SAS Learning Module
How to reshape data wide to long using proc transpose

(http://www.ats.ucla.edu/stat/sas/modules/wtol_transpose.htm)

1. Transposing one group of variables

For a data set in wide format such as the one below, we can reshape it into long format using **proc transpose**. From the first output of **proc print**, we see that the data now is in long format except that we don't have a numeric variable indicating year; instead; we have a character variable that has information on year in it. So we have to do a data step to extract the information on year. The second output of **proc print** shows that our data step after the **proc transpose** has successfully created a numeric variable **year** and has rename the variable COL1 to **faminc**.

```
data wide1;
input famid faminc96 faminc97 faminc98 ;
cards;
1 40000 40500 41000  2 45000 45400 45800  3 75000 76000 77000  ;
run;
proc transpose data=wide1 out=long1;
by famid;
run;
proc print data=long1;
run;
```

```
Obs    famid   _NAME_      COL1
1       1      faminc96    40000
2       1      faminc97    40500
3       1      faminc98    41000
4       2      faminc96    45000
5       2      faminc97    45400
6       2      faminc98    45800
7       3      faminc96    75000
8       3      faminc97    76000
9       3      faminc98    77000
```

```
data long1;
set long1 (rename=(col1=faminc));
year=input(substr(_name_, 7), 5.);
drop _name_; run;
proc print data=long1;
run;
```

```
Obs    famid    faminc    year
1       1       40000     96
2       1       40500     97
3       1       41000     98
4       2       45000     96
5       2       45400     97
6       2       45800     98
7       3       75000     96
8       3       76000     97
9       3       77000     98
```

2. Transposing two groups of variables

In the following data set we have two groups of variables that need to be transposed. The first group is family income across years and the second group is the spending across year. A simple approach here is to transpose one group of variables at a time and then merge them back together. In the data step where we merge the transposed data sets, we also create a numeric variable year based on the SAS automatic variable **_NAME_** from the second transposed data set.

```
data wide2 ;
input famid faminc96 faminc97 faminc98 spend96 spend97 spend98 ;
cards ;
1 40000 40500 41000 38000 39000 40000
```

```
2 45000 45400 45800 42000 43000 44000
3 75000 76000 77000 70000 71000 72000;
run ;

proc transpose data=wide2 out=longf prefix=faminc ;
by famid; var faminc96-faminc98;
run;
proc transpose data=wide2 out=longs prefix=spend ;
by famid; var spend96-spend98;
run;
data long2;
merge
longf (rename=(faminc1=faminc) drop=_name_)
longs (rename=(spend1=spend));
by famid;
year=input(substr(_name_, 6), 5.);
drop _name_;
run;
proc print data=long2;
run;
```

Obs	famid	faminc	spend	year
1	1	40000	38000	96
2	1	40500	39000	97
3	1	41000	40000	98
4	2	45000	42000	96
5	2	45400	43000	97
6	2	45800	44000	98
7	3	75000	70000	96
8	3	76000	71000	97
9	3	77000	72000	98

