

Creating variables by extracting and combining in Excel



Learning intentions

We will be learning how to create new variables, specifically to,

- understand what it means to **extract data to create a new variable**
- how to create simple new variables by extracting data in Excel
- understand what it means to **combine data to create a new variable**
- how to create simple new variables by combining data in Excel.

Creating new variables from strings

When you have variables that contain strings you can **combine** multiple data items or **extract data items** from it.

first_name	last_name	full_name
Hermione	Grainger	H. Grainger
Jay	Gatsby	J. Gatsby
Elizabeth	Bennet	E. Bennet
Scout	Finch	S. Finch

Creating new variables from dates

When you have variables that contain dates you can **extract data items** from them, to create new variables.

event	date	year
Rosa Parks arrested for not moving to the back of the bus	01-12-1955	1955
First successful human heart transplant performed	02-01-1968	1968
First Mickey Mouse comic strip appears	13-01-1930	1930
WWII begins as Germany invades Poland	01-09-1939	1939

Definition



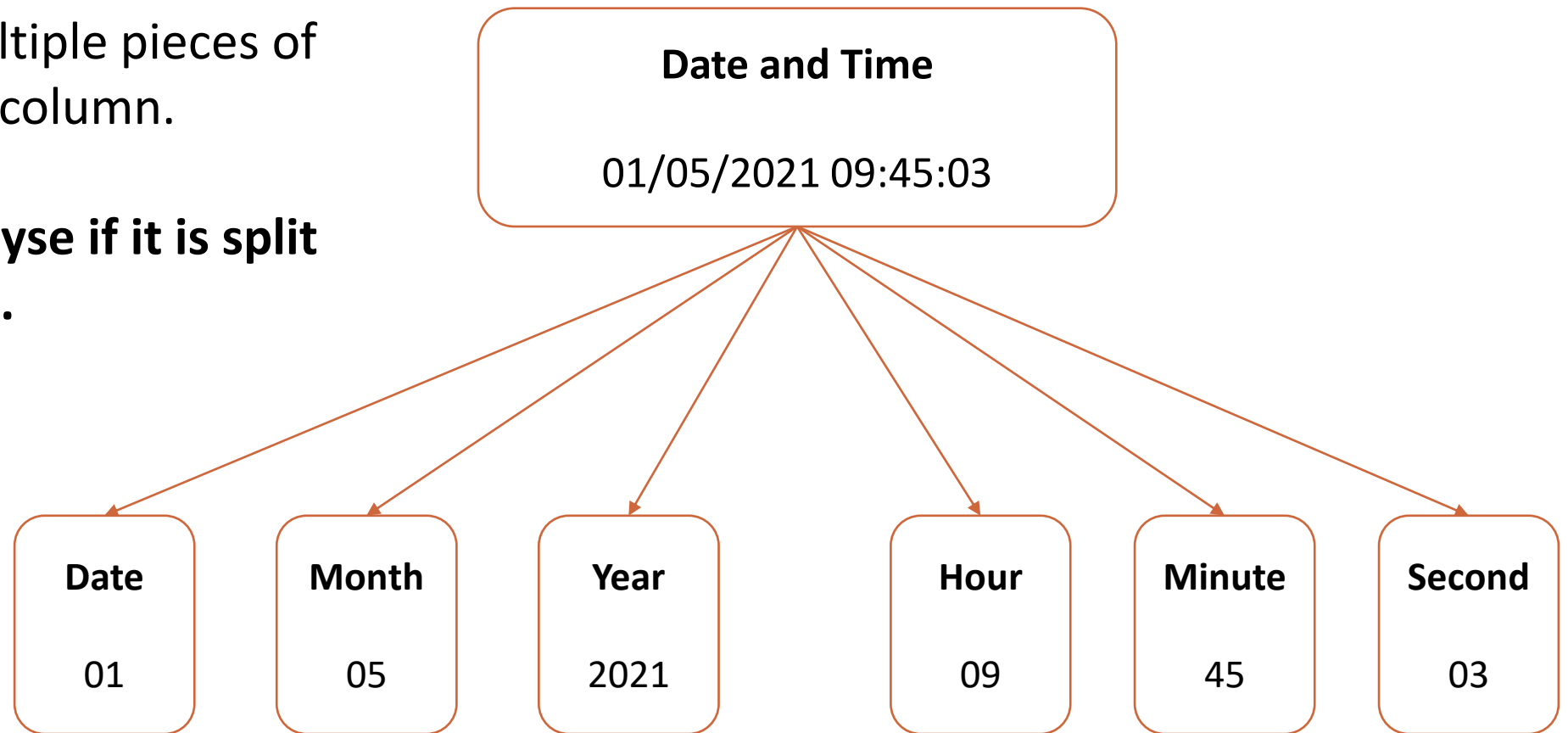
Extract

To split or select part of an existing data item in a dataset to make a new variable.

