**DEGREE INFORMATION SYSTEM**

**(DIS)**

**MHEC Internal Validation Code**

**STATA Syntax Code**

**Posted 11/15/2019**

# **Data Validation Syntax**

The following Stata code is used to validate all file submissions. Additional validation testing may be completed contingent upon the results of this baseline code. New validation steps may be added over time as the scope of the collection changes.

odbc load, table("dbo.COLL\_DIS\_VAL") lowercase noquote

\*rename from SQL name to SYSNAME\*

\*rename vars\*

rename coll\_tm coltm

rename coll\_yr colyr

rename opeid fice

rename sub\_camp\_cd subcamp

rename id\_num idn

rename id\_type idt

rename campus\_id campusid

rename deg\_sought degree

rename deg\_date degdate

rename prog\_tax program

rename frst\_maj\_link majorlk

rename cum\_gpa cumgpa

rename cum\_nat\_cr\_earn cncrhre

rename cum\_deg\_award ccrhra

rename prior\_deg\_flag pdflg

rename entry\_tm entrm

rename entry\_yr entyr

rename rev\_trans\_flag revtrn

rename cred\_hrs\_req\_aw rqcrhr

rename teach\_can teachcan

rename gender sex

rename dob birthdt

rename us\_citz citizen

rename blk\_aa rblack

rename white rwhite

rename asian rasian

rename am\_ind\_nat\_alsk raian

rename nat\_haw\_pac\_is rnhpi

rename hisp\_lat hispanic

destring, replace

\*construct sic and sector\*

\*first digit of sic is sector\*

\*second digit of sic is segment\*

\*construct sector and segment\*

\*step 1: convert to string\*

gen sic2=sic

tostring sic2, replace

\*step 2: sector - parse code to keep only first digit\*

gen sector=substr(sic2, 1, length(sic2)-5)

destring sector, replace

\*step 4: segment - parse code to keep only second digit\*

gen segment=substr(sic2, 2, length(sic2)-5)

destring segment, replace

\*step 3: drop sic2\*

drop sic2

\*construct majord and disp\*

\*first two digits are MAJORDS\*

\*first four digits are DISCPLN\*

\*step 1: convert to string\*

gen program2=program

tostring program2, replace

\*step 2: MAJORDS, first two digits\*

gen majords=substr(program2, 1, length(program2)-4)

destring majords, replace

\*step 3: DISCPLN - first four digits\*

gen discpln=substr(program2, 1, length(program2)-2)

destring discpln, replace

\*step 4: drop program2\*

drop program2

label variable coltm "Collection Term"

label variable colyr "Collection Year"

label variable fice "OPEID"

label variable sic "State ID"

label variable sector "Institution Ownership"

label variable segment "Education Segment"

label variable subcamp "Sub-Campus Code"

label variable tssn "Student ID"

label variable tssn2 "Student ID2"

label variable idt "ID Number Type"

label variable campusid "Local Campus Student ID"

label variable sasid "SASID"

label variable degdate "Degree Date"

label variable degree "Degree Sought"

label variable program "Program Taxonomy"

label variable majorlk "First Major Link"

label variable cumgpa "Cumulative GPA"

label variable cncrhre "Cumulative Native Credit Hours Earned"

label variable ccrhra "Cumulative Credit Hours Awarded"

label variable pdflg "Prior Degree Flag"

label variable entrm "Entry Term"

label variable entyr "Entry Year"

label variable revtrn "Reverse Transfer Flag"

label variable rqcrhr "Credit Hours Required to Earn Award"

label variable teachcan "Teacher Candidate"

label variable sex "Gender"

label variable citizen "US Citizenship"

label variable hispanic "Hispanic/Latino"

label variable rwhite "White"

label variable rblack "African American"

label variable rasian "Asian"

label variable raian "Amer Indian/Nat Alaskan"

label variable rnhpi "Native Hawaiian/Pac. Is."

label variable birthdt "Birthdate"

label define coltm ///

1 "Fall" ///

2 "Winter" ///

3 "Spring" ///

4 "Summer" ///

9 "Annual" ///

8 "Annual"

label values coltm entrm coltm

label define entrm ///

1 "Fall" ///

2 "Winter" ///

3 "Spring" ///

4 "Summer" ///

-9 "not collected"

