

```

. clear

. macro drop _all

.
. /*
> Back to Fan's Stata4Econ or other repositories:
>
> - http://fanwangecon.github.io
>
> - http://fanwangecon.github.io/Stata4Econ
> - http://fanwangecon.github.io/R4Econ
> - http://fanwangecon.github.io/REconTools
> - http://fanwangecon.github.io/M4Econ
> - http://fanwangecon.github.io/Tex4Econ
> - http://fanwangecon.github.io/CodeDynaAsset/
> - http://fanwangecon.github.io/Math4Econ/
> - http://fanwangecon.github.io/Stat4Econ/
>
> 1. given a discrete variable
> 2. recode the discrete variable to reduce the number of categories, generate larger category ca
>
> Note there are several ingredients to consider here:
> 1. current variable name
> 2. new variable name
> 3. new variable label
> 4. new value labels
> 5. new note
> */

. ///--- Start log
> set more off

. capture log close _all

. cd "${root_log}"
C:\Users\fan\Documents\Dropbox (UH-ECON)\Project Emily Minority Survey\Code

. global st_link "/gen/replace/fs_recode"

. global curlogfile "~/Stata4Econ/${st_link}"

. global st_logname "stata_recode_discrete_subset"

. log using "${curlogfile}" , replace name($st_logname)
(note: file C:\Users\fan\Stata4Econ//gen/replace/fs_recode.smcl not found)



---


      name: stata_recode_discrete_subset
      log: C:\Users\fan\Stata4Econ//gen/replace/fs_recode.smcl
  log type: smcl
opened on: 17 Apr 2020, 21:53:30

. log on $st_logname
(log already on)

.
. ///-- Site Link: Fan's Project Reusable Stata Codes Table of Content
> di "https://fanwangecon.github.io/"
https://fanwangecon.github.io/

. di "https://fanwangecon.github.io/Stata4Econ/"
https://fanwangecon.github.io/Stata4Econ/

.
. ///-- File Title
> global filetitle "Stata Recode a Discrete Variable with Alternative Labels and Values Subgroups
.

```

. sysuse auto, clear  
 (1978 Automobile Data)

```
.
. //----- Recode Method 1
> //----- Recode Method 1a: recode
> capture drop turn_m5
```

```
. recode turn ///
> (min/35 = 1 "Turn <35") ///
> (36 = 2 "Turn = 36") ///
> (37 = 3 "Turn = 37") ///
> (38/45 = 4 "Turn 38 to 45") ///
> (46/max = 5 "Turn > 45") ///
> (else =. ) ///
> , gen(turn_m5)
(74 differences between turn and turn_m5)
```

. tab turn\_m5

RECODE of turn (Turn Circle (ft.) )	Freq.	Percent	Cum.
Turn <35	<b>16</b>	<b>21.62</b>	<b>21.62</b>
Turn = 36	<b>9</b>	<b>12.16</b>	<b>33.78</b>
Turn = 37	<b>4</b>	<b>5.41</b>	<b>39.19</b>
Turn 38 to 45	<b>39</b>	<b>52.70</b>	<b>91.89</b>
Turn > 45	<b>6</b>	<b>8.11</b>	<b>100.00</b>
Total	<b>74</b>	<b>100.00</b>	

```
.
. //----- Recode Method 1b: egen cut
> capture drop turn_m5_cut
```

```
. egen turn_m5_cut = cut(turn), at(31, 36, 37, 38, 46, 51) label
(1 missing value generated)
```

. tab turn\_m5\_cut

turn_m5_cut	Freq.	Percent	Cum.
31-	<b>16</b>	<b>21.92</b>	<b>21.92</b>
36-	<b>9</b>	<b>12.33</b>	<b>34.25</b>
37-	<b>4</b>	<b>5.48</b>	<b>39.73</b>
38-	<b>39</b>	<b>53.42</b>	<b>93.15</b>
46-	<b>5</b>	<b>6.85</b>	<b>100.00</b>
Total	<b>73</b>	<b>100.00</b>	

```
. capture drop turn_m7_cut
```

```
. egen turn_m7_cut = cut(turn), at(31(3)52) label
```