Extracting data

When data is collected it can sometimes contain multiple pieces of information in a single column.

It can **be easier to analyse if it is split into multiple variables.**

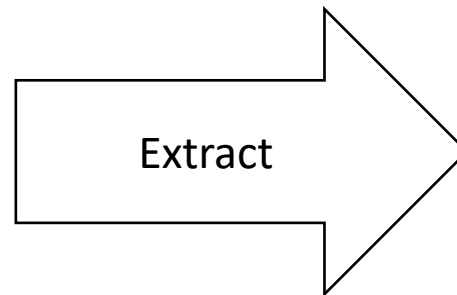


Show me...



Extract the athletes first names from these names.

name	date_of_birth
Usain Bolt	21-07-1986
Muhammad Ali	17-01-1942
Chris Hoy	23-03-1976
Liz McColgan	24-05-1964



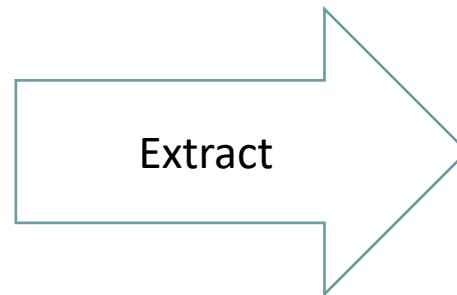
first_name
Usain
Muhammad
Chris
Liz

Show me...



Extract the year of birth from these dates of birth.

name	date_of_birth
Usain Bolt	21 Aug 1986
Muhammad Ali	17 Jan 1942
Chris Hoy	23 March 1976
Liz McColgan	24 May 1964



