# Reshaping Data: Long / Wide

# Repeated Measures

- In this workshop, we will be talking about reshaping data from long to wide forms (and back again)

- These forms of data occur most often for repeated measures of a variable (like blood pressure over 5 visits to the doctor, or temperature over twelve months)

- The need to do this usually occurs because it is far easier to record and enter data in a **wide format** (with the visits or months across the top of the page), but it is often easier to analyze and visualize data in *long format*

# Wide Data

- An example of a wide data set could be this one, comparing my temperature with that of a lizard

| Name | Temp1 | Temp2 | Temp3 | Temp4 | Temp5 | Temp6 |
|---|---|---|---|---|---|---|
| Cale | 98.4 | 98.7 | 98.6 | 98.3 | 98.5 | 98.6 |
| Lizard | 69 | 75 | 84 | 92 | 86 | 79 |

- This is a totally reasonable way to record/enter these temperatures in a spreadsheet, but as soon as we consider how to graph or analyze them…

can we **summarize**? can we **graph twoway line**?

# Long Data

- A much friendlier data format for commands is the long format
- It may seem like there is a lot of repetition, but now each observation represents the unit of analysis we care about — an animal/time/ temperature combination

| Name | Time | Temp |
|---|---|---|
| Cale | 1 | 98.4 |
| Cale | 2 | 98.7 |
| Cale | 3 | 98.6 |
| Cale | 4 | 98.3 |
| Cale | 5 | 98.5 |
| Cale | 6 | 98.6 |
| Lizard | 1 | 69 |
| Lizard | 2 | 75 |
| Lizard | 3 | 84 |
| Lizard | 4 | 92 |
| Lizard | 5 | 86 |
| Lizard | 6 | 79 |

# Reshaping W/L

- In Stata, the command **`reshape`** transforms wide data to long data and back again:

*long*

| *i* | *j* | *stub* |
|---|---|---|
| 1 | **1** | 4.1 |
| 1 | **2** | 4.5 |
| 2 | **1** | 3.3 |
| 2 | **2** | 3.0 |

reshape <⟷>

*wide*

| *i* | *stub***1** | *stub***2** |
|---|---|---|
| 1 | 4.1 | 4.5 |
| 2 | 3.3 | 3.0 |

To go from wide to long:

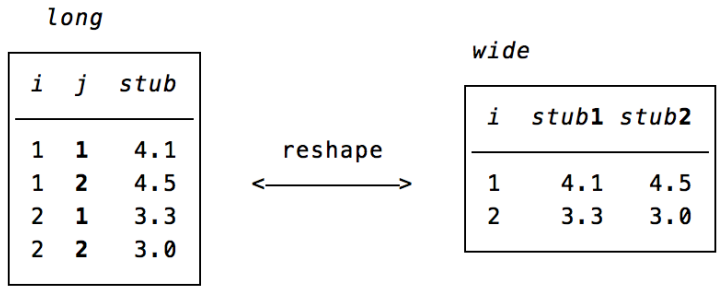
```
reshape long stub, i(i) j(j)
                        \
                         j new variable
```

# Reshaping W/L

```
import excel cale_lizard, clear firstrow
```

To go from wide to long:

```
reshape long stub, i(i) j(j)
                           \
                            j new variable
```




```
reshape long Temp, i(Name) j(Time)
```

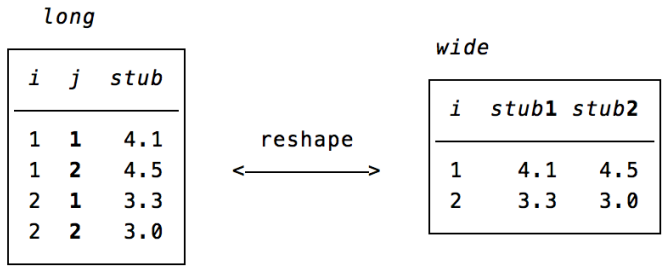
- `stub` is the repeated measure, `i` is the identifier, and `j` is the new variable we wish to create to identify the time points we peeled off of `stub`

# Reshaping L/W

To go from long to wide:

*j* existing variable

**reshape wide** *stub*, **i**(*i*) **j**(*j*)




```
reshape wide Temp, i(Name) j(Time)
```

- `stub` is the repeated measure, `i` is the identifier, and j the variable in our long dataset that indicates the repeated time or date

# Switching W/L L/W

- Once we have defined the reshape `stub`, `i`, and `j` parameters for a dataset with **`reshape`**, we can go back and forth easily:

```
reshape wide
```

```
reshape long
```

# Exercises (1)

1. Wide to Long

   A. Load the reshape1 dataset: `webuse reshape1`

   B. Take a look at this dataset in the Data Browser

   C. Reshape this data from wide format to long format

# Exercises (2)

1. Long to Wide

   A. Load the bplong dataset: **webuse** bplong

   B. Take a look at the data in the browser

   C. Transform this data from long to wide