## Data manipulation in Stata

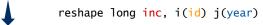
## Reshape command

The classic use of the reshape command is to split data into several variables according to some categorical variable\* (or to do the reverse):

reshape wide data, i(id) j(category)

id	year	sex	inc
1	80	0	5,000
1	81	0	2,000
1	82	0	3,000
2	80	1	5,500
2	81	1	2,200
2	82	1	2,000
3	80	0	6,000
3	81	0	3,300
3	82	0	1,000

reshape wide inc, i(id) j(year)



id	sex	inc80	inc81	inc82
1	0	5,000	5,500	6,000
2	1	2,000	2,200	3,300
3	0	3,000	2,000	1,000

id	year	sex	inc	ue
1	80	0	5,000	0
1	81	0	5,500	1
1	82	0	6,000	0
2	80	1	2,000	1
2	81	1	2,200	0
2	82	1	3,300	0
3	80	0	3,000	0
3	81	0	2,000	0
3	82	0	1,000	1

In some cases you may have more than one variable of data that you want to split according to your categorical variable:

reshape wide data1 data2, i(id) j(category)

reshape wide inc ue, i(id) j(year)



reshape long inc ue, i(id) j(year)

id	sex	inc80	inc81	inc82	ue80	ue81	ue82
1	0	5,000	5,500	6,000	0	1	0
2	1	2,000	2,200	3,300	1	0	0
3	0	3,000	2,000	1,000	0	0	1

\*This variable does not necessarily have to be a "categorical variable" in the traditional sense, but it is useful in this context to differentiate it somehow from the variable that contains the data.

reshape wide data, i(id category2) j(category1) reshape wide datacat1a datacat1b datacat1c, i(id) j(category2)



reshape wide inc, i(id ue) j(year)

reshape long inc, i(id ue) j(year)

id	sex	ue	inc80	inc81	inc82
1	0	0	5,000		6,000
1	0	1		5,500	
2	1	0		2,200	3,300
2	1	1	2,000		
3	0	0	3,000	2,000	
3	0	1			1,000

reshape wide inc80 inc81 inc82, i(id) j(ue)

Rather than having two variables of data that you want to split according to some category, you may instead want to sub-divide your data by two categories.

We can illustrate this with the same dataset by changing our interpretation of the 'ue' variable.

The process is a bit more complicated. It has to be done in two steps.