year_of_birth
1986
1942
1976
1964

Definition



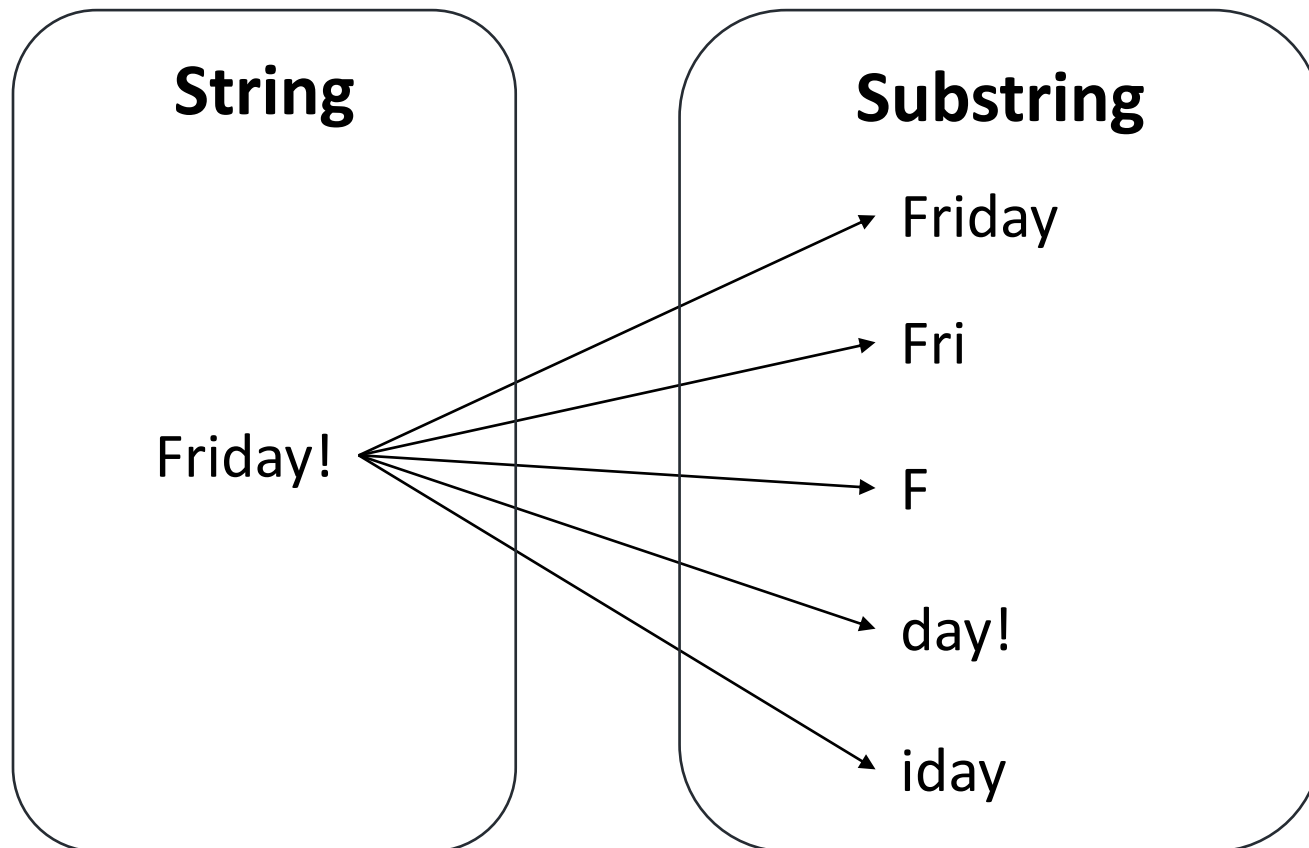
Substring

A sequence of characters that
exist inside a string

Show me...

Reminder

A string is a collection of characters combined to create alphanumeric text



The substring is any sequence of characters that are inside the original text.

We can use the substring text to create new variables.

Your turn...



Can you think of some examples of substrings in this string?

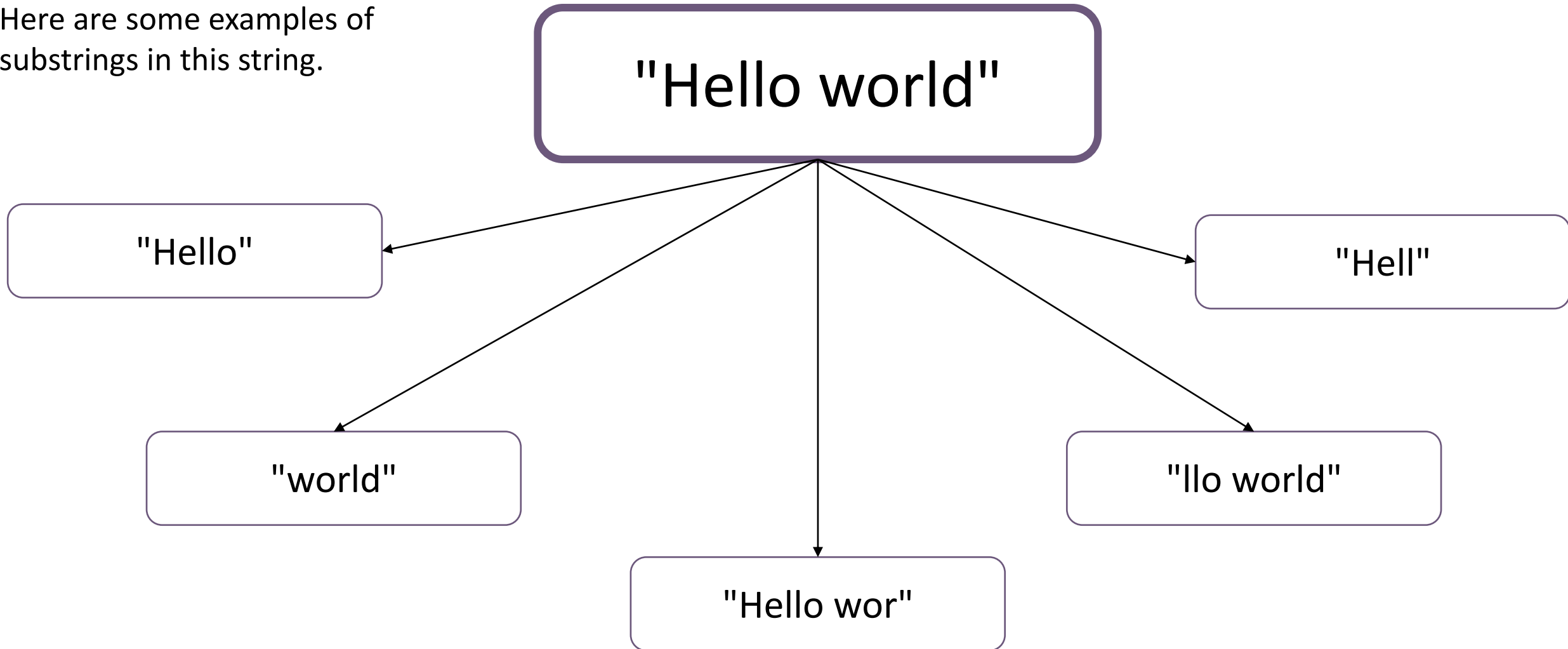
"Hello world"

