

Manipulating dataset columns in Excel



Learning intentions

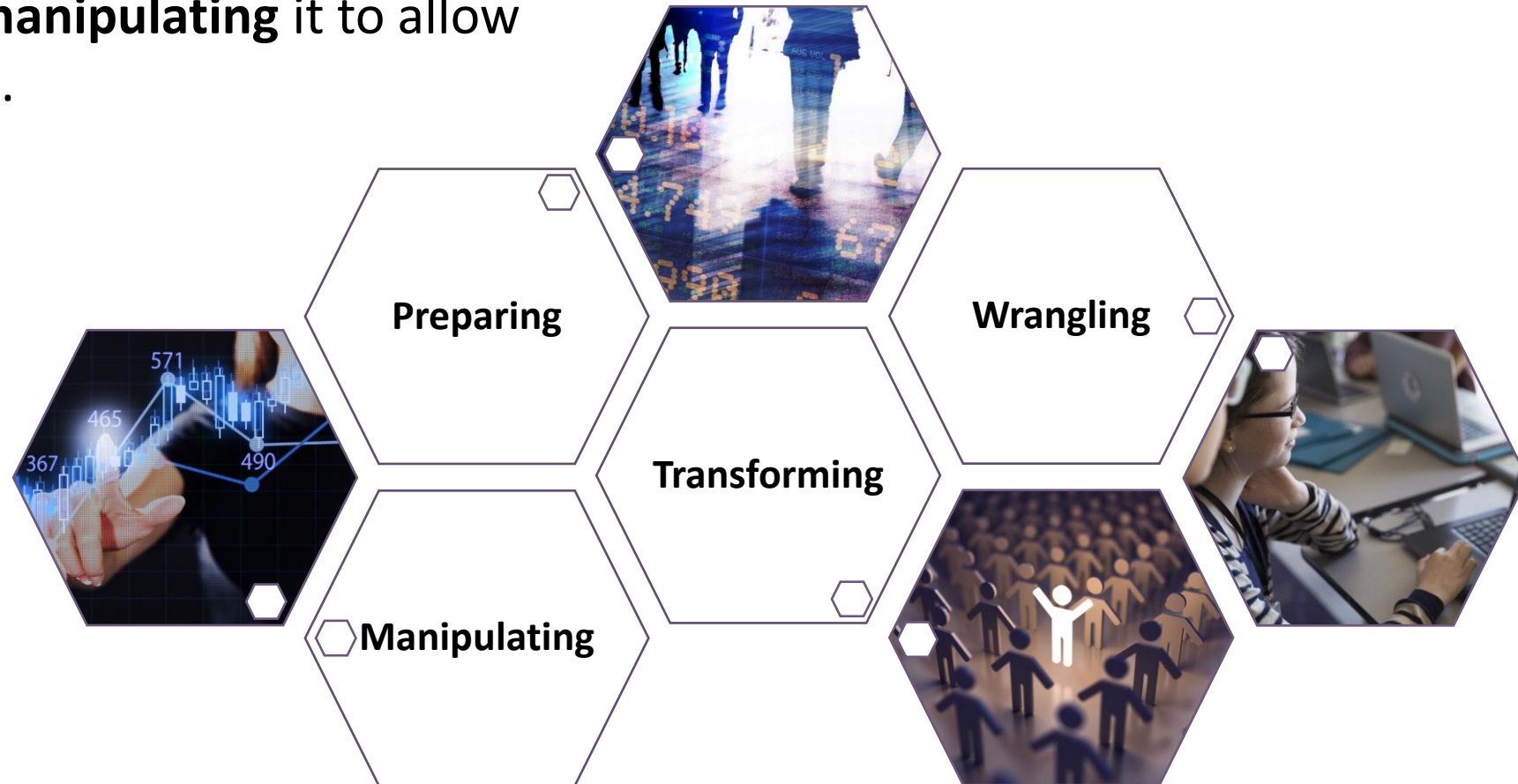
We will be learning how to manipulate data in Excel, specifically,

- to be able to **select** columns
- to be able to **reorder** columns
- to be able to **reformat** columns

Background

When a data analyst is given a dataset to analyse, **most of their time is spent manipulating** it to allow them to conduct the analysis.

Many words are used to describe this process, but they all mean the same thing.