label values entrm entrm

label define fice ///

00143600 "Capitol Technology" ///

00205700 "Allegany" ///

00205800 "Anne Arundel" ///

00206100 "Baltimore City" ///

00206200 "Bowie" ///

00206300 "Baltimore County" ///

00206400 "Southern Maryland" ///

00206500 "Notre Dame" ///

00206700 "Washington Adventist" ///

00206800 "Coppin" ///

00207100 "Frederick" ///

00207200 "Frostburg" ///

00207300 "Goucher" ///

00207400 "Hagerstown" ///

00207500 "Harford" ///

00207600 "Hood" ///

00207700 "JHU" ///

00207800 "Loyola" ///

00208000 "MICA" ///

00208300 "Morgan" ///

00208600 "Mount St. Mary's" ///

00208700 "Ner Israel Rabbinical College" ///

00208900 "Prince George's" ///

00209100 "Salisbury" ///

00209200 "St. John's" ///

00209500 "St. Mary's" ///

00209600 "St. Mary's Seminary" ///

00209900 "Towson" ///

00210200 "UB" ///

00210300 "UMCP" ///

00210400 "UMB" ///

00210500 "UMBC" ///

00210600 "UMES" ///

00210700 "Stevenson" ///

00210800 "Washington College" ///

00210900 "McDaniel" ///

00465000 "Chesapeake" ///

00691100 "Montgomery" ///

00795900 "USM Office" ///

00817500 "Howard" ///

00830800 "Cecil" ///

01001400 "Garrett" ///

01164400 "UMUC" ///

02073900 "Wor-Wic" ///

02083600 "Brightwood" ///

02127900 "Sojourner-Douglass" ///

02314800 "Baltimore Intnl. College" ///

02578400 "MD Integrative Health" ///

03100700 "Carroll" ///

03455500 "National Labor" ///

90795900 "UM -Environ. Sci." ///

99765900 "UM -Biotech"

label values fice fice

label define sic ///

110100 "Allegany" ///

110200 "Anne Arundel" ///

110770 "Carroll" ///

110900 "Cecil" ///

111000 "Southern Maryland" ///

111100 "Chesapeake" ///

111250 "Baltimore County" ///

111300 "Baltimore City " ///

111700 "Frederick" ///

111900 "Garrett" ///

112100 "Hagerstown" ///

112200 "Harford" ///

112400 "Howard" ///

112970 "Montgomery" ///

113600 "Prince George's" ///

115470 "Wor-Wic " ///

120600 "Bowie" ///

121400 "Coppin" ///

121800 "Frostburg" ///

123900 "Salisbury" ///

124200 "Towson" ///

124400 "UB" ///

124500 "UMB" ///

124600 "UMBC" ///

124700 "UMCP" ///

124800 "UMES" ///

124900 "UMUC" ///

124950 "USM Office" ///

124960 "UM -Environ. Sci." ///

124965 "UM -Biotech" ///

143000 "Morgan" ///

154000 "St. Mary's" ///

260400 "Baltimore Hebrew University" ///

260410 "Baltimore Intnl. College" ///

260469 "Bais Hamedrash and Mesivta of Balt" ///

260537 "Binah Institute of Adv Judiac Stdy" ///

260700 "Capitol Technology " ///

262000 "Goucher" ///

262050 "Hagerstown Business College" ///

262230 "Harry Lundeberg School Seamanship " ///

262300 "Hood" ///

262487 "ITT Technical Institute" ///

262500 "JHU" ///

262570 "Lincoln College of Technology" ///

262600 "Loyola" ///

262650 "Maryland College of Art & Design" ///

262700 "MICA" ///

262767 "Fortis College" ///

263100 "Mount St. Mary's " ///

263124 "National Labor College" ///

263200 "Ner Israel Rabbinical College" ///

263300 "Notre Dame " ///

263500 "Peabody Institute" ///

263700 "St. John's" ///

263915 "SANS Technology Institute" ///

264100 "St. Mary's Seminary" ///

264150 "Sojourner-Douglass College" ///

264182 "Talmudical Academy of Baltimore" ///

264187 "Brightwood" ///

264300 "MD Integrative Health" ///

265000 "Stevenson" ///

261200 "Washington Adventist" ///

265100 "Washington Bible" ///

265200 "Washington College" ///

265300 "Washington Theological" ///

265400 "McDaniel" ///

265420 "Women's Inst of Torah Seminary" ///

265600 "Yeshiva College of Nations Capital" ///

261575 "Eastern Christian" ///

263575 "Potomac College" ///

260500 "Bay College" ///

261150 "Columia cultural Institute" ///

261450 "DeSales Hall" ///

110800 "Catonsville Community College" ///

111500 "Dundalk Community College" ///

111600 "Essex Community College" ///

112800 "Montgomery - Rockville" ///

112900 "Montgomery - Takoma Park" ///

112950 "Montgomery - Germantown" ///

260250 "Antioch College"

label values sic sic

label define sector ///

1 "public" ///

2 "private"

label values sector sector

label define segment ///

1 "community college" ///

2 "usm" ///

4 "morgan" ///

5 "st marys" ///

6 "independent"

label values segment segment

label define idt ///

1 "ssn" ///

2 "inst assign id" ///

3 "itin"

label values idt idt

label define degree ///

01 "Pri Car Dip/Cert" ///

10 "Cert Lower" ///

20 "Associate" ///

30 "Cert Upper" ///

40 "Bachelor" ///

47 "Non-Deg/Hs Stud" ///

50 "Cert Post-Bach" ///

60 "Masters" ///

65 "Cert Post Ma" ///

70 "Cert Adv Study" ///

80 "Doctorate" ///

81 "Doc R/S" ///

85 "Doc Prof Prac" ///

86 "Doc Other" ///

87 "Ns Grad First Prof" ///

90 "First Prof" ///

95 "Cert First Prof" ///

97 "Nd First Prof" ///

99 "Multi Major" ///

00 "Unknown"

label values degree degree

label define sex ///

1 "Male" ///

2 "Female" ///

3 "Male Assigned" ///

4 "Female Assigned" ///

-9 "not collected"

label values sex sex

label define citizen ///

1 "US Citizen" ///

2 "Nonresident Alien" ///

-9 "not collected"

label values citizen citizen

label define race ///

1 "Yes" ///

2 "Yes" ///

3 "Yes" ///

4 "Yes" ///

5 "Yes" ///

-9 "not collected"

label values rwhite rblack rasian raian rnhpi race

\*save file for validation work\*

save "M:\path\DIS\DIS\_Xxxx\_20##\_working\_v#.dta"

