MATT’S AND SHUAI’S TOP STATA TIPS

#1 GET YOUR DO FILE OFF TO A GOOD START

- Put "clear all" at the top of your do file.
  - Stata needs a "clean state" to bring in a new dataset
  - Prevents you from accidently saving over your data
- set more off: Makes it so Stata's output will continuously print
- close log _all: Will close any/all open log files. *You don't have to specify the name!*

```
*********
// Affordable Housing
      // Analysis
      // 9-10-2020
clear all
set more off, perm
close log _all
```

QUICK TIPS #2&3

- Be careful with blanks in your working directory and file names, put quotation mark around your working directory if there are blanks

```
use X:Stats/Homework 1/ Data, clear
```

☹

```
use “X:Stats/Homework 1/ Data”, clear
```

☺

- Be careful with the types of the variable, always put quotation marks for string variable. String variables contain letters, but sometimes a dataset will also read in numbers as string.

```
drop if state != Pennsylvania
```

☹

```
drop if state != “Pennsylvania”
```

☺

#4 TOSTRING AND DESTRING

- Sometimes you download a dataset and a numerical variable is read in as a string. Stata can’t perform most tasks on string variables. However you can easily put the variable in the right type using the destring command

```
destring year, replace
```

```
destring fips_id, gen (fips_num)
```

^^^ Must use either the replace or generate option

```
tostring fips_num, replace
```

#5 USE THE FRE COMMAND

- The fre command shows a tab of a variable, both with and without labels.

- Helps you figure out how recode/relabel a variable.

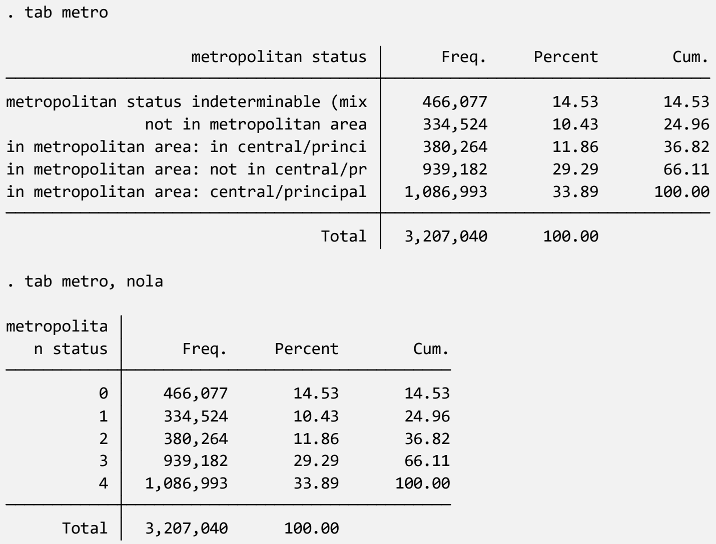
```
ssc install fre

checking fre consistency and verifying not already
> installed...

installing into c:\ado\plus\...

installation complete.
```

WHICH ONE IS EASIER TO UNDERSTAND?

