



# Reshape data using R/RStudio

(v. 2.0)

*Oscar Torres-Reyna*

*otorres@princeton.edu*



January 2011

<http://dss.princeton.edu/training/>

# 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

	year	country	gdppc
1	1990	Algeria	2544.48
2	1990	Argentina	3968.77
3	1990	Australia	25000.84
4	1990	Austria	28040.43
5	1990	Azerbaijan	1668.86
6	1990	Bahamas, The	22387.35
7	1990	Belgium	28076.39
8	1990	Belize	2570.91
9	1990	Bolivia	833.86
10	1990	Bosnia and Herzegovina	NA
11	1990	Botswana	3730.68
12	1990	Brazil	3999.43
13	1990	Bulgaria	2863.55
14	1990	Cambodia	NA
15	1990	Canada	27871.86
16	1990	Chile	4121.34
17	1990	Colombia	2832.47
18	1990	Costa Rica	3188.18
19	1990	Croatia	NA
20	1990	Cuba	3635.51
21	1990	Cyprus	16625.71
22	1990	Czech Republic	9775.77
23	1990	Denmark	36443.83
24	1990	Dominican Republic	2187.69
25	1990	Ecuador	2647.70
26	1990	Egypt, Arab Rep.	879.16
27	1990	El Salvador	1815.40
28	1990	Estonia	NA
29	1990	Finland	28122.87
30	1990	France	27801.05
31	1990	Germany	27900.87
32	1990	Greece	15384.04
33	1990	Guatemala	1761.60

```
# Data from World Bank (WDI) 1990-2013
```

```
gdppc = read.csv("http://www.princeton.edu/~otorres/GDPpc.csv",  
               header=TRUE,  
               stringsAsFactors = FALSE)
```

```
View(gdppc)
```

```
gdppc.wide = reshape(data = gdppc,  
                    idvar = "year",  
                    v.names = c("gdppc"),  
                    timevar = "country",  
                    direction = "wide")
```

```
View(gdppc.wide)
```

(see next page for a sample of the resulting data)



# Reshaping Long to Wide

	year	gdppc.Algeria	gdppc.Argentina	gdppc.Australia	gdppc.Austria	gdppc.Azerbaijan	gdppc.Bahamas, The	gdppc.Belgium	gdppc.Belize	gdppc.Bolivia
1	1990	2544.48	3968.77	25000.84	28040.43	1668.86	22387.35	28076.39	2570.91	833.86
96	1991	2452.81	4410.72	24653.75	28717.32	1631.65	21059.76	28485.00	2784.13	857.25
191	1992	2438.79	4871.76	24433.78	28997.63	1243.91	19873.91	28803.91	3062.79	850.99
286	1993	2334.80	5092.51	25196.52	28911.00	942.14	19566.83	28415.64	3197.19	866.71
381	1994	2265.89	5321.19	25937.92	29491.75	746.38	19837.96	29242.34	3140.07	886.37
476	1995	2306.85	5105.40	26583.03	30232.26	650.77	20393.85	29877.06	3089.84	906.99
571	1996	2358.81	5321.73	27271.06	30936.19	652.61	20983.35	30243.48	3054.11	925.74
666	1997	2345.44	5684.59	28031.52	31614.63	683.83	21179.63	31297.29	3073.96	950.71
761	1998	2427.07	5835.03	28970.25	32775.43	745.11	21942.87	31832.96	3095.46	977.49
856	1999	2468.23	5574.92	30074.07	33869.43	793.26	23228.80	32884.39	3268.83	961.35
951	2000	2487.29	5472.71	30869.66	35027.30	874.10	23841.87	34008.54	3590.88	965.41
1046	2001	2567.15	5179.45	31042.31	35192.65	953.22	24044.25	34165.49	3669.30	961.98
1141	2002	2675.73	4571.74	31852.06	35613.07	1046.42	24212.39	34475.14	3756.49	966.46
1236	2003	2830.97	4930.81	32430.05	35746.89	1154.84	23407.40	34608.11	4002.34	973.57
1331	2004	2912.29	5328.56	33389.17	36445.77	1261.54	23117.76	35586.93	4082.59	995.35
1426	2005	3038.75	5767.66	34011.74	37067.32	1578.37	23416.90	36011.47	4097.53	1020.79
1521	2006	3041.48	6195.38	34513.68	38237.95	2099.71	23538.91	36728.68	4181.37	1051.47
1616	2007	3092.04	6623.11	35591.49	39501.45	2596.06	23430.72	37511.18	4124.73	1081.39
1711	2008	3098.41	7050.84	36175.26	39895.11	2815.97	22472.88	37582.66	4177.56	1129.48
1806	2009	3091.06	7478.57	36054.88	38240.33	3017.66	21160.86	36237.39	4085.84	1148.79
1901	2010	3143.63	7906.30	36195.06	38803.24	3126.72	21013.90	36664.21	4138.97	1176.97
1996	2011	3171.74	8334.02	36495.04	39825.14	3088.28	21022.45	36876.92	4092.12	1217.75
2091	2012	3215.21	8761.75	37175.44	40058.38	3114.56	21083.65	36557.51	4154.33	1259.81
2186	2013	3240.83	9189.48	37492.85	40008.28	3252.83	21144.85	36410.53	4084.12	1323.12

# References

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

*Quick-R* <http://www.statmethods.net/>

*UCLA* <http://www.ats.ucla.edu/stat/R/>

*World Development Indicators (World Bank)*

<http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>