\*file population\*

\*run distribution and compare to prior year\*

\*review for population increases and decreases of more than 10%\*

\*cross check counts of fice vs sic to ensure sic is correctly constructed\*

table fice

table sic

table degree

\*counts of degrees\*

\*compare to prior years\*

\*review for population increases and decreases of more than 10%\*

table fice if degree==10 | degree==30

table fice if degree==20

table fice if degree==40

table fice if degree==50

table fice if degree==60

table fice if (degree>=65 & degree<99) | degree==47 | degree==00 | degree==01

table fice if degree==99

tab fice if degree==47

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Duplicate Check\*

\*this must be done first as variables will be constructed throughout validation that will

\*alter results when looking for duplicates across all variables\*

sort fice idn

\*dups on all variables\*

duplicates tag, generate(dups) /\*duplicates in data set, data set duplicates on all elements are likely errors\*/

tab dups

tab fice dups if dups>0

\*duplicate checking on ssn, fice, program, degree level\*

\*duplicates on program and degree level (two identical degrees) within a fice are possible errors\*

sort fice idn program degree

duplicates tag idn fice program degree, generate(ssn\_pg)

tab ssn\_pg

tab fice ssn\_pg if ssn\_pg>0

\*dups on ssn only\*

\*validate counts are same on idn, tssn and tssn2 to make sure idn is correctly converted\*

\*id may appear multiple times if student earns more than one degree in one or more FICE\*

duplicates tag idn, generate (dupssn) /\*not an error, student attending multiple fice\*/

duplicates tag tssn, generate (duptssn)/\*not an error, student attending multiple fice\*/

duplicates tag tssn2, generate (duptssn2)/\*not an error, student attending multiple fice\*/

tab dupssn

tab duptssn

tab duptssn2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*count unique populations\*

\*first instance of each unit record within an fice is marked as unique, additional instances as duplciate\*

\*add to excel tab and cross check against prior year totals\*

\*dis should be one row per idn per fice per degree, duplicates are students earning two degrees, not errors\*

\*counts should be the same on idn, tssn, tssn2\*

gsort fice idn -degree

egen idn\_unique=tag(fice idn)

gsort fice tssn -degree

egen tssn\_unique=tag(fice tssn)

gsort fice tssn2 -degree

egen tssn2\_unique=tag(fice tssn2)

label define counts 1 "unique" 0 "duplicate"

label values idn\_unique tssn\_unique tssn2\_unique counts

tab idn\_unique

tab tssn\_unique

tab tssn2\_unique

tab fice idn\_unique

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate FICE-SIC alignment\*

\*does the sic expected for the fice match\*

\*does every fice have a sic\*

gen sicfice=.

replace sicfice=1 if fice==143600& sic==260700

replace sicfice=1 if fice==205700& sic==110100

replace sicfice=1 if fice==205800& sic==110200

replace sicfice=1 if fice==206100& sic==111300

replace sicfice=1 if fice==206200& sic==120600

replace sicfice=1 if fice==206300& sic==111250

replace sicfice=1 if fice==206400& sic==111000

replace sicfice=1 if fice==206500& sic==263300

replace sicfice=1 if fice==206700& sic==261200

replace sicfice=1 if fice==206800& sic==121400

replace sicfice=1 if fice==207100& sic==111700

replace sicfice=1 if fice==207200& sic==121800

replace sicfice=1 if fice==207300& sic==262000

replace sicfice=1 if fice==207400& sic==112100

replace sicfice=1 if fice==207500& sic==112200

replace sicfice=1 if fice==207600& sic==262300

replace sicfice=1 if fice==207700& sic==262500

replace sicfice=1 if fice==207800& sic==262600

replace sicfice=1 if fice==208000& sic==262700

replace sicfice=1 if fice==208300& sic==143000

replace sicfice=1 if fice==208600& sic==263100

replace sicfice=1 if fice==208900& sic==113600

replace sicfice=1 if fice==209100& sic==123900

replace sicfice=1 if fice==209200& sic==263700

replace sicfice=1 if fice==209500& sic==154000

replace sicfice=1 if fice==209900& sic==124200

replace sicfice=1 if fice==210200& sic==124400

replace sicfice=1 if fice==210300& sic==124700

replace sicfice=1 if fice==210400& sic==124500

replace sicfice=1 if fice==210500& sic==124600

replace sicfice=1 if fice==210600& sic==124800

replace sicfice=1 if fice==210700& sic==265000

replace sicfice=1 if fice==210800& sic==265200

replace sicfice=1 if fice==210900& sic==265400

replace sicfice=1 if fice==465000& sic==111100

replace sicfice=1 if fice==691100& sic==112970

replace sicfice=1 if fice==817500& sic==112400

replace sicfice=1 if fice==830800& sic==110900

replace sicfice=1 if fice==1001400& sic==111900

replace sicfice=1 if fice==1164400& sic==124900

replace sicfice=1 if fice==2073900& sic==115470

replace sicfice=1 if fice==3100700& sic==110770

tab sicfice

\*validate individual values in data elements\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid fice, 1 valid, -9 not valid\*

gen vfice=.