. tab turn\_m7\_cut

turn_m7_cut	Freq.	Percent	Cum.
31-	<b>4</b>	<b>5.41</b>	<b>5.41</b>
34-	<b>21</b>	<b>28.38</b>	<b>33.78</b>
37-	<b>8</b>	<b>10.81</b>	<b>44.59</b>
40-	<b>17</b>	<b>22.97</b>	<b>67.57</b>
43-	<b>18</b>	<b>24.32</b>	<b>91.89</b>
46-	<b>5</b>	<b>6.76</b>	<b>98.65</b>
49-	<b>1</b>	<b>1.35</b>	<b>100.00</b>
Total	<b>74</b>	<b>100.00</b>	

```
. ///--- Recode Method 1c: inrange and inlist
> capture drop turn_m5_alt
```

```
. clonevar turn_m5_alt = turn
```

```
. label variable turn_m5_alt "Recode using inlist and inrange"
```

```
. replace turn_m5_alt = 1 if inrange(turn, 31, 35)
(16 real changes made)
```

```
. replace turn_m5_alt = 2 if inlist(turn, 36)
(9 real changes made)
```

```
. replace turn_m5_alt = 3 if inlist(turn, 37)
(4 real changes made)
```

```
. replace turn_m5_alt = 4 if inrange(turn, 38, 45)
(39 real changes made)
```

```
. replace turn_m5_alt = 5 if inlist(turn, 46, 48, 51)
(6 real changes made)
```

```
. label define turn_m5_alt 1 "Turn <35" 2 "Turn = 36" 3 "Turn = 37" 4 "Turn 38 to 45" 5 "Turn > 45"
```

```
. label values turn_m5_alt turn_m5_alt
```

```
. tab turn_m5_alt
```

Recode using inlist and inrange	Freq.	Percent	Cum.
Turn <35	<b>16</b>	<b>21.62</b>	<b>21.62</b>
Turn = 36	<b>9</b>	<b>12.16</b>	<b>33.78</b>
Turn = 37	<b>4</b>	<b>5.41</b>	<b>39.19</b>
Turn 38 to 45	<b>39</b>	<b>52.70</b>	<b>91.89</b>
Turn > 45	<b>6</b>	<b>8.11</b>	<b>100.00</b>
Total	<b>74</b>	<b>100.00</b>	

```
. ///--- compare
> tab turn_m5 turn_m5_cut
```

RECODE of turn (Turn Circle (ft.) )	31-	36-	turn_m5_cut 37-	38-	46-	Total
Turn <35	<b>16</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>16</b>
Turn = 36	<b>0</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>9</b>
Turn = 37	<b>0</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>4</b>
Turn 38 to 45	<b>0</b>	<b>0</b>	<b>0</b>	<b>39</b>	<b>0</b>	<b>39</b>
Turn > 45	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>5</b>
Total	<b>16</b>	<b>9</b>	<b>4</b>	<b>39</b>	<b>5</b>	<b>73</b>

```
. tab turn_m5 turn_m5_alt
```

RECODE of turn (Turn Circle (ft.) )	Recode using inlist and inrange					Total
	Turn <35	Turn = 36	Turn = 37	Turn 38 to 45	Turn > 45	
Turn <35	<b>16</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>16</b>
Turn = 36	<b>0</b>	<b>9</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>9</b>
Turn = 37	<b>0</b>	<b>0</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>4</b>
Turn 38 to 45	<b>0</b>	<b>0</b>	<b>0</b>	<b>39</b>	<b>0</b>	<b>39</b>
Turn > 45	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>6</b>	<b>6</b>
Total	<b>16</b>	<b>9</b>	<b>4</b>	<b>39</b>	<b>6</b>	<b>74</b>

RECODE of turn (Turn Circle (ft.) )	31-	34-	37-	turn_m7_cut 40-	43-	46-	49-
Turn <35	4	12	0	0	0	0	0
Turn = 36	0	9	0	0	0	0	0
Turn = 37	0	0	4	0	0	0	0
Turn 38 to 45	0	0	4	17	18	0	0
Turn > 45	0	0	0	0	0	5	1
Total	4	21	8	17	18	5	1

```
.
. //----- Recode Method 2a: Recode based on single variable,
> /// slightly less typing, compose ingredients together
> //-----
> /*
> Define string using local strings to avoid some retyping.
> try to make variable label not longer than width limit.
> */

.
. //-- Set Variable Strings
. global svr_newv "trunk_new"

. global svr_oldv "trunk"

. global slb_labl "this is the new version of the trunk variable"

. global slb_note "we reset this variable be grouping values 5 to 10, 11 to 13, 14 "

. global slb_note "$slb_note to 18, 20 to 22, and 23 into subgroups. We did this "

. global slb_note "$slb_note test things out for resetting variables"

.
. //-- value resetting
. #delimit;
delimiter now ;
. global slb_valv "
> (min/4 = 1 "trunk <5")
> (5/10 = 2 "Turn = 36")
> (11/13 = 3 "Turn = 37")
> (14/18 = 4 "Turn 38 to 45")
> (20/22 = 5 "Turn > 45")
> (23 = 5 "Turn > 45")
> (else =. )
> ";

. #delimit cr
delimiter now cr

.
. //-- recode
. * generate
. capture drop $svr_newv

. recode $svr_oldv $slb_valv, gen($svr_newv)
(74 differences between trunk and trunk_new)

. label variable $svr_newv "$slb_labl"

. notes $svr_newv: $slb_note

. * summ
. d $svr_oldv $svr_newv, f
```

variable name	storage type	display format	value label	variable label
<b>trunk</b>	int	%8.0g		<b>Trunk space (cu. ft.)</b>
<b>trunk_new</b>	int	%13.0g	trunk_new	* <b>this is the new version of the trunk variable</b>

