRECT refers to a cell with the country name in it we can access all the information for a country – just by typing its name.

University of Leicester

Calculations on data arrays

- As well as providing functions that handle and extract data, Excel provides functions that ease calculations when dealing with arrays.
- We will look at SUMPRODUCT, SUMIF and SUMIFS.
- We will also look at other ways to ease calculations, such as calculating on whole columns and three dimensional calculations.
- You can also use *array formulae* but we will not cover those here.

University of Leicester

SUMPRODUCT

- What it does: multiplies arrays together term by term and adds the results together.
- Syntax: =SUMPRODUCT(array1, [array2], ...). SUMPRODUCT multiplies the corresponding elements of all the arrays (which must all have the same dimensions) and adds the results
- Uses: simplifies this type of calculation, avoiding having to calculate the products in a separate array before they can be added
- There is also SUMSQ, which squares numbers and adds the results.

University of Leicester

SUMIF and SUMIFS

- We saw COUNTIF and COUNTIFS before. SUMIF and SUMIFS work the same way, adding only the contents of those cells that satisfy the given condition(s).
- SUMIF (and COUNTIF) are legacy functions – you only need the IFS versions.
- There is also an AVERAGEIFS function but not for other functions such as MIN or MAX, though you can construct your own easily enough.

University of Leicester

Calculating with whole columns

- For some of the functions we have looked you have to change the formulae if you add rows to your data but many functions ignore blank cells and so can be used with whole columns:
  - VLOOKUP, INDEX, MATCH
  - SUM, SUMPRODUCT, SUMSQ, SUMIFS
  - Count functions
  - AVERAGE, MIN, MAX, MEDIAN
  - Standard deviation and Variance functions
- To use whole columns, click on the headers for the column(s), rather than selecting a range.

University of Leicester

Multiple worksheets

- So far, we have used a workbook with just a single worksheet. Suppose now those data are for January and there are sheets in an identical format for all the other months, with sheets named ‘January’, ‘February’ etc
- We can use INDIRECT to pick out individual months in the same way as for individual countries, by making its argument a month name.
- Using INDIRECT across multiple sheets is probably its most important use.

University of Leicester

Three dimensional formulae

- We can also do calculations across the different sheets. For example, =SUM(January:December!C2) adds together the values in C2 over all 12 sheets, giving the total annual sales for Argentina.
- Many other functions will also work across sheets, such as:
  - SUM, SUMSQ
  - Count functions
  - Descriptive statistics functions

University of Leicester

Wild cards

- Many data handling functions accept wild cards in text strings, including:

  AVERAGEIF(S)
  COUNTIF(S)
  H(V)LOOKUP
  MATCH
  SUMIF(S)

- The wild cards are ?, * and ~.
  - ? can stand for any (single) letter, so B?Y will select BAY, BOY or BUY
  - * can be any string of letters, so BO* can be BOA, BODY, BODY LANGUAGE, BOY etc
  - ~ is used when you want to include ?, * or ~ in a string, so BAY~? picks up BAY?

University of Leicester

Find and replace

- 'Find and Replace' can be very useful – particularly if you have a lot of replacements to do, when you can choose ‘Replace All’.
- For example, if your data contains unwanted hyphens in a column you can replace them all with spaces or insert spaces in front of them.
- Note that you have to be careful to replace with exactly what you want (especially in terms of spaces) and about the order in which you carry out find and replace if you are doing more than one.

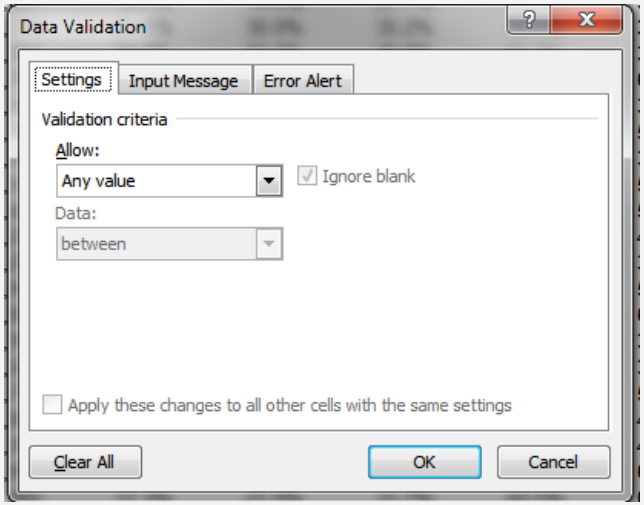
University of Leicester

Data Validation

University of Leicester

Data validation

- One of the problems with functions like MATCH is that if you ask for an exact match and mistype the name then it will return an error.
- To overcome this, you restrict the range of entries that the user can make to those that are valid – data validation.
- Data validation can be found on the data tab. Select this and choosing data validation brings up this dialogue box.



University of Leicester

Data validation options

- Selecting the drop down menu brings up a range of options. Data can be restricted to certain ranges of numbers and to integers only, or to text or dates, again within given ranges.
- A useful option is to restrict entries to those in a list, by selecting the list option. The entries can be restricted to those in a list somewhere in the spreadsheet.
- For a list, these can be displayed as a drop down menu, so the user does not have to type anything at all!

University of Leicester