replace vfice=1 if fice==143600 | fice==205700 | fice==205800 | fice==206100 | fice==206200 | fice==206300 | fice==206400 | ///

fice==206500 | fice==206700 | fice==206800 | fice==207100 | ///

fice==207200 | fice==207300 | fice==207400 | fice==207500 | fice==207600 | fice==207700 | ///

fice==207800 | fice==208000 | fice==208600 | fice==208700 | fice==208900 | ///

fice==209100 | fice==209200 | fice==209500 | fice==209600 | fice==209900 | fice==210200 | fice==210300 | ///

fice==210400 | fice==210500 | fice==210600 | fice==210700 | fice==210800 | fice==210900 | fice==465000 | ///

fice== 691100 | fice==795900 | fice==817500 | fice==830800 | fice==1001400 | fice==1164400 | fice== 2073900 | ///

fice==208300 | fice==2127900 | fice==2314800 | fice==2578400 | fice==3100700 | fice==3455500 | fice==90795900 | fice==99765900

mvdecode vfice, mv(-9)

tab vfice /\*valid fice, 1 valid, -9 not valid\*/

tab fice if vfice==-9 /\*list of invalid fice numbers\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid sic, 1 valid, -9 not valid\*

gen vsic=.

replace vsic=1 if sic==110100 | sic==110200 | sic==110770 | sic==110900 | sic==111000 | ///

sic==111100| sic==111250| sic==111300 | sic==111700 | sic==111900 | sic==112100 | ///

sic==112200 | sic==112400| sic==112970 | sic==113600 | sic==115470| sic==120600 | ///

sic==121400 | sic==121800| sic==123900| sic==124200 | sic==124400| sic==124500 | ///

sic==124600 | sic==124700 | sic==124800| sic==124900 | sic==124960 | sic==124965 | ///

sic==143000 | sic==154000 | sic==260400 | sic==260410 | sic==260469 | sic==260537| ///

sic==260700 | sic==262000 | sic==262050 | sic==262230 | sic==262300| sic==262487| ///

sic==262500| sic==262570 | sic==262600| sic==262650 | sic==262700 | sic==262767 | ///

sic==263100 | sic==263124 | sic==263200| sic==263300| sic==263500 | sic==263700| ///

sic==263915 | sic==264100| sic==264150 | sic==264182| sic==264187 | sic==264300 | ///

sic==265000| sic==261200| sic==265100 | sic==265200 | sic==265300| sic==265400| ///

sic==265420| sic==265600 | sic==261575| sic==263575| sic==260500 | sic==261150| ///

sic==261450| sic==110800| sic==111500 | sic==111600 | sic==112800| sic==112900| ///

sic==112950 | sic==260250

mvdecode vsic, mv(-9)

tab vsic, missing /\*1 valid, -9 not valid\*/

tab sic if vsic==-9 /\*list of invalid sic numbers\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Validate collection term\*

\*valid collection term, 1 is valid, 2 is not valid\*

gen vct = .

\*select one of the following\*

replace vct = 1 if coltm ==9

replace vct = 2 if coltm !=9

tab vct /\*1 is valid, 2 is not valid\*/

tab fice if vct==2 /\*1 is valid, 2 is not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid collection year, 1 is valid, 2 is not valid\*

\*change year as needed\*

gen vcy = .

\*select one of the following\*

replace vcy = 1 if colyr == 2018

replace vcy = 2 if colyr != 2018

replace vcy = 1 if colyr == 2019

replace vcy = 2 if colyr != 2019

replace vcy = 1 if colyr == 2020

replace vcy = 2 if colyr != 2020

tab vcy /\*1 is valid, 2 is not valid\*/

tab fice if vcy==2 /\*1 is valid, 2 is not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*sex, 1 valid\*

gen vsex=.

replace vsex=1 if sex==1| sex==2| sex==3| sex==4

replace vsex=2 if sex>=5 | sex==. | sex<1

tab sex

tab vsex /\*1 is valid, 2 not valid\*/

tab fice if vsex==2 /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*citizenship\*

gen vcit=.

replace vcit=1 if citizen==1|citizen==2

replace vcit=2 if citizen<1 | citizen>2 | citizen==.

tab vcit /\*1 is valid, 2 not valid\*/

tab citizen

tab fice if vcit==2 /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*race/ethnicity\*

encode hispanic, gen (hislat)

gen vhisp=.

replace vhisp=1 if hislat==1 | hislat==2 |hislat==.

replace vhisp=2 if hislat!=1 & hislat!=2 & hislat!=.

