```
. 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

. 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:  16 Apr 2020, 03:33:05

. 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

.
```

```
. ///--- Load Data
> set more off

. sysuse auto, clear
(1978 Automobile Data)

.
. ////////////////////////////////////////////////////////////////////////////
> ///--- Recode Method 1
> ////////////////////////////////////////////////////////////////////////////
> 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)

.
. ////////////////////////////////////////////////////////////////////////////
> ///--- Recode Method 2
> ////////////////////////////////////////////////////////////////////////////
> 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 > 4

. label values turn_m5_alt turn_m5_alt

.
. ////////////////////////////////////////////////////////////////////////////
> ///--- Recode Method 3: 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 reseting 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

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

. 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 int

. 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 |     this is the new version of the trunk
space (cu. |                  variable
      ft.) | Turn = 36  Turn = 37  Turn 38 t  Turn > 45 |     Total
-----------+--------------------------------------------+----------
         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 4: same as method 3, 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"
  4.                 global svr_oldv "price"
  5.                 global slb_labl "price discretized 2 levels"
  6.                 global slb_note "reset the price variable into two groups, original variable
  7.                 global slb_note "$slb_note 74 observations with 74 unique values. "
  8.
.                 //-- value resetting
.                 #delimit;
delimiter now ;
.                 global slb_valv "
>                         (min/6000 = 1 "price <= 6000")
>                         (6001/max = 2 "price >  6000")
>                         (else  =. )
>                   ";
  9.                  #delimit cr
delimiter now cr
.
.                 //-- states verbose show or not
.                 global bl_verbose_print = 0
 10.         }
 11.         if (`it_var' == 2) {
 12.                 //-- Set Variable Strings
.                 global svr_newv "price_3m"
 13.                 global svr_oldv "price"
 14.                 global slb_labl "price discretized 3 levels"
 15.                 global slb_note "reset the price variable into two groups, original variable
 16.                 global slb_note "$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 (relabled, 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
.
.                 //-- states verbose show or not
.                 global bl_verbose_print = 1
 28.          }
 29.
.          //-- recode
.          di "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
 30.          di "Generate the `it_var'th variable: Generates $svr_newv based on $svr_oldv"
 31.          di "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
 32.          * generate
.          global svr_oldv_use "${svr_oldv}"
 33.          if ("$svr_newv" == "$svr_oldv") {
 34.                * allows for relabeling the same variable keeping name
.                 global svr_oldv_use "_prev_${svr_oldv}"
 35.                clonevar _prev_${svr_oldv} = $svr_oldv
 36.                notes $svr_oldv_use: "this variable $svr_oldv_use is replaced by $svr_newv"
 37.          }
 38.          capture drop $svr_newv
 39.          recode $svr_oldv_use $slb_valv, gen($svr_newv)
 40.          label variable $svr_newv "$slb_labl"
 41.          notes $svr_newv: $slb_note
 42.
.          //-- summarize
.          d $svr_newv, f
 43.          summ $svr_oldv_use $svr_newv
 44.          tab $svr_newv
 45.          pwcorr $svr_oldv_use $svr_newv, sig
 46.          if ($bl_verbose_print) {
 47.                d $svr_oldv_use $svr_newv, f
 48.                notes $svr_oldv_use $svr_newv
 49.                tab $svr_oldv_use $svr_newv
 50.                label list $svr_newv
 51.          }
 52.
.          //-- Store all strings for easier later retrieval
.          global svr_newv_all `"$svr_newv_all $svr_newv"'
 53.
. }
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Generate the 1th variable: Generates price_2m based on price
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
(74 differences between price and price_2m)

              storage   display    value
variable name   type    format     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** **0.0000**	**1.0000**

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Generate the 2th variable: Generates price_3m based on price
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
(74 differences between price and price_3m)

variable name	storage type	display format	value label	variable label
price_3m	int	%21.0g	price_3m *	**price discretized 3 levels**

Variable	Obs	Mean	Std. Dev.	Min	Max
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** **0.0000**	**1.0000**

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Generate the 3th variable: Generates foreign based on foreign
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
(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 (relabled, previous 1 is foreign no**

Variable	Obs	Mean	Std. Dev.	Min	Max
_prev_fore~n	**74**	**.2972973**	**.4601885**	**0**	**1**
foreign	**74**	**.7027027**	**.4601885**	**0**	**1**

is car domestic (relabled, 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_fore~n	**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 (relabled, previous 1 is foreign no**

_prev_foreign:
1. "this variable _prev_foreign is replaced by foreign"

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

Car type	is car domestic (relabled, previous 1 is foreign now 0) foreign c	domestic	Total
Domestic	**0**	**52**	**52**
Foreign	**22**	**0**	**22**
Total	**22**	**52**	**74**

foreign:
0 foreign car
1 domestic car

```
. //-- recode
. di "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

. di "We just finished Generating `it_var' Variables, here is their joint summary"
We just finished Generating  Variables, here is their joint summary

. di "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

. d $svr_newv_all, f
```

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 (relabled, 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**	**1.0000**	
	0.0000		
foreign	**-0.1381**	**-0.1233**	**1.0000**
	0.2406	**0.2953**	

```
.
. ///--- 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:  16 Apr 2020, 03:33:05
-------------------------------------------------------------------------------

. capture noisily {
.         log2html "${curlogfile}", replace title($filetitle (<a href="https://github.com/FanWang
> .github.io/">Fan</a> and <a href="https://fanwangecon.github.io/Stata4Econ">Stata4Econ</a>))

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)
```