3. A more realistic example

```
data wide3;    input id inc90 inc91 inc92 inc93 inc94 inc95 ;
cards;
1  66483 69146 74643 79783 81710 86143  2  17510 17947 19484 20979 21268 22998
3  57947 62964 68717 70957 75198 75722  4  64831 71060 71918 72514 73100 74379
5  18904 19949 21335 22237 23829 23913  6  32057 34770 35834 37387 40899 42372
7  60551 64869 67983 70498 71253 75177  8  16553 18189 18349 19815 21739 22980
9  32611 33465 35961 36416 37183 40627  10 61379 66002 67936 70513 74405 76009
11 24065 24229 25709 26121 26617 28142  12 32975 36185 37601 41336 43399 43670
13 69548 71341 72455 76552 80538 85330  14 50274 53349 55900 59375 61216 63911
15 72011 73334 76248 77724 78638 80582  16 18911 20046 21343 21630 22330 23081
17 68841 75410 80806 81327 81571 86499  18 28099 30716 32986 36097 39124 39866
19 17302 18778 18872 19884 20665 21855  20 16291 16674 16770 17182 17979 18917
21 43244 46545 47633 50744 54734 59075  22 56393 59120 60801 61404 63111 69278
23 47347 49571 50101 51345 56463 56927  24 16076 17217 17296 17900 18171 18366
25 65906 69679 76131 77676 81980 85426  26 58586 61188 66542 69267 71063 74549
27 61674 66584 69185 75193 78647 81898  28 31673 31883 32774 34485 36929 39751
29 63412 67593 69911 73092 80105 81840  30 27684 28439 30861 31406 32960 35530
31 71873 76449 80848 88691 94149 97431  32 62177 63812 64235 65703 69985 71136
33 37684 38258 39208 39489 39745 41236  34 64013 66398 71877 75610 76395 79644
35 16011 16847 17746 19123 19183 19996  36 49215 52195 52343 56365 58752 59354
37 15774 16643 17605 18781 18996 19685  38 29106 31693 31852 34505 35806 36179
39 25147 26923 28785 30987 34036 34106  40 71978 79144 80453 86580 95164 96155
41 46166 47579 49455 53849 56630 57473  42 55810 59443 65291 66065 69009 74365
43 49642 50603 53917 54858 58470 59767  44 21348 22361 23412 24038 24774 25828
45 44361 48720 51356 54927 56670 58800  46 56509 60517 61532 65077 69594 73089
47 39097 40293 43237 44809 48782 53091  48 18685 19405 20165 20316 22197 23557
49 73103 76243 76778 82734 86279 86784  50 48129 49267 53799 58768 63011 66461
;
run ;
```

```
proc transpose data=wide3 out=long3;
by id; run;  data long3;
set long3 (rename=(col1=inc));
year=input(substr(_name_, 4), 5.);
drop _name_;
run;

proc print data=long3 (obs=20);
run;
```

Obs	id	inc	year
1	1	66483	90
2	1	69146	91
3	1	74643	92
4	1	79783	93
5	1	81710	94
6	1	86143	95
7	2	17510	90
8	2	17947	91
9	2	19484	92
10	2	20979	93
11	2	21268	94
12	2	22998	95
13	3	57947	90
14	3	62964	91
15	3	68717	92
16	3	70957	93
17	3	75198	94
18	3	75722	95
19	4	64831	90
20	4	71060	91

4. Reshape wide to long with a character variable

In the following data set we have three groups of variables that needs to be transposed. One of the groups is the indicator of debt across years. The approach is the same with either numeric variables or character variables. Since there are three groups of variables, we need to use **proc transpose** three times, one for each group. Then we merge them back together. In the data step where we merge the transposed data files together, we also create a numeric variable for year and rename each of the variables properly. The variable **year** is created based on the SAS automatic variable **_NAME_** from the last transposed data set.

```
data wide4;
input famid faminc96 faminc97 faminc98 spend96 spend97 spend98 debt96 $ debt97 $
debt98 $ ;
cards;
1 40000 40500 41000 38000 39000 40000 yes yes no
2 45000 45400 45800 42000 43000 44000 yes no  no
3 75000 76000 77000 70000 71000 72000 no  no  no
;
run ;
```

```
proc transpose data=wide4 out=longf prefix=faminc;
by famid;
var faminc96-faminc98;
run;

proc transpose data=wide4 out=longs prefix=spend;
by famid;
var spend96-spend98;
run;

proc transpose data=wide4 out=longd prefix=debt;
by famid;
var debt96-debt98;
run;
data long4;
merge longf (rename=(faminc1=faminc) drop=_name_)
longs (rename=(spend1=spend) drop=_name_)
longd (rename=(debt1=debt));
by famid;
year=input(substr(_name_, 5), 5.);
drop _name_;
run;
proc print data=long4;
run;
```

Obs	famid	faminc	spend	debt	year
1	1	40000	38000	yes	96
2	1	40500	39000	yes	97
3	1	41000	40000	no	98
4	2	45000	42000	yes	96
5	2	45400	43000	no	97
6	2	45800	44000	no	98
7	3	75000	70000	no	96
8	3	76000	71000	no	97
9	3	77000	72000	no	98