mvencode hislat, mv(9)

tab hispanic

tab hislat

tab vhisp /\*1 is valid, 2 not valid\*/

tab fice if vhisp==2 /\*1 is valid, 2 not valid\*/

gen vrwhite=.

replace vrwhite=1 if rwhite==1 |rwhite==.

replace vrwhite=2 if rwhite!=1 & rwhite!=.

tab vrwhite /\*1 is valid, 2 not valid\*/

tab fice if vrwhite==2 /\*1 is valid, 2 not valid\*/

gen vrblack=.

replace vrblack=1 if rblack==2|rblack==.

replace vrblack=2 if rblack!=2 & rblack!=.

tab vrblack /\*1 is valid, 2 not valid\*/

tab fice if vrblack==2 /\*1 is valid, 2 not valid\*/

gen vrasian=.

replace vrasian=1 if rasian==3 |rasian==.

replace vrasian=2 if rasian!=3 & rasian!=.

tab vrasian /\*1 is valid, 2 not valid\*/

tab fice if vrasian==2 /\*1 is valid, 2 not valid\*/

gen vamer=.

replace vamer=1 if raian==4 |raian==.

replace vamer=2 if raian!=4 & raian!=.

tab vamer /\*1 is valid, 2 not valid\*/

tab fice if vamer==2 /\*1 is valid, 2 not valid\*/

gen vhawi=.

replace vhawi=1 if rnhpi==5 |rnhpi==.

replace vhawi=2 if rnhpi!=5 & rnhpi!=.

tab vhawi /\*1 is valid, 2 not valid\*/

tab fice if vhawi==2 /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*birthdt - must be 8 digits with at least a valid 4 digit year + 0000 or can be 00000000 if unknown\*

\*considering birthdt between 1900 and 2017 valid, tagging unknowns as 3\*

\*1 and 4 valid, 2 and 3 not valid\*

\*review frequency distribution at fice level for 3 (unknown birthdate)\*

gen vbirth=.

replace vbirth=1 if birthdt >=19170000 & birthdt <=20100000 /\*valid\*/

replace vbirth=2 if birthdt <19170000 & birthdt!=0 /\*not valid\*/

replace vbirth=2 if birthdt >20100000 & birthdt!=0 /\* not valid\*/

replace vbirth=3 if birthdt==. /\*not valid\*/

replace vbirth=4 if birthdt==0 /\*review\*/

tab vbirth

tab fice vbirth if vbirth>=2 /\*not valid or requires review\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid ID Type\*

gen vtype=.

replace vtype = 1 if idt ==1 | idt==2| idt==3

replace vtype = 2 if idt !=1 & idt !=2 & idt!=3 & idt !=.

tab vtype /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid SSN, 1 is valid, 2 is 8 digits, 3 is 10 or more digits, 4 is 7 or less digits\*

\*8 digits may indicate missing leading 0\*

/\*SSA will not issue SSNs beginning with the number “9”. SSA will not issue SSNs

beginning with the number “666” in positions 1 – 3. SSA will not issue SSNs beginning

with the number “000” in positions 1 – 3. ... SSA will not issue SSNs with the

number “0000” in positions 6 – 9.\*/

gen vssn =idn

tostring vssn, replace

\*testing for lenght of string\*

\*note, on import, the string is destrung, which will drop leading 0s\*

\*when the string is re-instated the 0s remain dropped\*

\*this helps identify IDNs with too many leading zeros but an IDT of SSN\*

gen vssn2=.

replace vssn2=1 if strlen(vssn)==9

replace vssn2=2 if strlen(vssn)==8

replace vssn2=3 if strlen(vssn)>=10

replace vssn2=4 if strlen(vssn)<=7

tab vssn2

\*validate against id type\*

tab vssn2 if idt==1 &(vssn2==3|vssn2==4) /\*number of invalid IDNs with ID Type =SSN\*/

tab fice if idt==1 &(vssn2==3|vssn2==4) /\*fice with number of invalid IDNs with ID Type =SSN\*/

tab vssn2 idt

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate TSSN\*

\*tssn should be 10 characters\*

gen vtssn=.

replace vtssn=3 if strlen(tssn)>10

replace vtssn=2 if strlen(tssn)<10

replace vtssn=1 if strlen(tssn)==10

tab vtssn /\*1 is valid, 2 & 3 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate TSSN2\*

\*tssn2 should be 7 characters\*

gen vtssn2=.

replace vtssn2=3 if strlen(tssn2)>7

replace vtssn2=2 if strlen(tssn2)<7

replace vtssn2=1 if strlen(tssn2)==7

tab vtssn2 /\*1 is valid, 2 & 3 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid campusID\*

\*campus id is 10 or less characters\*

gen vcamp=.

replace vcamp=2 if strlen(tssn)>=11

replace vcamp=1 if strlen(tssn)<=10

tab vcamp /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid SASID, code missing values as 9\*

gen vsasid =sasid

mvencode vsasid, mv(9)

tostring vsasid, replace

\*valid SASID\*, 1 and 2 are valid, 3 and 4 are not valid

gen vsasid2=.