Your turn...

Substring

A sequence of characters that exist inside a string

Here are some examples of substrings in this string.



Example of extracting from a string

The Community Health Index Number (CHI) helps NHS Scotland uniquely identify patients. The **first 6 digits** are the patients date of birth (DDMMYY).

Extract the patients date of birth from the **chi_number** variable.

CHI
1205651234
0608214587
0112054785
3001855478

Extract 1st 6
characters

date_of_birth
120565
060821
011205
300185



Your turn...



You need to find out the towns these people live in. What part of the address do you think you need to extract?

name	Address
Sherlock Holmes	221B Baker Street, London
Wallace and Gromit	62 West Wallaby Street, Wigan
The Banks Family	17 Cherry Tree Lane, London
The Dursley Family	4 Privet Drive, Little Whinging

Your turn...



You need to find out the towns these people live in. What part of the address do you think you need to extract?

name	Address
Sherlock Holmes	221B Baker Street, London
Wallace and Gromit	62 West Wallaby Street, Wigan
The Banks Family	17 Cherry Tree Lane, London
The Dursley Family	4 Privet Drive, Little Whinging

Extract after the
comma

Town
London
Wigan
London
Little Whinging

Extract in Excel

When you are extracting data to create a new variable in Excel you can split the data based on,

- Number of characters
- Searching for symbols or characters e.g. comma, space



Extract in Excel

You can use the formula below to keep only the number of characters you state, counting from the left of the text.

`=LEFT(text,[num_chars])`

	A	B
1	CHI	
2	1205651234	=LEFT(A2,6)
3	0608214587	
4	112054785	
5	3001855478	
6		

Extract using a formula
to keep the 1st 6
characters

	A	B
1	CHI	
2	1205651234	120565
3	0608214587	060821
4	112054785	112054
5	3001855478	300185
6		

Extract in Excel

Here are some examples of calculations you can use in Excel to extract data.

Formula	Description	Example	Result
LEFT	Extracts a given number of characters from the left side	=LEFT("apple",3)	"app"
MID	Extracts a given number of characters from a defined starting point	=MID("abcde",2,3)	"bcd"
RIGHT	Extracts a given number of characters from the right side	=RIGHT("red",2)	"ed"

For more details of the formulas in Excel, see <https://support.microsoft.com/en-us/excel>

Extract in Excel

If you need to **extract data based on a character** such as a comma or space, you will need to combine formulas.

Part 1: **Search the text** for the character.

The SEARCH formula returns the position of the character you are searching for.

`=SEARCH(find_text,text)`

There is a comma 5 characters from the left.

Name	
Jo,Bloggs	=SEARCH(",",A35)
Jane,Smith	5
Liz,Brown	4
Jack,Red	5

Extract in Excel

Part 2: Use the **SEARCH** formula to define the number of characters in the **LEFT** formula.

=LEFT(text,SEARCH(find_text,text))

Name		
Jo,Bloggs	3	=LEFT(A35,SEARCH(",",A35))
Jane,Smith	5	Jane,
Liz,Brown	4	Liz,
Jack,Red	5	Jack,

The left function has kept the first 5 characters.

Extract in Excel

Part 3: We don't need to keep the character we are splitting the data by (e.g. the comma). Therefore we need to **take away 1** from the SEARCH function output.