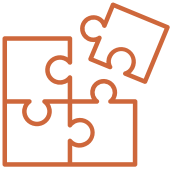


Why this is important?

Some benefits of manipulating data are,



Makes it easier for you to understand the dataset you're working with



Helps break the task into **manageable chunks**



Speeds up data processing



It **helps you to focus** on what you are planning to do with the data by choosing which columns you need to keep.

Show me...



Here is an example of a dataset that would need the columns to be manipulated.

There is a column that is **no longer needed**

It would be easier to understand if the start_time and end_time **columns were next to each other**

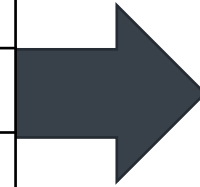
end_time	date	old_title	id	start_time
13:06	44467	0	1A	12:30
02:56	29/09/2021	0	2B	21:18
23:53	44,469	0	3C	14:09
16:04	44,470	0	4D	21:49

Date column has a **mixture of formats**

Show me...



end_time	date	old_title	id	start_time
13:06	44467	0	1A	12:30
02:56	29/09/2021	0	2B	21:18
23:53	44,469	0	3C	14:09
16:04	44,470	0	4D	21:49



id	date	start_time	end_time
1A	28/9/2021	12:30	13:06
2B	29/09/2021	21:18	02:56
3C	30/9/2021	14:09	23:53
4D	1/10/2021	21:49	16:04

This lesson will show you **how to manipulate columns of data**, which will make any analysis easier to complete and less prone to errors.

Definition



Select

To choose some of the
columns from a dataset

Show me...



Here is an example of selecting columns.

address	price_2017	price_2018	price_2019	price_2020
5 Little Gardens	50,000	55,000	60,500	66,550
17 Cherry Tree Lane	65,000	71,500	78,650	86,515
42 Granny Clarks Wynd	100,000	110,000	121,000	133,100

Select 'address'
and 'price_2020'
columns

address	price_2020
5 Little Gardens	66,550
17 Cherry Tree Lane	86,515
42 Granny Clarks Wynd	133,100

Example

Select

To choose some of the columns from a dataset

Select the name of the river and its length (km) from this dataset.

river	start_point	length_km
Tay	Ben Lui	188
Spey	Loch Spey	172
Clyde	Lowther Hill	170
Forth	Loch Ard	47
Nith	Enoch Hill	114
Tweed	Tweeds Well	156

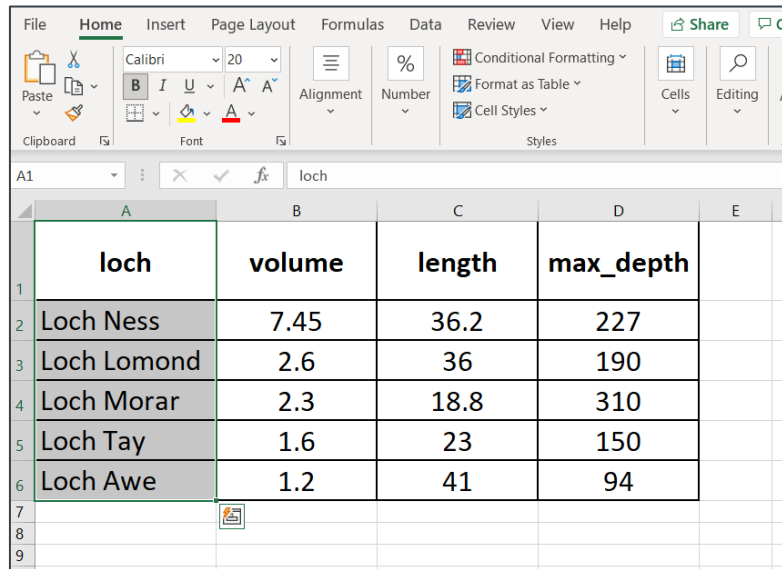
Select 'river' and
'length_km'

river	length_km
Tay	188
Spey	172
Clyde	170
Forth	47
Nith	114
Tweed	156

How to Select in Excel

Step 1.