replace vsasid2=4 if strlen(vsasid)>10 /\* not valid, too long\*/

replace vsasid2=3 if strlen(vsasid)<10 &strlen(vsasid)>1 /\* not valid, too short\*/

replace vsasid2=2 if strlen(vsasid)==10 /\*valid\*/

replace vsasid2=1 if strlen(vsasid)==1 /\*valid coded as 9 missing value\*/

tab vsasid2

tab fice if vsasid2>=3 /\*fice with SASID not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid program taxomony\*

/\*program taxonomy is 4 digits; however, leading 0 is dropped so validate all

leading 0 pt as 3 digit codes; excel truncates last two digits from cell so the 3 remaining are 0### even though

0 is not in cell\*/

\*step 1: convert to string\*

rename program program\_orig

tostring program\_orig, gen(ptax)

\*step 2: parse code to remove last two digits from ptax to retain only first 3 or 4 of program taxonomy\*

\*so, start with the first number and keep 2 digits\*

gen program\_taxonomy=substr(ptax, 1, length(ptax)-2)

\*finally, destring to reinstate numbers for validation both degree code and program taxonomy\*

destring program\_taxonomy, replace

gen vtax=.

replace vtax=1 if program\_taxonomy==100 |program\_taxonomy==101 | program\_taxonomy==102 | program\_taxonomy==103 | program\_taxonomy==104 | program\_taxonomy==105 | program\_taxonomy==106 | ///

program\_taxonomy==107 | program\_taxonomy==108 | program\_taxonomy==109 | program\_taxonomy==110 | program\_taxonomy==111 | program\_taxonomy==112 | program\_taxonomy==113 | ///

program\_taxonomy==114 | program\_taxonomy==115 | program\_taxonomy==116 | program\_taxonomy==117 | program\_taxonomy==199 | ///

program\_taxonomy==200 | program\_taxonomy==201 | program\_taxonomy==202 | program\_taxonomy==203 | program\_taxonomy==204 | program\_taxonomy==205 | program\_taxonomy==206 | program\_taxonomy==299 | ///

program\_taxonomy==300 | program\_taxonomy==301 | program\_taxonomy==302 | program\_taxonomy==303 | program\_taxonomy==304 | program\_taxonomy==305 | program\_taxonomy==306 | program\_taxonomy==307 | ///

program\_taxonomy==308 | program\_taxonomy==309 | program\_taxonomy==310 | program\_taxonomy==311 | program\_taxonomy==312 | program\_taxonomy==313 | program\_taxonomy==314 | program\_taxonomy==399 | ///

program\_taxonomy==400 | program\_taxonomy==401 | program\_taxonomy==402 | program\_taxonomy==403 | program\_taxonomy==404 | program\_taxonomy==405 | program\_taxonomy==406 | program\_taxonomy==407 | ///

program\_taxonomy==408 | program\_taxonomy==409 | program\_taxonomy==410 | program\_taxonomy==411 | program\_taxonomy==412 | program\_taxonomy==413 | program\_taxonomy==414 | program\_taxonomy==415 | ///

program\_taxonomy==416 | program\_taxonomy==417 | program\_taxonomy==418 | program\_taxonomy==419 | program\_taxonomy==420 | program\_taxonomy==421 | program\_taxonomy==422 | program\_taxonomy==423 | ///

program\_taxonomy==424 | program\_taxonomy==425 | program\_taxonomy==426 | program\_taxonomy==427 | program\_taxonomy==499 | ///

program\_taxonomy==500 | program\_taxonomy==501 | program\_taxonomy==502 | program\_taxonomy==503 | program\_taxonomy==504 | program\_taxonomy==505 | program\_taxonomy==506 | program\_taxonomy==507

replace vtax=1 if program\_taxonomy==508 | program\_taxonomy==509 | program\_taxonomy==510 | program\_taxonomy==511 | program\_taxonomy==512 | program\_taxonomy==513 | program\_taxonomy==514 | program\_taxonomy==515 | ///

program\_taxonomy==516 | program\_taxonomy==517 | program\_taxonomy==599 | ///

program\_taxonomy==600 | program\_taxonomy==601 | program\_taxonomy==602 | program\_taxonomy==603 | program\_taxonomy==604 | program\_taxonomy==605 | program\_taxonomy==699 | ///

program\_taxonomy==700 | program\_taxonomy==701 | program\_taxonomy==702 | program\_taxonomy==703 | program\_taxonomy==704 | program\_taxonomy==705 | program\_taxonomy==799 | ///

program\_taxonomy==800 |program\_taxonomy==801 | program\_taxonomy==802 | program\_taxonomy==803 | program\_taxonomy==804 | ///

program\_taxonomy==805 | program\_taxonomy==806 | program\_taxonomy==807 | program\_taxonomy==808 | ///

program\_taxonomy==809 | program\_taxonomy==810 | program\_taxonomy==811 | ///

program\_taxonomy==812 | program\_taxonomy==813 | program\_taxonomy==814 | program\_taxonomy==815 | ///

program\_taxonomy==816 | program\_taxonomy==817 | program\_taxonomy==818 | program\_taxonomy==819 | ///

program\_taxonomy==820 | program\_taxonomy==821 | program\_taxonomy==822 | program\_taxonomy==823 | program\_taxonomy==824 | ///