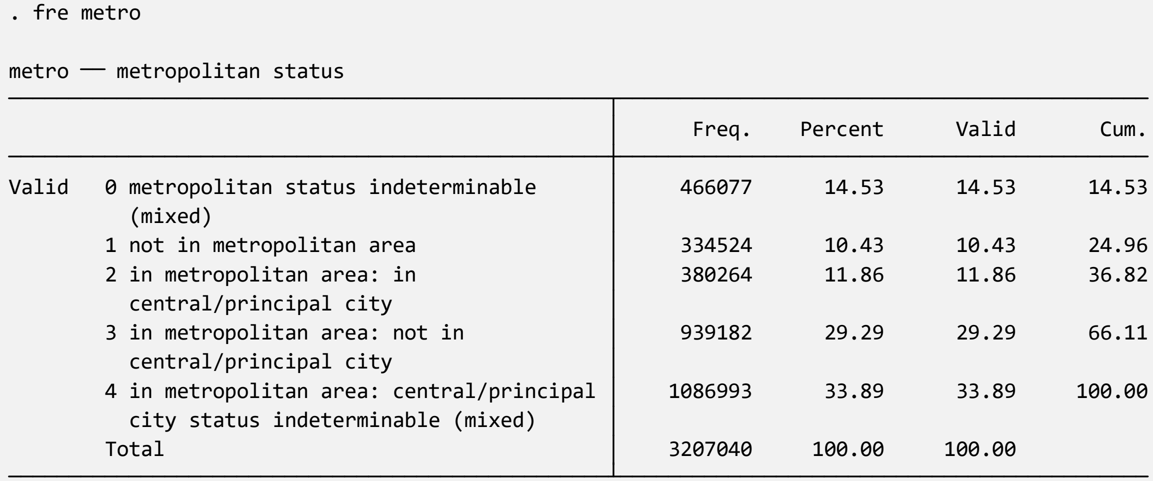


```
. tab metro

                  metropolitan status |      Freq.     Percent        Cum.
--------------------------------------+-----------------------------------
metropolitan status indeterminable (mix |    466,077       14.53       14.53
             not in metropolitan area |    334,524       10.43       24.96
in metropolitan area: in central/princi |    380,264       11.86       36.82
in metropolitan area: not in central/pr |    939,182       29.29       66.11
in metropolitan area: central/principal |  1,086,993       33.89      100.00
--------------------------------------+-----------------------------------
                                Total |  3,207,040      100.00

. tab metro, nola

metropolita |
   n status |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |    466,077       14.53       14.53
          1 |    334,524       10.43       24.96
          2 |    380,264       11.86       36.82
          3 |    939,182       29.29       66.11
          4 |  1,086,993       33.89      100.00
------------+-----------------------------------
      Total |  3,207,040      100.00
```



```
. fre metro

metro —— metropolitan status
-----------------------------------------------------------------------------------------------
                                                  |      Freq.    Percent      Valid       Cum.
--------------------------------------------------+--------------------------------------------
Valid   0 metropolitan status indeterminable      |     466077      14.53      14.53      14.53
          (mixed)                                 |
        1 not in metropolitan area                |     334524      10.43      10.43      24.96
        2 in metropolitan area: in                |     380264      11.86      11.86      36.82
          central/principal city                  |
        3 in metropolitan area: not in            |     939182      29.29      29.29      66.11
          central/principal city                  |
        4 in metropolitan area: central/principal |    1086993      33.89      33.89     100.00
          city status indeterminable (mixed)      |
        Total                                     |    3207040     100.00     100.00
-----------------------------------------------------------------------------------------------
```

#6 MAKE VAR NAMES MAKE SENSE

- I like to make my new variables be `var_recode` if I am reorganizing an existing variable, `var_01` if the new variable is binary, or `var_per` if it’s a new percentage/rate variable
- Variable labels can help you stay organized!

```
lab var X “Description of X”
```

```
gen metro_recode =  metro

lab var metro_recode “Metro
Status Recode”
```

```
gen child_01 = age

lab var child_01 “Child 0-1”
```

#7 DEFAULT TO . WHEN MAKING A NEW VARIABLE

- When you are creating a new variable—especially categorical variables—it can be helpful to start a new variable that is "." (missing)
- Helps you make sure the new var is coded just the way you want.
- Helps you remember what values you haven't done yet.

```
fre metro


gen metro_recode = .

replace metro_recode = 0 if metro == 1

replace metro_recode = 1 if metro >= 2
```

```
. fre metro

metro — metropolitan status
-----------------------------------------------------------------------------------------------------------
                                                      |      Freq.    Percent      Valid       Cum.
------------------------------------------------------+----------------------------------------------------
Valid   0 metropolitan status indeterminable (mixed)  |     466077      14.53      14.53      14.53
        1 not in metropolitan area                    |     334524      10.43      10.43      24.96
        2 in metropolitan area: in central/principal city |  380264      11.86      11.86      36.82
        3 in metropolitan area: not in central/principal |    939182      29.29      29.29      66.11
          city                                        |
        4 in metropolitan area: central/principal city |   1086993      33.89      33.89     100.00
          status indeterminable (mixed)               |
        Total                                         |    3207040     100.00     100.00
-----------------------------------------------------------------------------------------------------------

. gen metro_recode = .
(3,207,040 missing values generated)

. replace metro_recode = 0 if metro == 1
(334,524 real changes made)

. replace metro_recode = 1 if metro >= 2
(2,406,439 real changes made)
```

#8 PUT CHECKS WITHIN YOUR DO FILE!