`=LEFT(text,SEARCH(find_text,text)-1)`

Name			
Jo,Bloggs	3	Jo,	=LEFT(A35,SEARCH(",",A35)-1)
Jane,Smith	5	Jane,	Jane
Liz,Brown	4	Liz,	Liz
Jack,Red	5	Jack,	Jack

Extract in Excel

If you would like to keep the data after the character. You need to combine the LEN, RIGHT and SEARCH functions.

=RIGHT(text,LEN(text)-SEARCH(find_text,text))

43	Name	Before	After
44	Jo,Bloggs	Jo	=RIGHT(A44,LEN(A44)-SEARCH(",",A44))
45	Jane,Smith	Jane	Smith
46	Liz,Brown	Liz	Brown
47	Jack,Red	Jack	Red
48			

Next steps

Complete **section 1** of the
'Creating variables by extracting
and combining in Excel' workbook.

Definition



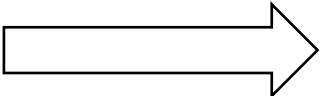
Combine

Join or merge variables to create a single variable.

Combining data

Sometimes we may wish to create a variable by *combining* two or more data items from the same dataset.

The data items being joined are usually **strings**.

"Usain" + "Bolt"  "Usain Bolt"

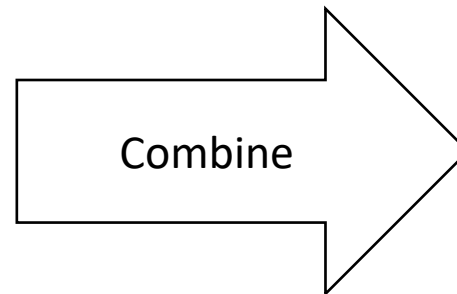


Show me...



Combine the first and last names to make the athletes' names.

first_name	last_name
Usain	Bolt
Muhammad	Ali
Chris	Hoy
Liz	McColgan



full_name
Usain Bolt
Muhammad Ali
Chris Hoy
Liz McColgan

Example

Combine the name and team of each cyclist, and format like this: **Tadej Pogačar (UAE Team Emirates)**

cyclist	team
Tadej Pogačar	UAE Team Emirates
Egan Bernal	Team Ineos
Bradley Wiggins	Team Sky
Chris Froome	Team Sky



Example

Combine the name and team of each cyclist, and format like this: **Tadej Pogačar (UAE Team Emirates)**

cyclist	team	cyclist_and_team
Tadej Pogačar	UAE Team Emirates	Tadej Pogačar (UAE Team Emirates)
Egan Bernal	Team Ineos	Egan Bernal (Team Ineos)
Bradley Wiggins	Team Sky	Bradley Wiggins (Team Sky)
Chris Froome	Team Sky	Chris Froome (Team Sky)



Combining strings in Excel

In Excel there are many different ways to combine data. The simplest way is to use **& sign** between the strings.

= cell1 & cell2 & cell3 & cell4.....

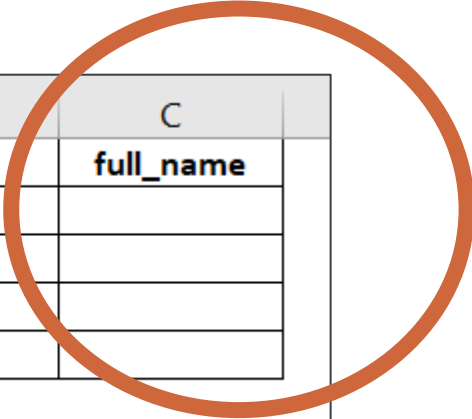
For more options, please see

<https://support.microsoft.com/en-us/office/combine-text-from-two-or-more-cells-into-one-cell-81ba0946-ce78-42ed-b3c3-21340eb164a6>

Create a new variable by combining in Excel

We want to create a new variable **full_name** by combining **first_name** and **last_name**.

