

# Data Reshaping in R

Using the furniture package's `long()` and `wide()` functions

Tyson S. Barrett | t.barrett@aggiemail.usu.edu | tysonstanley.github.io

Changing data from wide to long format is an essential tool in most data analyses. Long ("tidy") form is generally needed for plotting, statistical and other analyses.

Wide Format					Long Format				
ID	x1	x2	z1	z2	ID	Time	x	z	
a	1	8	.	.	a	1	1	.	
b	2	9	.	.	a	2	8	.	
c	3	10	.	.	b	1	2	.	
d	4	11	.	.	b	2	9	.	
					c	1	3	.	
					c	2	10	.	

## Wide to Long Form

df					long_df				
ID	x1	x2	z1	z2	ID	Time	x	z	
a	1	8	.	.	a	1	1	.	
b	2	9	.	.	a	2	8	.	
c	3	10	.	.	b	1	2	.	
d	4	11	.	.	b	2	9	.	
					c	1	3	.	
					c	2	10	.	

```

long_df <- long(df,
  c("x1", "x2"), c("z1", "z2"),
  v.names = c("x", "z"),
  timevar = "Time")
  
```

Newly created data.frame  
 Function name  
 Original data.frame  
 The time varying variables from original data.frame  
**long()** means we are going from wide to long form  
 The name of the new "time" variable  
 The new variable names for each group of measured variables

Note: there can be any number of vectors like the `c("x1", "x2")` and `c("z1", "z2")` but they must have the same number of elements. For unbalanced data (where some have more "times" than others, see Unbalanced Measures to the right).

## Long to Wide Form

df				wide_df				
ID	Time	x	z	ID	x1	x2	z1	z2
a	1	1	.	a	1	8	.	.
a	2	8	.	b	2	9	.	.
b	1	2	.	b	3	10	.	.
b	2	9	.	c	4	11	.	.
c	1	3	.					
c	2	10	.					

```

wide_df <- wide(df,
  v.names = c("x", "z"),
  timevar = "Time",
  idvar = "ID")
  
```

Newly created data.frame  
 Function name  
 Original data.frame  
 The variables that will be split into the different times (e.g., "x1" and "x2" stem from "x")  
 The ID of the individual. The function automatically looks for a variable with the letters "ID" within it.  
 The name of the time variable in the data frame.

## Unbalanced Measures

If the various measures have differing number of measurement points ("times"), then use the placeholder "miss". Say, our "x" measure had only one time point but our "z" measure had two, this code would work.

```

long(df,
  c("x1", "miss"), c("z1", "z2"),
  v.names = c("x", "z"),
  timevar = "time")
  
```

### Additional Notes Regarding Data Reshaping:

- Recommendation:** Save only one data set (in wide or long format) that is your "main" data set—to avoid confusion about what different data sets contain. Then, use these reshaping functions in R to manipulate the data to get it in the form needed to plot and analyze. Finally, save the code ("syntax") and any output (you can save the reshaped data but it is not necessary).
- Term Definitions:** *measure* implies any distinct variable in the data set (e.g., `x2`), *time* implies anything that distinguishes the observations (e.g., time, cluster, location), *time varying variable* is any variable that has or could have a different value at each observation occasion.