- After I create new variables, I put in some “checks’ to make sure that I did everything correct
- The easiest check is `tab X Y`
- By default the tab command doesn’t include missing values.Which can cause you some problems. Instead you must use `tab X Y, m`

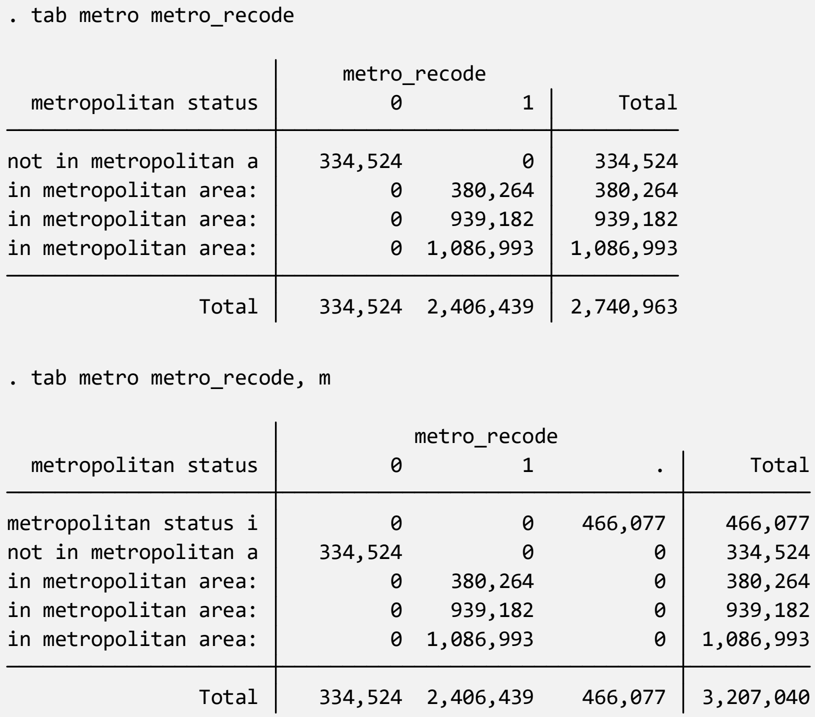
```
gen metro_recode = .

replace metro_recode = 0 if metro == 1

replace metro_recode = 1 if metro >= 2


tab metro metro_recode
```

WHAT DID I FORGET TO CODE?



```
. tab metro metro_recode

                      |     metro_recode
 metropolitan status  |         0          1 |     Total
----------------------+----------------------+----------
not in metropolitan a |   334,524          0 |   334,524
in metropolitan area: |         0    380,264 |   380,264
in metropolitan area: |         0    939,182 |   939,182
in metropolitan area: |         0  1,086,993 | 1,086,993
----------------------+----------------------+----------
                Total |   334,524  2,406,439 | 2,740,963

. tab metro metro_recode, m

                      |           metro_recode
 metropolitan status  |         0          1          . |     Total
----------------------+---------------------------------+----------
metropolitan status i |         0          0    466,077 |   466,077
not in metropolitan a |   334,524          0          0 |   334,524
in metropolitan area: |         0    380,264          0 |   380,264
in metropolitan area: |         0    939,182          0 |   939,182
in metropolitan area: |         0  1,086,993          0 | 1,086,993
----------------------+---------------------------------+----------
                Total |   334,524  2,406,439    466,077 | 3,207,040
```

#9 = VS. ==

- In STATA, = and == are not the same, and Stata won’t run your line of code if they aren’t the correct kind.

  = : Is used to set the value

  == :Tests for equality between two things. “Logical/Boolean” Operator.

```
replace educ_recode = 2 if educ == 4 | educ == 5
```

Rule of thumb: = goes before if, while == is used after if

QUICK TIPS FOR DO FILE

#10: Global macros are your friend

Instead of

```
use "$C:\Users\mfb5341\OneDrive - The
Pennsylvania State
University\Documents\Diss_Estimates\Data\Unforma
tted\Sample_18.dta", clear”
```

Use

```
global data "C:\Users\mfb5341\OneDrive - The
Pennsylvania State
University\Documents\Diss_Estimates\Data\Unformatted"
```

```
use “$data\Sample_18.dta,” clear

use “$data\Tax_data_18.dta,” clear
```

#11: Comment out your do file

```
*********

*Analysis


* New Variables

replace metro_recode = 0 if metro == 1

        // Obs in nonmetro

        // counties now = 0
```

#10 MAKE LONG LINES INTO SMALL LINES

Adding /// to commands in the do file, tells Stata to read multiple lines of code as one line

Too Long!

