This command file was created for Stata commands used in CFDR workshop on "Data Management in Stata" on Jan. 25, 2021.

The command file includes five groups of Stata commands, including
(1) general commands;
(2) commands for Managing data sets,
(3) commands for checking variables,
(4) commands for modifying variables,
(5) Generating new variables.

The final command and log files (i.e., data management.do and data management.log) are saved in the folder (c:\temp\workshop\data management).

No new data files were created.

**************************************************************
***************
* The findit command
***************

findit mim //the mim is a user-written ado file for analyzed imputed data
findit esttab //the esttab command is for creating Publication-Quality Tables in Stata

**************************************************************
***************
* Create a log file
***************
*log using "d:\temp\workshop\data management\data management.log", replace
log using "d:\temp\data management.log", replace

**************************************************************
***************
* Test some Stata commands
***************
clear
set maxvar 10000 // set the maxium number of variables to be read in Stata

*set more on //Tell Stata to pause for --more-- messages
set more off //Tell Stata not to pause for --more-- messages

***************
* Open the data file
***************
use d:\temp\auto.dta, clear

**************************************************************
***************
* The list command
***************
list
list if foreign ==1
list if inlist(rep78,1,2,3)
list if inrange(rep78,1,3)

**************************************************************
***************
* Commands for managing dta sets
***************
use d:\temp\auto.dta, clear
des
sum

**************************************************************
***************
* Check duplicate records
***************
duplicates report foreign rep78
duplicates tag foreign rep78, generate(same)

label variable same "respondents have the same values of foreign and rep78"

by foreign rep78, sort: gen n = _n
by foreign rep78, sort: gen N = _N
list foreign rep78 same n N, sepby(foreign rep78)

********************
* Merge data
********************

********************
* 1:1 match merge
********************

webuse autosize, clear
list
webuse autoexpense, clear
list
webuse autosize, clear
merge 1:1 make using http://www.stata-press.com/data/r15/autoexpense
list

********************
* m:1 match merge
********************

webuse dollars, clear
list
webuse sforce, clear
list
merge m:1 region using http://www.stata-press.com/data/r15/dollars
list

********************
* 1:m match merge
********************

webuse sforce, clear
list
webuse dollars, clear
list
merge 1:m region using http://www.stata-press.com/data/r15/sforce
list

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

webuse reshape1, clear
list
reshape long inc ue, i(id) j(year)
list, sepby(id)
reshape wide inc ue, i(id) j(year)
list, sepby(id)

********************
* Tabulations of frequencies
********************

webuse auto.dta, clear
sum
tab1 rep78
tab1 rep78, mis

********************
* Commands for Modifying Existing Variables
********************
webuse auto.dta, clear

*******************
* rename variable
*******************
rename *, upper
rename *, lower
rename make make2

des

*******************
* recast variables
*******************
recast str30 make

des make

*****************
* recode variables
*****************
recode rep78 (1=7) (2/5=8) (.=9), gen(rep78_2)

*************************************************************************
* Create and drop variable/value labels
*************************************************************************
label variable rep78_2 "new rep78"
label define rep78_2 "1 repair" 8 "2-5 repairs" 9 "missing values"
label value rep78_2 rep78_2
label drop rep78_2

des

tab1 rep78, mis

des

tab1 rep78_2, mis

*************************************************************************
* replace the values of a variable
*************************************************************************
replace rep78_2 = 1 if rep78 ==1
replace rep78_2 = 2 if inrange(rep78,2,5)
replace rep78_2 = . if rep78 ==.

des

tab2 rep78 rep78_2, mis

*************************************************************************
* Convert numeric variables to string variables and vice versa
*************************************************************************
webuse auto.dta, clear
tostring rep78, gen(c_rep78)
label variable c_rep78 "Rep78 as a categorical variable"

des

tab1 rep78, mis

tab1 c_rep78, mis

*************************************************************************
* code missing values
*************************************************************************
tab1 rep78, mis
mvencode rep78 , mv(999)
tab1 rep78, mis
mvdecode rep78 , mv(999)
tab1 rep78, mis

* clone a variable
clonevar make2 = make
des make make2

* Generating different types of variables
gen rep78_3 = rep78

* Quickly generate dummy variables
tab1 rep78, gen(rep78_c)

* generate a new variable using a substring from a string variable
gen newmake=substr(make,1,8)
list newmake make in 1/20

* Create variable containing the running sum of an existing variable
by foreign, sort: gen sum_rep78 =sum(rep78)
list foreign rep78 sum_rep78, sepby(foreign)

* Use the egen command
egen missing =rowmiss(price mpg rep78 foreign)
egen value1  =anycount(price mpg rep78 foreign), value(1)
list price mpg rep78 foreign missing value1

egen value1_2 =anyvalue(rep78), value(1,2)
tab2 rep78 value1_2,mis

* Find the maximum and minimum values among observations
by foreign, sort: egen max_rep78 =max(rep78)
by foreign, sort: egen min_rep78 =min(rep78)
list foreign rep78 max_rep78 min_rep78

* Use the egen command to quickly recode continuous variables into groups
egen new_price = cut(price), at (3000,6000,9000,12000,15000, 18000)
tab1 new_price, mis

* Combine variables into a string varaible
egen make_rep78 = concat( make rep78 ) , punct("_")
list make_rep78 make make_rep78

log close