Reshape data using RStudio

(v. 1.5)

Oscar Torres-Reyna

otorres@princeton.edu

January 2011

Reshaping Wide to Long

```
gdp.wide = read.csv("http://www.princeton.edu/~otorres/GDP.csv", header=TRUE)
```

	country	varA2001	varA2002	varA2003	varA2004	varA2005	varB2001	varB2002	varB2003	varB2004	varB2005
1	A	NA	NA	8000.01	8212.90	7847.36	7702.89	7288.48	6430.98	6932.45	7486.24
2	B	18268.01	18738.99	19360.46	20151.42	20715.54	20866.90	21364.02	21801.41	22404.59	22676.26
3	C	21088.14	21608.14	21988.64	22739.28	23436.61	24194.85	24300.57	24411.48	24650.02	25076.01

For the reshape to work it is important to reorder the variables first according to the numeric sequence, second in alphabetical order.

```
# Reorder variables for reshape

gdp.wide = gdp.wide[c("country", "varA2001", "varB2001",
                                 "varA2002", "varB2002",
                                 "varA2003", "varB2003",
                                 "varA2004", "varB2004",
                                 "varA2005", "varB2005")]
```

	country	varA2001	varB2001	varA2002	varB2002	varA2003	varB2003	varA2004	varB2004	varA2005	varB2005
1	A	NA	7702.89	NA	7288.48	8000.01	6430.98	8212.90	6932.45	7847.36	7486.24
2	B	18268.01	20866.90	18738.99	21364.02	19360.46	21801.41	20151.42	22404.59	20715.54	22676.26
3	C	21088.14	24194.85	21608.14	24300.57	21988.64	24411.48	22739.28	24650.02	23436.61	25076.01

Reshaping Wide to Long

	country	varA2001	varB2001	varA2002	varB2002	varA2003	varB2003	varA2004	varB2004	varA2005	varB2005
1	A	NA	7702.89	NA	7288.48	8000.01	6430.98	8212.90	6932.45	7847.36	7486.24
2	B	18268.01	20866.90	18738.99	21364.02	19360.46	21801.41	20151.42	22404.59	20715.54	22676.26
3	C	21088.14	24194.85	21608.14	24300.57	21988.64	24411.48	22739.28	24650.02	23436.61	25076.01

```
gdp = reshape(data = gdp.wide,
              idvar = "country",
              varying = c("varA2001", "varB2001",
                          "varA2002", "varB2002",
                          "varA2003", "varB2003",
                          "varA2004", "varB2004",
                          "varA2005", "varB2005"),
              sep = "",
              timevar = "year",
              times = c(2001,2002,2003,2004,2005),
              new.row.names= 1:10000,
              direction = "long")

# Note on sep = "". You can use it to distinguish
# the variables to be reshaped from the numeric sequence.
# The option above will separate the common stems
# when it finds the first number.

# Sorting by country and year (optional)

gdp = gdp[order(gdp$country,gdp$year),]

# Removing row.names (optional)

row.names(gdp) = NULL
```

	country	year	varA	varB
1	A	2001	NA	7702.89
2	B	2001	18268.01	20866.90
3	C	2001	21088.14	24194.85
4	A	2002	NA	7288.48
5	B	2002	18738.99	21364.02
6	C	2002	21608.14	24300.57
7	A	2003	8000.01	6430.98
8	B	2003	19360.46	21801.41
9	C	2003	21988.64	24411.48
10	A	2004	8212.90	6932.45
11	B	2004	20151.42	22404.59
12	C	2004	22739.28	24650.02
13	A	2005	7847.36	7486.24
14	B	2005	20715.54	22676.26
15	C	2005	23436.61	25076.01

Reshaping Long to Wide

	country	year	varA	varB
1	A	2001	NA	7702.89
2	B	2001	18268.01	20866.90
3	C	2001	21088.14	24194.85
4	A	2002	NA	7288.48
5	B	2002	18738.99	21364.02
6	C	2002	21608.14	24300.57
7	A	2003	8000.01	6430.98
8	B	2003	19360.46	21801.41
9	C	2003	21988.64	24411.48
10	A	2004	8212.90	6932.45
11	B	2004	20151.42	22404.59
12	C	2004	22739.28	24650.02
13	A	2005	7847.36	7486.24
14	B	2005	20715.54	22676.26
15	C	2005	23436.61	25076.01

```
gdp2 =  reshape(data = gdp,
                idvar = "country",
                v.names = c("varA","varB"),
                timevar = "year",
                direction = "wide")

# Notice the dot between variable name and year
```

	country	varA.2001	varB.2001	varA.2002	varB.2002	varA.2003	varB.2003	varA.2004	varB.2004	varA.2005	varB.2005
1	A	NA	7702.89	NA	7288.48	8000.01	6430.98	8212.90	6932.45	7847.36	7486.24
2	B	18268.01	20866.90	18738.99	21364.02	19360.46	21801.41	20151.42	22404.59	20715.54	22676.26
3	C	21088.14	24194.85	21608.14	24300.57	21988.64	24411.48	22739.28	24650.02	23436.61	25076.01

References

John Fox 's site: http://socserv.mcmaster.ca/jfox/

Quick-R http://www.statmethods.net/

UCLA http://www.ats.ucla.edu/stat/R/