

# Reshaping datasets



Worksheet section	Contents
1	Wide and long datasets
2	Process for reshaping

This lesson has been created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

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



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.

# 1. Wide and long datasets

## Section 1.1

1) Fill in the gaps of these descriptions of the columns in a **long** dataset.

id                      labels the \_\_\_\_\_  
key                    \_\_\_\_\_ of the data item  
value                contains the value of the \_\_\_\_\_

2) Select the correct definition of a **wide** dataset.

- ☐ Each different data variable is in a separate row
- ☐ Each different data variable is in a separate column

## Section 1.2

3) Label these datasets as wide or long datasets

Country	population _1900	population _2000
Scotland	4,437,000	5,063,000
Canada	5,301,000	31,689,000
Greece	2,504,070	11,120,000
New Zealand	802,200	4,028,000

ID	test_1	test_2
GF15421	45%	41%
SD14582	21%	95%
WS12452	33%	93%
AW5248	96%	62%

building_or_statue	measurement	value
Scott Monument	height	61
The Glasgow Tower	height	127
Shanghai Tower	height	632
Empire State building	height	381
Nelson's column	height	52
The Shard	height	310

name	address_part	value
A. Horn	town	Selkirk
G. Davies	town	Crail
V. Corrs	town	Wigtown
R. Jones	town	Lanark
A. Horn	street	Ashby Close
G. Davies	street	Avon Street
V. Corrs	street	Law View
R. Jones	street	Birks Street

# 1. Wide and long datasets

## Section 1.3

4) Explain why this dataset from Traffic Scotland is wide.

Destination	Current Journey Time	Typical Journey Time	Delay
M8 J19 Anderston	55 mins	54 mins	Less than 1 minute
M8 J1 Hermiston Gait (WB)	11 mins	12 mins	No delay
M8 J25A Braehead	59 mins	59 mins	Less than 1 minute
M8 J28 Glasgow Airport	1 hour, 2 mins	1 hour, 2 mins	Less than 1 minute
M8 J8 Baillieston	45 mins	45 mins	Less than 1 minute

<https://trafficscotland.org/journeytimes/list/>

5) What are the **id**, **key** and **value** columns in this **long** dataset? This dataset shows Data Science and STEM Salaries from top companies.

timestamp	company	title	# totalyearly...
8/17/2021 8:28:57	Adobe	Product Designer	300000
8/17/2021 8:26:21	HSBC	Software Engineer	159000
8/17/2021 8:24:56	Cisco	Software Engineer	154000
8/17/2021 8:22:17	Fidelity Investments	Software Engineer	98000
8/17/2021 8:16:36	Amazon	Product Manager	241000
8/17/2021 7:55:47	AT&T	Data Scientist	175000
8/17/2021 7:50:25	Raytheon Technologies	Software Engineer	95000

id

key (or keys)

value

Source: <https://www.kaggle.com/jackogozaly/data-science-and-stem-salaries>

6) Can you give a benefit of wide dataset and a benefit of long dataset?

Benefit of wide data	Benefit of long data

# 1. Wide and long datasets

## Section 1.4

- 7) Can you go online and find an example of a wide and long dataset? Copy and paste them below and explain why they are wide or long.

### Wide

Explain why it is wide

Paste your dataset here.

### Long

Explain why it is long

Paste your dataset here.

## 2. Process for reshaping datasets

### Section 2.1

- 1) Explain what it means to reshape a dataset

### Section 2.2

- 2a) Calculate how many years these TV shows ran for in the wide and long datasets.

Long

TV_show	date_type	date
friends	first show	1994
friends	last show	2004
Game of Thrones	first show	2011
Game of Thrones	last show	2019
The Big Bang Theory	first show	2007
The Big Bang Theory	last show	2019
M*A*S*H	first show	1972
M*A*S*H	last show	1983

TV_show	years
friends	
Game of Thrones	
The Big Bang Theory	
M*A*S*H	

Wide

TV_show	first_show	last_show	years
Game of Thrones	2011	2019	
M*A*S*H	1972	1983	
The Big Bang Theory	2007	2019	
friends	1994	2004	

- 2b) Which did format did you find easier to calculate how long these TV shows ran for and why?

## 2. Process for reshaping datasets

3a) Summarise this dataset in the wide and long format to work out the **total\_population** by year and **% urban population** by year

Long

population_type	year	world_population
urban	2020	4,378,993,944
rural	2020	3,415,804,795
urban	2000	2,868,307,513
rural	2000	3,275,186,310
urban	1980	1,754,201,029
rural	1980	2,703,802,485
urban	1960	1,023,845,517
rural	1960	2,011,104,231

year	total_population	%_urban
2020		
2000		
1980		
1960		

Wide

population_type	1960	1980	2000	2020
rural	2,011,104,231	2,703,802,485	3,275,186,310	3,415,804,795
urban	1,023,845,517	1,754,201,029	2,868,307,513	4,378,993,944
<b>Total_population</b>				
<b>% urban</b>				

Source:

<https://www.worldometers.info/world-population/world-population-by-year/>

3b) Which did format did you find easier to summarise and why?

## 2. Process for reshaping datasets

### Section 2.3

- 4) This dataset is being reshaped from wide to long. Can you fill in the gaps in the long dataset?

planet	distance_from_sun	radius_km	mass_vs_earth
Jupiter	748,000,000	69911	1321
Saturn	1,483,000,000	58232	95.159

planet	measurement	
Jupiter	distance_from_sun	748,000,000
Saturn	distance_from_sun	1,483,000,000
Jupiter		69,911
Saturn		58,232
Jupiter	mass_vs_earth	
Saturn	mass_vs_earth	95

- 5) This dataset is being reshaped from long to wide. Can you fill in the gaps in the wide dataset?

song	information	value
Happy	artist	Pharrell
Happy	year released	2013
Umbrella	artist	Rihanna
Umbrella	year released	2008
Smells Like Teen Spirit	artist	Nirvana
Smells Like Teen Spirit	year released	1991
Don't Stop Believin'	artist	Journey
Don't Stop Believin'	year released	1981
Sweet Home Alabama	artist	Lynyrd Skynyrd
Sweet Home Alabama	year released	1974
She Loves You	artist	The Beatles
She Loves You	year released	1963
Jailhouse Rock	artist	Elvis Presley
Jailhouse Rock	year released	1957

song		
Happy	Pharrell	2013
Umbrella	Rihanna	2008
Smells Like Teen Spirit	Nirvana	
Don't Stop Believin'	Journey	1981
Sweet Home Alabama	Lynyrd Skynyrd	1974
She Loves You		1963
Jailhouse Rock		1957

## 2. Process for reshaping datasets

### Section 2.4

- 6) Find the weather forecast in a location any where in the world that interests you and put the data into a wide and long dataset.

This should include,

- forecast for 4 days or times
- temperature
- wind speed
- change of rain (precipitation)

Here are some websites where you can find weather forecasts

<https://www.bbc.co.uk/weather>

<https://www.metoffice.gov.uk/>

<https://weather.com/en-GB/>

<https://www.accuweather.com/>

Wide

forecast				

Long

	forecast	