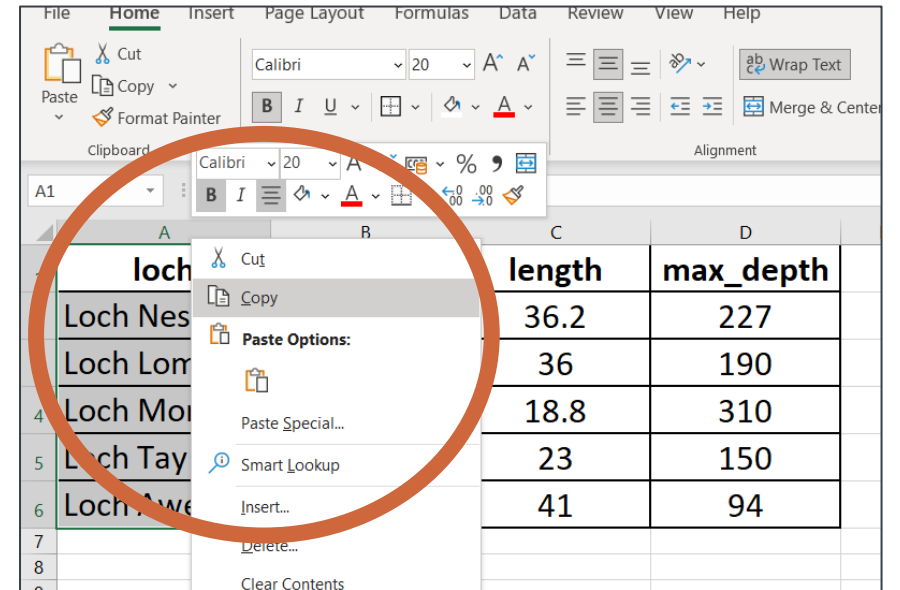
Highlight the column you want to **select** then 'Copy' the column.



An Excel spreadsheet with the following data:

	A	B	C	D	E
	loch	volume	length	max_depth	
1					
2	Loch Ness	7.45	36.2	227	
3	Loch Lomond	2.6	36	190	
4	Loch Morar	2.3	18.8	310	
5	Loch Tay	1.6	23	150	
6	Loch Awe	1.2	41	94	
7					
8					
9					

Select



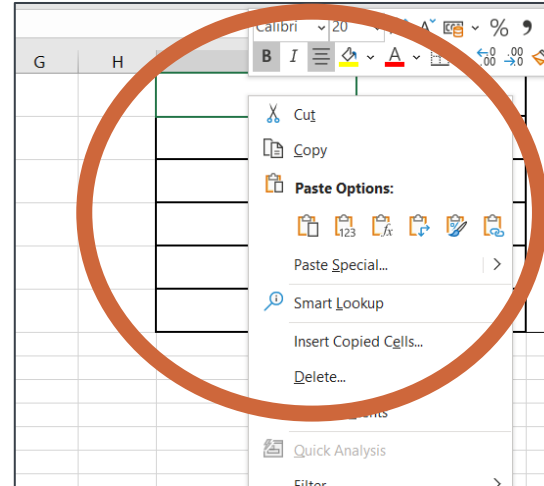
An Excel spreadsheet showing the 'Copy' option highlighted in the context menu for the selected 'loch' column. The data is the same as the previous table.

	A	B	C	D	E
	loch	volume	length	max_depth	
1					
2	Loch Ness	7.45	36.2	227	
3	Loch Lomond	2.6	36	190	
4	Loch Morar	2.3	18.8	310	
5	Loch Tay	1.6	23	150	
6	Loch Awe	1.2	41	94	
7					
8					
9					

Select in Excel

Step 2.

Click on an empty space of your workbook where you would like new dataset. **Paste** the selected column.

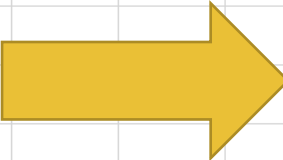
[illegible]

Select in Excel

Step 3.

Repeat steps 1 and 2 until you have **selected all the columns** you want like to keep.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	loch	volume	length	max_depth					loch	length				
2	Loch Ness	7.45	36.2	227					Loch Ness	36.2				
3	Loch Lomond	2.6	36	190					Loch Lomond	36				
4	Loch Morar	2.3	18.8	310					Loch Morar	18.8				
5	Loch Tay	1.6	23	150					Loch Tay	23				
6	Loch Awe	1.2	41	94					Loch Awe	41				
7														



Your turn...



Which variables do you think you would need to select if you wanted to have the loch name and its length?

loch	volume	length	max_depth
Loch Ness	7.45	36.2	227
Loch Lomond	2.6	36.0	190
Loch Morar	2.3	18.8	310
Loch Tay	1.6	23.0	150
Loch Awe	1.2	41.0	94



Your turn...