Step 1. Create a new variable with a heading name of **full_name**

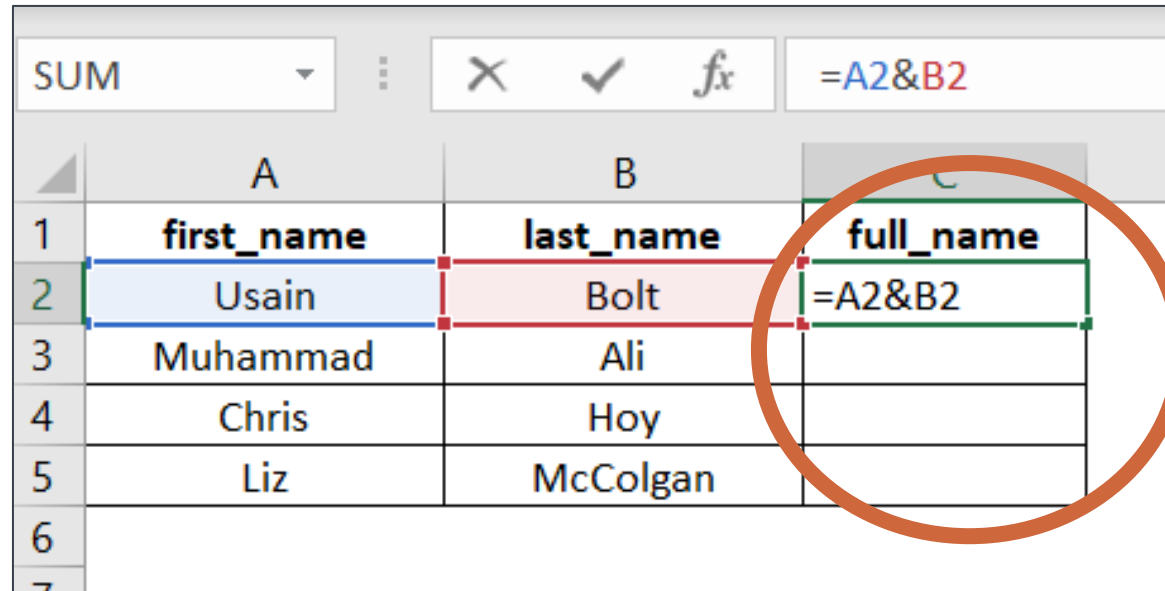


	A	B	C
1	first_name	last_name	full_name
2	Usain	Bolt	
3	Muhammad	Ali	
4	Chris	Hoy	
5	Liz	McColgan	
6			
7			

Using & sign

Step 2. In the first row, use the & sign to combine the names.

= first_name & last_name



	A	B	C
1	first_name	last_name	full_name
2	Usain	Bolt	=A2&B2
3	Muhammad	Ali	
4	Chris	Hoy	
5	Liz	McColgan	
6			
7			

Adding in other characters

Sometimes we want other characters added into the string such as spaces, full stops or numbers.

You can add any characters (or strings) by putting them in quotes (“ ”).

= cell1 & **“Text”** & cell2 & cell3 & cell4.....

Adding in other characters

= first_name &" "& last_name

Without added space

SUM			
=A3&B3			
	A	B	C
1	first_name	last_name	full_name
2	Usain	Bolt	UsainBolt
3	Muhammad	Ali	=A3&B3
4	Chris	Hoy	ChrisHoy
5	Liz	McColgan	LizMcColgan
6			
7			

With added space

SUM			
=A3&" "&B3			
	A	B	C
1	first_name	last_name	full_name
2	Usain	Bolt	Usain Bolt
3	Muhammad	Ali	=A3&" "&B3
4	Chris	Hoy	Chris Hoy
5	Liz	McColgan	Liz McColgan
6			
7			

Combining date parts in Excel

If you have the parts of a date (e.g. day, month and year) you can combine them into a single date using the Excel function DATE()

= date(year,month,day)

Combining date parts in Excel

Reminder: Date/Time data type
the number of days or seconds
passed since the 'epoch' date

= date(year,month,day)

	A	B	C	D
1	day	month	year	date
2	1	4	2025	=DATE(C2,B2,A2)
3	4	12	2002	04/12/2002
4	19	1	1965	19/01/1965
5				

The DATE() function
has combined the
numbers into a new
data item.

Next steps

Complete **section 2** of the
‘Creating variables by extracting
and combining in Excel’ workbook.

Then the complete
section 3 - Extension

Learning checklist

I can *describe* how to create a new variable by extracting data.

I can *create* new variables in Excel by extracting data.

I can *describe* how to create a new variable by combining data.

I can *create* new variables in Excel by combining data.

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.