. notes \$svr\_oldv \$svr\_newv

trunk\_new:

1. we reset this variable be grouping values 5 to 10, 11 to 13, 14 to 18, 20 to 22, and 23 into

. summ \$svr\_oldv \$svr\_newv

Variable	Obs	Mean	Std. Dev.	Min	Max
trunk	74	13.75676	4.277404	5	23
trunk_new	74	3.418919	1.020432	2	5

. tab \$svr\_oldv \$svr\_newv

Trunk space (cu. ft.)	this is the new version of the trunk variable				Total
	Turn = 36	Turn = 37	Turn 38 t	Turn > 45	
5	1	0	0	0	1
6	1	0	0	0	1
7	3	0	0	0	3
8	5	0	0	0	5
9	4	0	0	0	4
10	5	0	0	0	5
11	0	8	0	0	8
12	0	3	0	0	3
13	0	4	0	0	4
14	0	0	4	0	4
15	0	0	5	0	5
16	0	0	12	0	12
17	0	0	8	0	8
18	0	0	1	0	1
20	0	0	0	6	6
21	0	0	0	2	2
22	0	0	0	1	1
23	0	0	0	1	1
Total	19	15	30	10	74

. tab \$svr\_newv

this is the new version of the trunk variable	Freq.	Percent	Cum.
Turn = 36	19	25.68	25.68
Turn = 37	15	20.27	45.95
Turn 38 to 45	30	40.54	86.49
Turn > 45	10	13.51	100.00
Total	74	100.00	

```
.
.
. //////////////////////////////////////////////////
> ///--- Recode Method 2b: same as method 2a, but do it for multiple variables loop loop
> //////////////////////////////////////////////////
> /*
> 1. Define string using local strings to avoid some retyping.
> 2. Summarize outputs iteration by iteration, verbose or not
> 3. Summarize outputs at the end overall
> 4. if new and old variables have the same name, understand we want to use the
>     same name, will relabel generate a new variable with the same variable name
>     and keep old variable as old_abc, where abc is the current var name
> */
. global svr_newv_all ""
```

```

. foreach it_var of numlist 1 2 3 {
2.
.     //-- Variable by Variable Naming Settings
.     if (`it_var' == 1) {
3.         //-- Set Variable Strings
.         global svr_newv "price_2m"
.         global svr_oldv "price"
4.         global slb_labl "price discretized 2 levels"
5.         global slb_note "reset the price variable into two groups, original variable
6.         global slb_note "$slb_note 74 observations with 74 unique values. "
7.
8.         //-- value resetting
.         #delimit;
.     delimiter now ;
.         global slb_valv "
>             (min/6000 = 1 "price <= 6000")
>             (6001/max = 2 "price > 6000")
>             (else =. )
>         ";
9.         #delimit cr
.     delimiter now cr
.
.         //-- states verbose show or not
.         global bl_verbose_print = 0
10.     }
11.     if (`it_var' == 2) {
12.         //-- Set Variable Strings
.         global svr_newv "price_3m"
13.         global svr_oldv "price"
14.         global slb_labl "price discretized 3 levels"
15.         global slb_note "reset the price variable into two groups, original variable
16.         global slb_note "$slb_note 74 observations with 74 unique values. "
17.
.         //-- value resetting
.         #delimit;
.     delimiter now ;
.         global slb_valv "
>             (min/5500 = 1 "price <= 5500")
>             (5501/8500 = 2 "5501 <= price <= 8500")
>             (8501/max = 3 "8501 <= price")
>             (else =. )
>         ";
18.         #delimit cr
.     delimiter now cr
.
.         //-- states verbose show or not
.         global bl_verbose_print = 0
19.     }
20.     if (`it_var' == 3) {
21.         //-- Set Variable Strings
.         * this is an example where I relabel and revalue names, but keep variable name
.         * auto keep an old version
.         global svr_newv "foreign"
22.         global svr_oldv "foreign"
23.         global slb_labl "is car domestic (relabelled, previous 1 is foreign now 0)"
24.         global slb_note "reseting the foreign variable previously 1 is foreign 0"
25.         global slb_note "$slb_note is domestic, now 1 is domestic 0 is foreign"
26.
.         //-- value resetting
.         #delimit;
.     delimiter now ;
.         global slb_valv "
>             (1 = 0 "foreign car")
>             (0 = 1 "domestic car")
>             (else =. )
>         ";
27.         #delimit cr
.     delimiter now cr
.