Select the name of the **loch** and its **length** from this dataset

loch	volume	length	max_depth
Loch Ness	7.45	36.2	227
Loch Lomond	2.6	36.0	190
Loch Morar	2.3	18.8	310
Loch Tay	1.6	23.0	150
Loch Awe	1.2	41.0	94

select

loch	length
Loch Ness	36.2
Loch Lomond	36.0
Loch Morar	18.8
Loch Tay	23.0
Loch Awe	41.0

Next steps

Complete **questions 1 to 8**
in **section 1** of the
'Manipulating dataset columns in Excel' workbook.

Definition



Reorder

To change the order of
columns in a dataset

Show me...



Here is an example of reordering columns.

id	fruit	colour
1	Apples	Red
3	Bananas	Yellow
5	Grapes	Green

Reorder the **fruit**
column to be 1st

fruit	id	colour
Apples	1	Red
Bananas	3	Yellow
Grapes	5	Green



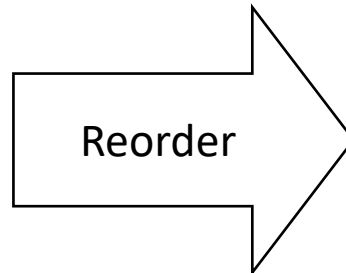
Example

Reorder

To change the order of columns in a dataset

Reorder this dataset so that 'locality' is the first column.

status	population	locality
Town	75,120	East Kilbride
Town	57,030	Livingston
City	148,280	Dundee
Town	50,010	Kirkcaldy



locality	status	population
East Kilbride	Town	75,120
Livingston	Town	57,030
Dundee	City	148,280
Kirkcaldy	Town	50,010

Your turn...



What order do you think these columns could be reordered into so that they are in a logical order?

temperature	dusk	dawn	location
15	22:50	03:30	Edinburgh
21	22:28	05:09	Paris
17	17:21	06:25	Sydney
17	20:55	04:54	New York

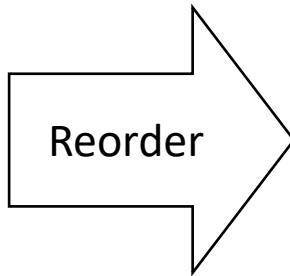


Your turn...



One option for reordering this dataset is,

temperature	dusk	dawn	location
15	22:50	03:30	Edinburgh
21	22:28	05:09	Paris
17	17:21	06:25	Sydney
17	20:55	04:54	New York