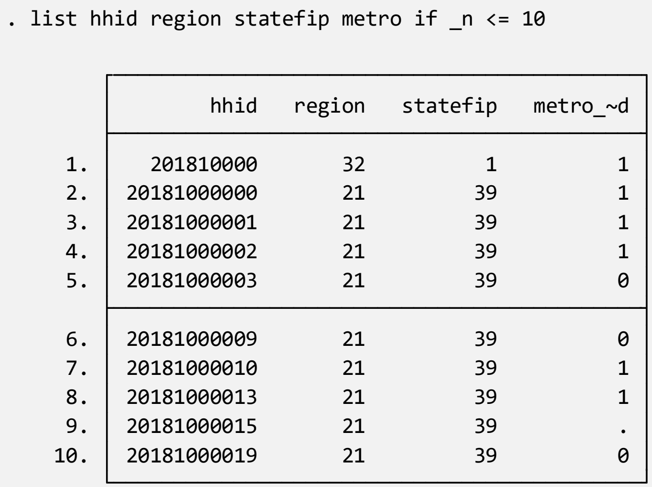
```
drop cluster countyfip density met2013 puma strata cpuma0010 farm rentmeal condofee moblhome
costelec costgas costwatr  costfuel foodstmp valueh builtyr2 unitsstr bedrooms vehicles bpl bpld
ancestr1 ancestr1d ancestr2
```

Much More Readable!

```
drop cluster countyfip density met2013 puma strata cpuma0010 ///
        farm rentmeal condofee moblhome costelec costgas ///
        costwatr  costfuel foodstmp valueh builtyr2 unitsstr
        bedrooms vehicles bpl bpld ancestr1 ancestr1d ancestr2 ///
```

#12 USE THE LIST AND BROWSE COMMANDS

Using the list and browse commands helps you debug why a new variable isn't working the way you want.



```
. list hhid region statefip metro if _n <= 10

     +---------------------------------------------+
     |        hhid   region   statefip   metro_~d |
     |---------------------------------------------|
  1. |   201810000       32          1          1 |
  2. | 20181000000       21         39          1 |
  3. | 20181000001       21         39          1 |
  4. | 20181000002       21         39          1 |
  5. | 20181000003       21         39          0 |
     |---------------------------------------------|
  6. | 20181000009       21         39          0 |
  7. | 20181000010       21         39          1 |
  8. | 20181000013       21         39          1 |
  9. | 20181000015       21         39          . |
 10. | 20181000019       21         39          0 |
     +---------------------------------------------+
```

Using "`if _n <= X`", only shows the first X obs.

The browse of "br" command brings up the data editor.

```
br hhid region statefip metro
```

#13 NEVER SAVE OVER THE ORIGINAL DATASET

```
use “$data\usa_001.dta.”
       // Has over 30 million obs. Too big to run all the time!

drop if year != 2018
       // Analysis only focuses on 2018, Removing all other obs.

save "$data\Sample_18.dta", replace
use "$data\Sample_18.dta", clear
       // Loads only the needed data, and not the whole original source.

*** Fancy Stata Stuff
save "$data\Sample_18_Formatted.dta", replace
       // Now that the dataset is formatted, I will use
       // this version for the analysis.
```

#14 RECODE VS. REPLACE

- Recode is for simply turning existing values into other values. This is useful for creating binary variables and when you only need the existing values of a variable to create the new values.

```
gen metro_status = rucc

recode metro_status (0=1) (2=1) (3=1) (4/9 = 2)
```

- Replace is must more powerful, but involves the use of "arguments" (==, >=, &, etc.) New values can be created based on the values of multiple variables. Each line of code can only replace 1 value.

```
gen imm_status = .

replace imm_status = 1 if nativ == 1 & citizen == 0 | citizen == 4

replace imm_status = 2 if nativ >= 1 & citizen == 1

replace imm_status = 3 if native >=1 & citizen == 2 | citizen == 3
```

#15 STATA RESOURCES

Stata has the built-in `help` command.

This command will bring up, in Stata, the pdf help sheet for the given command. Here there is the guide of all command options, requirements, and examples

```
help hist
help tab
```

Stata manual (Stata 16)

https://www.stata.com/manuals/u.pdf

UCLA

https://stats.idre.ucla.edu/stata/

Princeton

http://www.princeton.edu/~otorres/Stata/

University of Wisconsin

https://www.ssc.wisc.edu/sscc/pubs/sfr-intro.htm