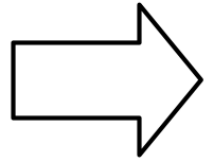


# Reshaping data

# Last time: `pivot_longer`

id	bp1	bp2
A	100	120
B	140	115
C	120	125



id	measurement	value
A	bp1	100
A	bp2	120
B	bp1	140
B	bp2	115
C	bp1	120
C	bp2	125

```
1 df |>
2   pivot_longer(
3     cols = bp1:bp2,
4     names_to = "measurement",
5     values_to = "value"
6   )
```

# Example from last time

```
1 df_3
```

	id	bp_1	bp_2	hr_1	hr_2
1	1	100	120	60	77
2	2	120	115	75	81
3	3	125	130	80	93

```
1 df_3 |>  
2   pivot_longer(cols = -id,  
3                 names_to = c("measurement", "stage"),  
4                 names_sep = "_",  
5                 values_to = "value")
```

```
# A tibble: 12 × 4
```

	id	measurement	stage	value
	<dbl>	<chr>	<chr>	<dbl>
1	1	bp	1	100
2	1	bp	2	120
3	1	hr	1	60
4	1	hr	2	77
5	2	bp	1	120
6	2	bp	2	115

7	2 hr	1	75
8	2 hr	2	81
9	3 bp	1	125
10	2 hr	2	120

# Warmup activity

Work on the activity (handout) with a neighbor, then we will discuss as a class

# Warmup

```
1 df_3
2
3 df_3 |>
4   pivot_longer(cols = -id,
5                 names_to = c(".value", "stage"),
6                 names_sep = "_")
```

	id	bp_1	bp_2	hr_1	hr_2
1	1	100	120	60	77
2	2	120	115	75	81
3	3	125	130	80	93

	id	stage	bp	hr
1	1	1	100	60
2	1	2	120	77
3	2	1	120	75
4	2	2	115	81
5	3	1	125	80
6	3	2	130	93

What is `names_to = c(".value", "stage")` doing?

# Going the other way

Data on air quality in two locations (BETR801, London Westminster) on different days:

```
1 air_quality
```

```
# A tibble: 1,825 × 3
```

	date.utc	location	value
	<dtm>	<chr>	<dbl>
1	2019-06-18 06:00:00	BETR801	18
2	2019-06-17 08:00:00	BETR801	6.5
3	2019-06-17 07:00:00	BETR801	18.5
4	2019-06-17 06:00:00	BETR801	16
5	2019-06-17 05:00:00	BETR801	7.5
6	2019-06-17 04:00:00	BETR801	7.5
7	2019-06-17 03:00:00	BETR801	7
8	2019-06-17 02:00:00	BETR801	7
9	2019-06-17 01:00:00	BETR801	8
10	2019-06-16 01:00:00	BETR801	15

What if I want a separate column for each location?

# pivot\_wider

# A tibble: 3 × 3

	date.utc	location	value
	<dtm>	<chr>	<dbl>
1	2019-06-18 06:00:00	BETR801	18
2	2019-06-17 08:00:00	BETR801	6.5
3	2019-06-17 07:00:00	BETR801	18.5

```
1 air_quality |>
2   pivot_wider(id_cols = ...,
3               names_from = ...,
4               values_from = ...)
```



# pivot\_wider

```
# A tibble: 3 × 3
```

	date.utc	location	value
	<dtm>	<chr>	<dbl>
1	2019-06-18 06:00:00	BETR801	18
2	2019-06-17 08:00:00	BETR801	6.5
3	2019-06-17 07:00:00	BETR801	18.5

```
1 air_quality |>
2   pivot_wider(id_cols = "date.utc",
3               names_from = "location",
4               values_from = "value")
```

```
# A tibble: 1,670 × 3
```

	date.utc	BETR801	`London Westminster`
	<dtm>	<dbl>	<dbl>
1	2019-06-18 06:00:00	18	7
2	2019-06-17 08:00:00	6.5	6
3	2019-06-17 07:00:00	18.5	6
4	2019-06-17 06:00:00	16	6
5	2019-06-17 05:00:00	7.5	6
6	2019-06-17 04:00:00	7.5	6
7	2019-06-17 03:00:00	7	6

8	2019-06-17	02:00:00	7	6
9	2019-06-17	01:00:00	8	6
10	2019-06-16	01:00:00	15	7

# Class activity

[https://sta279-f25.github.io/class\\_activities/ca\\_05.html](https://sta279-f25.github.io/class_activities/ca_05.html)

- Work with a neighbor on the class activity
- At the end of class, submit your work as an HTML file on Canvas (one per group, list all your names)

**For next time, read:**

- Chapter 5 in *Modern Data Science with R*