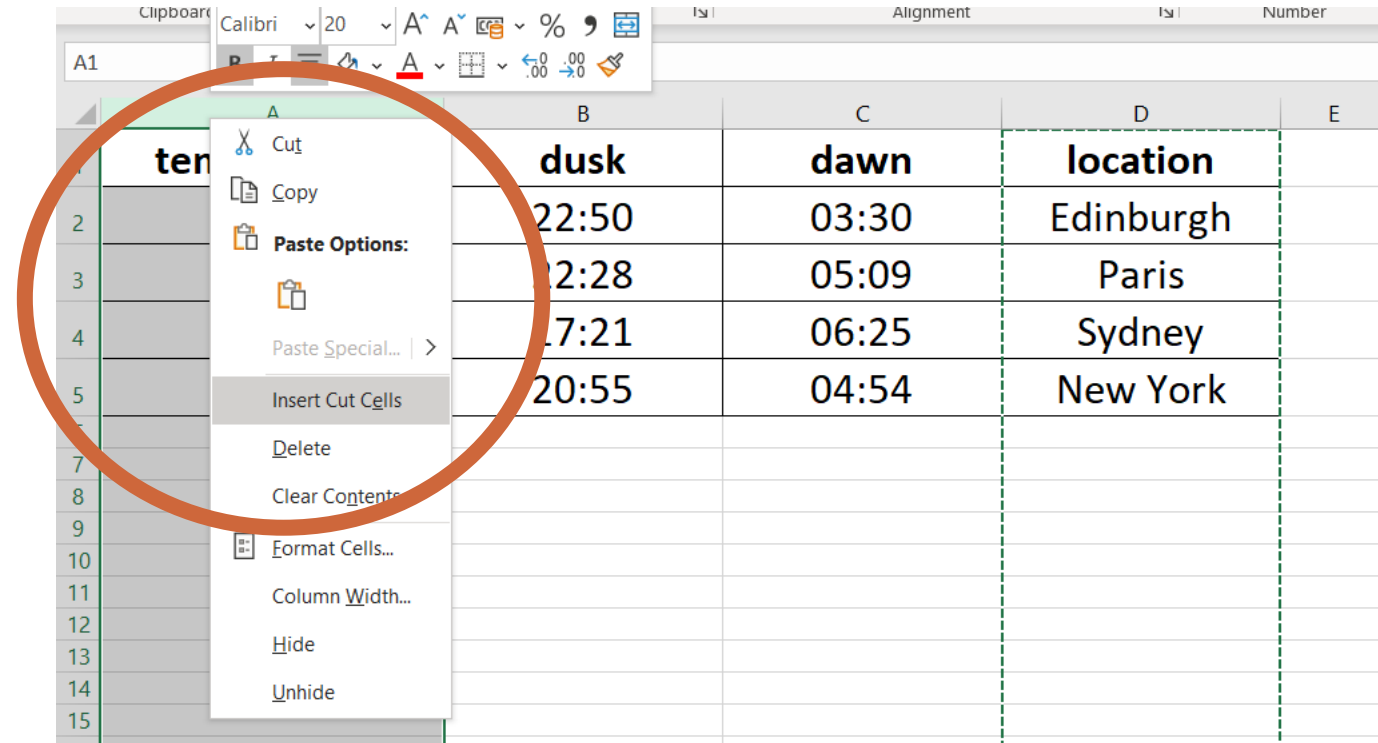


location	temperature	dawn	dusk
Edinburgh	15	03:30	22:50
Paris	21	05:09	22:28
Sydney	17	06:25	17:21
New York	17	04:54	20:55

Reorder in Excel

Step 2.

Right click on the column header to **the right of where you would like the data to go** and click on 'Insert Cut Cells'



Reorder in Excel

Step 3.

Repeat steps 1 and 2 until the **columns are all reordered**.

	A	B	C	D	E
1	location	temperature	dawn	dusk	
2	Edinburgh	15	03:30	22:50	
3	Paris	21	05:09	22:28	
4	Sydney	17	06:25	17:21	
5	New York	17	04:54	20:55	
6					
7					
8					

Alternatively, you can follow the **steps for selecting columns** and copy and paste them in the **new order** you need.

Next steps

Complete **questions 1 to 8**
in **section 2** of the
'Manipulating dataset columns in Excel' workbook.