program\_taxonomy==825 | program\_taxonomy==826 | program\_taxonomy==827 | program\_taxonomy==828 | ///

program\_taxonomy==829 | program\_taxonomy==830 | program\_taxonomy==831 | program\_taxonomy==832 | ///

program\_taxonomy==833 | program\_taxonomy==834 | program\_taxonomy==835 | program\_taxonomy==836 | ///

program\_taxonomy==837 | program\_taxonomy==838 | program\_taxonomy==839 | program\_taxonomy==899

replace vtax=1 if program\_taxonomy==900 |program\_taxonomy==901 | program\_taxonomy==902 | ///

program\_taxonomy==903 | program\_taxonomy==904 | program\_taxonomy==905 | program\_taxonomy==906 | ///

program\_taxonomy==907 | program\_taxonomy==908 | program\_taxonomy==909 | program\_taxonomy==910 | ///

program\_taxonomy==911 | program\_taxonomy==912 | program\_taxonomy==913 | program\_taxonomy==914 | ///

program\_taxonomy==915 | program\_taxonomy==916 | program\_taxonomy==917 | program\_taxonomy==918 | ///

program\_taxonomy==919 | program\_taxonomy==920 | program\_taxonomy==921 | program\_taxonomy==922 | ///

program\_taxonomy==923 | program\_taxonomy==924 | program\_taxonomy==925 | program\_taxonomy==999 | ///

program\_taxonomy==1000 |program\_taxonomy==1001 | program\_taxonomy==1002 | program\_taxonomy==1003 | program\_taxonomy==1004 | ///

program\_taxonomy==1005 | program\_taxonomy==1006 | program\_taxonomy==1007 | program\_taxonomy==1008 | ///

program\_taxonomy==1009 | program\_taxonomy==1010 | program\_taxonomy==1011 | program\_taxonomy==1099

replace vtax=1 if program\_taxonomy==1100 |program\_taxonomy==1101 | program\_taxonomy==1102 | ///

program\_taxonomy==1103 | program\_taxonomy==1104 | program\_taxonomy==1105 | program\_taxonomy==1106 | ///

program\_taxonomy==1107 | program\_taxonomy==1108 | program\_taxonomy==1109 | program\_taxonomy==1110 | ///

program\_taxonomy==1111 | program\_taxonomy==1112 | program\_taxonomy==1113 | ///

program\_taxonomy==1114 | program\_taxonomy==1115 | program\_taxonomy==1116 | program\_taxonomy==1199 | program\_taxonomy==1201 | ///

program\_taxonomy==1200 |program\_taxonomy==1202 | program\_taxonomy==1203 | program\_taxonomy==1204 | program\_taxonomy==1205 | program\_taxonomy==1206 | ///

program\_taxonomy==1207 | program\_taxonomy==1208 | program\_taxonomy==1209 | program\_taxonomy==1210 | ///

program\_taxonomy==1211 | program\_taxonomy==1212 | program\_taxonomy==1213 | program\_taxonomy==1214 | ///

program\_taxonomy==1215 | program\_taxonomy==1216 | program\_taxonomy==1217 | program\_taxonomy==1218 | ///

program\_taxonomy==1219 | program\_taxonomy==1220 | program\_taxonomy==1221 | program\_taxonomy==1222 | ///

program\_taxonomy==1223 | program\_taxonomy==1224 | program\_taxonomy==1225 | program\_taxonomy==1299 | ///

program\_taxonomy==1300 |program\_taxonomy==1301 | program\_taxonomy==1302 | program\_taxonomy==1303 | program\_taxonomy==1304 | ///

program\_taxonomy==1305 | program\_taxonomy==1306 | program\_taxonomy==1307 | program\_taxonomy==1399

replace vtax=1 if program\_taxonomy==1400 |program\_taxonomy==1401 | program\_taxonomy==1499 | ///

program\_taxonomy==1500 |program\_taxonomy==1501 | program\_taxonomy==1502 | program\_taxonomy==1503 | ///

program\_taxonomy==1504 | program\_taxonomy==1505 | program\_taxonomy==1506 | program\_taxonomy==1507 | ///

program\_taxonomy==1508 | program\_taxonomy==1509 | program\_taxonomy==1510 | program\_taxonomy==1599 | ///

program\_taxonomy==1600 |program\_taxonomy==1601 | program\_taxonomy==1699 | ///

program\_taxonomy==1700 |program\_taxonomy==1701 | program\_taxonomy==1702 | program\_taxonomy==1703 | program\_taxonomy==1799 | ///

program\_taxonomy==1800 |program\_taxonomy==1801 | program\_taxonomy==1802 | program\_taxonomy==1803 | program\_taxonomy==1899

replace vtax=1 if program\_taxonomy==1900 |program\_taxonomy==1901 | program\_taxonomy==1902 | program\_taxonomy==1903 | program\_taxonomy==1904 | program\_taxonomy==1905 | ///

program\_taxonomy==1906 | program\_taxonomy==1907 | program\_taxonomy==1908 | program\_taxonomy==1909 | ///

program\_taxonomy==1910 | program\_taxonomy==1911 | program\_taxonomy==1912 | program\_taxonomy==1913 | ///

