

```
log using "k:\debug\roster_good.log", replace

set more 1

/*****
* This command file identify the ID and gender of the respondent's spouse/partner
* The command file was created on April 12, 2015

* The original data, roster.dta, has 7,517 household and 30 variables. The data were in wide format.
* the final data, roster_good.dta, were in long format and had 112,260 observations and 13 variables

* The command file accomplished the following things
(1) Check if the original data had empty records
(2) check if the original data had duplicate records
(3) Reshape the data from the wide format to the long format
(4) Remove 33 families where multiple families have the same family members as their spouses or partners
(5) Generate five variables to identify the spouse or partner of a family member
    1) reid_rs - ID for the respondent, respondent report, Roster
    2) resx_rs - gender for the respondent, respondent report, Roster
    3) spid_rs - ID for the spouses, respondent report, Roster
    4) spsx_rs - gender for the respondent, spouse/partner report, Roster
    5) spty_rs - union type, spouse/partner report, Roster
    6) check_inconsistent_report on the different-sex unions
    7) check_inconsistent_report on the different-sex unions
*****/

use "k:\debug\roster.dta", clear

*****/
* Look at the data
* The data has 7,517 observation and 30 variable
*****/

des

*****/
* Check if there were empty records
*****/

count if caseid ==.

*****/
* check if there were duplicate IDs
*****/

duplicates report caseid

*****/
* Generate a new relation variable for the reference person, so I could reshape the data later
*****/

gen q112_1 = .

*****/
* Reshape the data
*****/

reshape long q109_ q112_ , i(caseid) j(person)

*****/
* Add variable and value label
*****/

label variable person "personal ID within the household"

rename q109_gender
label variable gender "Gender"

rename q112_relation
label variable relation "the relation to the reference person"
```

label value relation relation

```
*****  
* Possible problem: Two or more family members have the same family member as their spouses  
*****
```

```
*****  
* Extract the relationship based on the household roster  
* The union type was determined by how family member is related with the reference person  
*****
```

```
*****  
* Theoretically, reference person should be noted in the spouse/partner relations once.
```

```
* 33 families where reference person were in the union with multiple family members, creating a problem identifying spouse  
*****
```

```
gen in_union = 1 if inlist(relation,1,2,3,4)
```

```
sort caseid  
by caseid: gen i_in_union = sum(in_union)  
by caseid: egen s_in_union = sum(in_union)
```

```
tab1 s_in_union, mis
```

```
tab2 person s_in_union if s_in_union >= 2, mis
```

```
list caseid person gender relation in_union i_in_union s_in_union if s_in_union >= 2, sepby(caseid) nol
```

```
*****  
* After dropping 33 families, the data file had 7,484 families  
* Of these families, 4,727 had spouses or partners  
*****  
drop if s_in_union >= 2
```

```
tab1 s_in_union, mis
```

```
tab2 person s_in_union , mis
```

```
*****  
* extract spousal information from the spouse's perspective  
*****
```

```
gen reid_rs_temp = 1 if person ==1  
gen resx_rs_temp = gender if person ==1  
  
gen spid_rs_temp = person if inlist(relation,1,2,3,4)  
gen spsx_rs_temp = gender if inlist(relation,1,2,3,4)  
gen spty_rs_temp = relation if inlist(relation,1,2,3,4)
```

```
*****  
* Expand the data  
*****
```

```
sort caseid
```

```
by caseid: egen m_reid_rs_temp = max(reid_rs_temp)  
by caseid: egen m_resx_rs_temp = max(resx_rs_temp)  
by caseid: egen m_spid_rs_temp = max(spid_rs_temp)  
by caseid: egen m_spsx_rs_temp = max(spsx_rs_temp)  
by caseid: egen m_spty_rs_temp = max(spty_rs_temp)
```

```
*****  
* Generate variables  
*****
```

```
gen reid_rs = .  
gen resx_rs = .  
gen spid_rs = .  
gen spsx_rs = .  
gen spty_rs = .