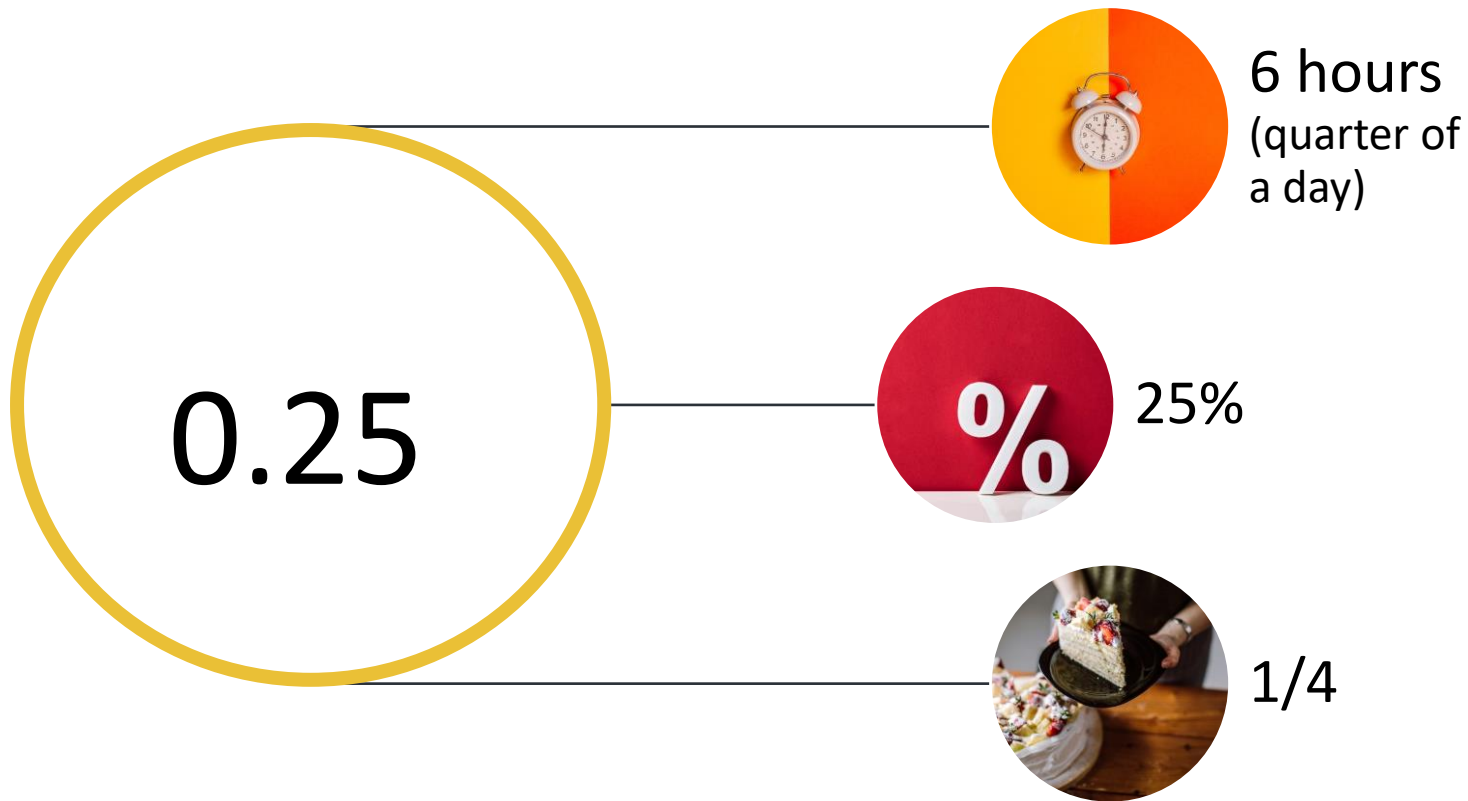
Definition



Reformat

To change how data in a column is displayed

Show me...



A number can be reformatted to appear in different ways.

However, the underlying number stays the same.

Changing the display format can make the data easier to understand.

Show me...



Value
0.25
0.5
0.75
1

Value
25%
50%
75%
100%

Value
£ 0.25
£ 0.50
£ 0.75
£ 1.00

Value
1/4
1/2
3/4
1

All of these columns of data have the same stored values, but they have been **reformatted to be displayed differently**.

Example

Reformat

To change how data in a column is displayed

Reformat the data in this column so that the values are in £s.