```

variable name	storage type	display format	value label	variable label	
price_2m	int	%13.0g	price_2m * price discretized 2 levels		
Variable	Obs	Mean	Std. Dev.	Min	Max
price	74	6165.257	2949.496	3291	15906
price_2m	74	1.310811	.4659848	1	2
price discretized 2 levels	Freq.	Percent	Cum.		
price <= 6000	51	68.92	68.92		
price > 6000	23	31.08	100.00		
Total	74	100.00			
	price	price_2m			
price	1.0000				
price_2m	0.8001	1.0000			
	0.0000				

variable name	storage type	display format	value label	variable label
price 3m	int	%21.0g	price 3m *	price discretized 3 levels

price	74	6165.257	2949.496	3291	15906
price_3m	74	1.581081	.7764824	1	3
price discretized 3 levels		Freq.	Percent	Cum.	
price <= 5500		44	59.46	59.46	
5501 <= price <= 8500		17	22.97	82.43	
8501 <= price		13	17.57	100.00	
Total	74	100.00			

	price	price_3m
price	1.0000	
price_3m	0.9085	1.0000
	0.0000	

Generate the 3th variable: Generates foreign based on foreign (74 differences between \_prev\_foreign and foreign)

variable name	storage type	display format	value label	variable label
---------------	--------------	----------------	-------------	----------------

**foreign** byte %12.0g foreign \* **is car domestic (relabeled, previous 1 is foreign now 0)**

Variable	Obs	Mean	Std. Dev.	Min	Max
_prev_fore~n	74	.2972973	.4601885	0	1
foreign	74	.7027027	.4601885	0	1
is car domestic (relabeled, previous 1 is foreign now 0)					
	Freq.	Percent	Cum.		
foreign car	22	29.73	29.73		
domestic car	52	70.27	100.00		
Total	74	100.00			

	_prev_~n	foreign
_prev_foreign	1.0000	
foreign	-1.0000	1.0000
	1.0000	

variable name	storage type	display format	value label	variable label
---------------	--------------	----------------	-------------	----------------

**\_prev\_foreign** byte %8.0g origin \* **Car type**  
**foreign** byte %12.0g foreign \* **is car domestic (relabeled, previous 1 is foreign now 0)**

**\_prev\_foreign:**  
1. "this variable \_prev\_foreign is replaced by foreign"

**foreign:**  
1. resetting the foreign variable previously 1 is foreign 0 is domestic, now 1 is domestic 0 is

	is car domestic (relabeled, previous 1 is foreign now 0)	
Car type	foreign c domestic	Total
Domestic	0	52
Foreign	22	0
Total	22	52



```
0 foreign car
1 domestic car
```

[illegible]

variable name	storage type	display format	value label	variable label
price_2m	int	%13.0g	price_2m *	price discretized 2 levels
price_3m	int	%21.0g	price_3m *	price discretized 3 levels
foreign	byte	%12.0g	foreign *	is car domestic (relabeled, previous 1 is foreign no

```
. summ $svr_newv_all
```

Variable	Obs	Mean	Std. Dev.	Min	Max
price_2m	74	1.310811	.4659848	1	2
price_3m	74	1.581081	.7764824	1	3
foreign	74	.7027027	.4601885	0	1

```
. pwcorr $svr newv all, sig
```

	price_2m	price_3m	foreign
price_2m	1.0000		
price_3m	0.8570 0.0000	1.0000	
foreign	-0.1381 0.2406	-0.1233 0.2953	1.0000

```
.
. ///--- End Log and to HTML
> log close _all
      name: stata_recode_discrete_subset
      log:  C:\Users\fan\Stata4Econ\gen\replace\fs_recode.smcl
      log type: smcl
      closed on: 17 Apr 2020, 21:53:30
```

```
. capture noisily {
.     log2html "${curlogfile}", replace title($filetitle (<a href="https://github.com/FanWang
> s://fanwangecon.github.io/">Fan</a> and <a href="https://fanwangecon.github.io/Stata4Econ">Stat
HTML log file ~/Stata4Econ//gen/replace/fs_recode.html created
. }
```

```

. ///--- to PDF
> capture noisily {
.   translator set Results2pdf logo off
.   translator set Results2pdf fontsize 10
.   translator set Results2pdf pagesize custom
.   translator set Results2pdf pagewidth 8.27
.   translator set Results2pdf pageheight 11.69
.   translator set Results2pdf lmargin 0.2
.   translator set Results2pdf rmargin 0.2
.   translator set Results2pdf tmargin 0.2
.   translator set Results2pdf bmargin 0.2
.   translate @Results "${curlogfile}.pdf", replace translator(Results2pdf)

```