```

```
label variable reid_rs "ID for the respondent, respondent report, Roster"
```

```
label variable resx_rs "gender for the respondent, respondent repor, Roster"
```

```
label variable spid_rs "ID for the spouses, respondent report, Roster"
```

```
label variable spsx_rs "gender for the respondent, spouse/partner report, Roster"
```

```
label variable spty_rs "union type, spouse/partner report, Roster"
```

```
label define gender -1 "Refused" ///  
1 "Male" ///  
2 "Female "
```

```
label value resx_rs gender  
label value spsx_rs gender  
label value spty_rs relation
```

```
replace reid_rs = m_reid_rs if person ==1 & s_in_union ==1  
replace resx_rs = m_resx_rs if person ==1 & s_in_union ==1  
replace spid_rs = m_spid_rs if person ==1 & s_in_union ==1  
replace spsx_rs = m_spsx_rs if person ==1 & s_in_union ==1  
replace spty_rs = m_spty_rs if person ==1 & s_in_union ==1
```

```
replace reid_rs = m_spid_rs if person ~1 & m_spid_rs == person  
replace resx_rs = m_spsx_rs if person ~1 & m_spid_rs == person  
replace spty_rs = m_spty_rs if person ~1 & m_spid_rs == person  
replace spid_rs = m_reid_rs if person ~1 & m_spid_rs == person  
replace spsx_rs = m_resx_rs if person ~1 & m_spid_rs == person
```

```
*****  
* Check the data  
*****
```

```
tab2 reid_rs spid_rs, mis
```

```
sum reid_rs resx_rs spid_rs spsx_rs spty_rs if s_in_union ==1  
sum reid_rs resx_rs spid_rs spsx_rs spty_rs if s_in_union ~1
```

```
*****  
* select variables  
*****
```

```
keep caseid person gender relation reid_rs resx_rs spid_rs spsx_rs spty_rs
```

```
save "k:\debug\roster2.dta", replace
```

```
*****  
* Check the accuracy in the different- or same-sex union  
*****  
use "k:\debug\roster2.dta", clear
```

```
tab1 spty_rs if person ==1, mis  
tab2 resx_rs spsx_rs if person ==1, mis
```

```
*****  
* For different-sex unions  
*****  
tab2 resx_rs spsx_rs if person ==1 & inlist(spty_rs,1,2), mis
```

```
*****  
* Generate the indicator for inconsistent report on the different-sex union  
*****  
gen diffsex= .  
replace diffsex = 1 if person ==1 & inlist(spty_rs,1,2)  
replace diffsex = 0 if person ==1 & inlist(spty_rs,1,2) & (resx_rs ==1 & spsx_rs ==2)  
replace diffsex = 0 if person ==1 & inlist(spty_rs,1,2) & (resx_rs ==2 & spsx_rs ==1)
```

```
label variable diffsex "indicator for problem with the report of different-sex union"

label define problem 0 "no problem" 1 "problem"
label value diffsex problem

*****
* check the accuracy of constructing the variables
*****
tab1 diffsex if person ==1 & inlist(spty_rs,1,2), mis

*****
* Expand the indicator to the whole family
*****
sort caseid
by caseid: egen m_diffsex = max(diffsex)
label value m_diffsex problem

tab1 m_diffsex if person ==1 & inlist(spty_rs,1,2), mis

*****
* List families with inconsistent reports on the different-sex unions
*****

list caseid- diffsex if m_diffsex ==1, sepby(caseid) nol

*****
* For the same-sex unions
*****

tab2 resx_rs spsx_rs if person ==1 & inlist(spty_rs,3,4), mis

*****
* Generate the indicator for inconsistent report on the same-sex union
*****
gen samesex= .
replace samesex = 1 if person ==1 & inlist(spty_rs,3,4)
replace samesex = 0 if person ==1 & inlist(spty_rs,3,4) & (resx_rs ==1 & spsx_rs ==1)
replace samesex = 0 if person ==1 & inlist(spty_rs,3,4) & (resx_rs ==2 & spsx_rs ==2)

label variable samesex "indicator for problem with the report of same-sex union"
label value samesex problem

*****
* check the accuracy of constructing the variables
*****
tab1 diffsex if person ==1 & inlist(spty_rs,3,4), mis

*****
* Expand the indicator to the whole family
*****
sort caseid
by caseid: egen m_samesex = max(samesex)
label value m_samesex problem

tab1 m_diffsex if person ==1 & inlist(spty_rs,3,4), mis

*****
* List families with inconsistent reports on the different-sex unions
*****

list caseid- diffsex if m_samesex ==1, sepby(caseid) nol

save "k:\debug\race_good.dta", replace

log close
```