Value
£12.35
0.1233
29/4/1901
154.1259

Reformat

Value
£12.35
£0.12
£485.00
£154.13

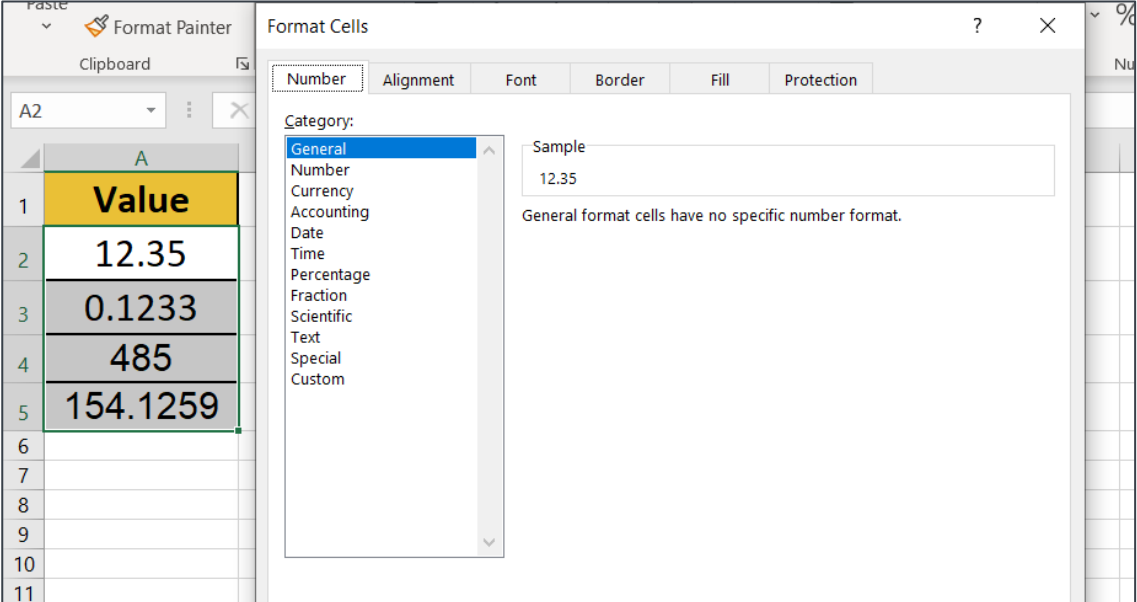
How to Reformat in Excel

Step 1.

Highlight the column of data you would like to reformat

Step 2.

Press **'Ctrl' and 1** if you are using Windows or **'Command' and 1** for Mac at the same time to open the format cells box



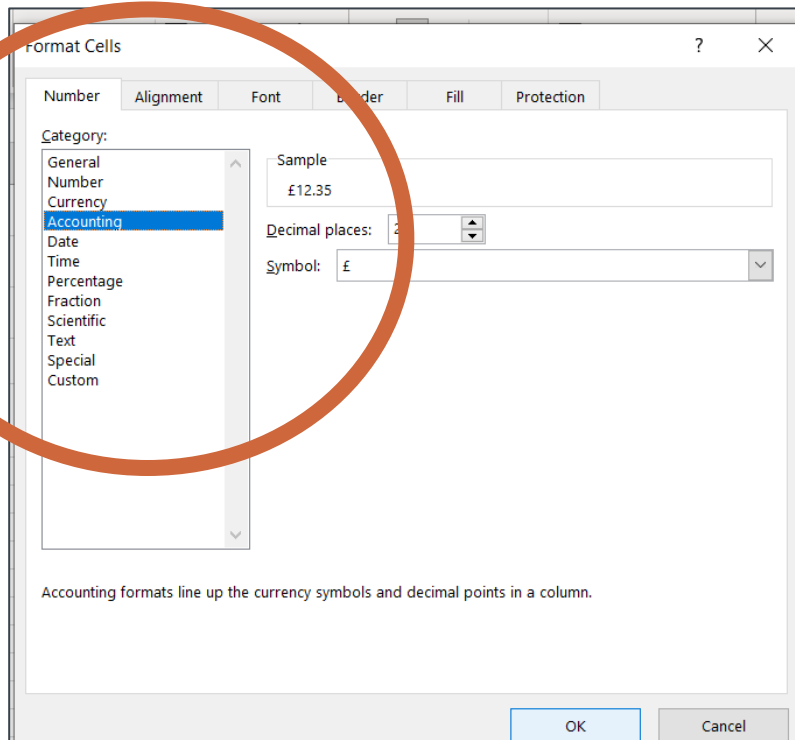
The screenshot shows an Excel spreadsheet with column A highlighted. The cells in column A contain the following values: "Value", "12.35", "0.1233", "485", and "154.1259". The "Format Cells" dialog box is open, showing the "Number" tab. The "Category" list on the left has "General" selected. The "Sample" box on the right shows "12.35".

Windows	Ctrl + 1
Mac	Command + 1

Reformat in Excel

Step 3.

Select the format option you need and press OK.



	A	B	C	D	E
1	Value			Value	
2	12.35			£ 12.35	
3	0.1233			£ 0.12	
4	485			£ 485.00	
5	154.1259			£ 154.13	
6					
7					
8					

Next steps

Complete **questions 1 to 6**
in **section 3** of the
'Manipulating dataset columns in Excel' workbook.

Additional information

In this lesson we have covered some of the procedures for manipulating data in Microsoft Excel.

However, there are different options (including keyboard shortcuts) available.

For more information, please see

<https://support.microsoft.com/en-us/excel>

Learning checklist

I can *describe* what it means to select, reorder and reformat columns in a data frame.

I can *manipulate* data by selecting, reordering and reformatting columns in Excel.

How you can use this lesson



You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform and build upon the material

Under the following terms:

- **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for [commercial purposes](#).
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

© 2021. This work is licensed under a [CC BY-NC-SA 4.0 license](#).

Created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