program\_taxonomy==1914 | program\_taxonomy==1915 | program\_taxonomy==1916 | program\_taxonomy==1917 | ///

program\_taxonomy==1918 | program\_taxonomy==1919 | program\_taxonomy==1920 | program\_taxonomy==1999 | ///

program\_taxonomy==2000 |program\_taxonomy==2001 | program\_taxonomy==2002 | program\_taxonomy==2003 | ///

program\_taxonomy==2004 | program\_taxonomy==2005 | program\_taxonomy==2006 | program\_taxonomy==2007 | ///

program\_taxonomy==2008 | program\_taxonomy==2009 | program\_taxonomy==2010 | program\_taxonomy==2099 | ///

program\_taxonomy==2100 |program\_taxonomy==2101 | program\_taxonomy==2102 | program\_taxonomy==2103 | ///

program\_taxonomy==2104 | program\_taxonomy==2105 | program\_taxonomy==2106 | program\_taxonomy==2199

replace vtax=1 if program\_taxonomy==2200 |program\_taxonomy==2201 | program\_taxonomy==2202 | program\_taxonomy==2203 | program\_taxonomy==2204 | ///

program\_taxonomy==2205 | program\_taxonomy==2206 | program\_taxonomy==2207 | program\_taxonomy==2208 | ///

program\_taxonomy==2209 | program\_taxonomy==2210 | program\_taxonomy==2211 | program\_taxonomy==2212 | ///

program\_taxonomy==2213 | program\_taxonomy==2214 | program\_taxonomy==2215 | program\_taxonomy==2299 | program\_taxonomy==2301 | ///

program\_taxonomy==2300 |program\_taxonomy==2302 | program\_taxonomy==2303 | program\_taxonomy==2304 | program\_taxonomy==2399

replace vtax=1 if program\_taxonomy==4900 |program\_taxonomy==4901 | program\_taxonomy==4902 | program\_taxonomy==4903 | program\_taxonomy==4904 | program\_taxonomy==4910 | ///

program\_taxonomy==4920 | program\_taxonomy==4930 | program\_taxonomy==4940 | program\_taxonomy==4950 | ///

program\_taxonomy==4960 | program\_taxonomy==4970 | program\_taxonomy==4980 | program\_taxonomy==4999 | ///

program\_taxonomy==5000 |program\_taxonomy==5001 | program\_taxonomy==5002 | program\_taxonomy==5003 | ///

program\_taxonomy==5004 | program\_taxonomy==5005 | program\_taxonomy==5006 | program\_taxonomy==5007 | program\_taxonomy==5008 | ///

program\_taxonomy==5009 | program\_taxonomy==5010 | program\_taxonomy==5011 | program\_taxonomy==5012 | program\_taxonomy==5099

replace vtax=1 if program\_taxonomy==5100 |program\_taxonomy==5101 | program\_taxonomy==5102 | ///

program\_taxonomy==5103 | program\_taxonomy==5104 | program\_taxonomy==5105 | program\_taxonomy==5199 | ///

program\_taxonomy==5200 |program\_taxonomy==5201 | program\_taxonomy==5202 | ///

program\_taxonomy==5203 | program\_taxonomy==5204 | program\_taxonomy==5205 | program\_taxonomy==5206 | ///

program\_taxonomy==5207 | program\_taxonomy==5208 | program\_taxonomy==5209 | program\_taxonomy==5210 | ///

program\_taxonomy==5211 | program\_taxonomy==5212 | program\_taxonomy==5213 | program\_taxonomy==5214 | program\_taxonomy==5215 | ///

program\_taxonomy==5216 | program\_taxonomy==5217 | program\_taxonomy==5218 | program\_taxonomy==5219 | program\_taxonomy==5299

replace vtax=1 if program\_taxonomy==5300 |program\_taxonomy==5301 | program\_taxonomy==5302 | program\_taxonomy==5303 | ///

program\_taxonomy==5304 | program\_taxonomy==5305 | program\_taxonomy==5306 | program\_taxonomy==5307 | program\_taxonomy==5308 | ///

program\_taxonomy==5309 | program\_taxonomy==5310 | program\_taxonomy==5311 | program\_taxonomy==5312 | program\_taxonomy==5313 | ///

program\_taxonomy==5314 | program\_taxonomy==5315 | program\_taxonomy==5316 | program\_taxonomy==5317 | program\_taxonomy==5399 | ///

program\_taxonomy==5400 |program\_taxonomy==5401 | program\_taxonomy==5402 | program\_taxonomy==5403 | program\_taxonomy==5404 | ///

program\_taxonomy==5405 | program\_taxonomy==5406 | program\_taxonomy==5407 | program\_taxonomy==5408 | program\_taxonomy==5499 | ///

program\_taxonomy==5501 | program\_taxonomy==5502 | program\_taxonomy==5503 | program\_taxonomy==5292 | ///

program\_taxonomy==5504 | program\_taxonomy==5505 | program\_taxonomy==5506 | program\_taxonomy==5507 | program\_taxonomy==5508 | program\_taxonomy==5599 | program\_taxonomy==5601 | program\_taxonomy==9099

mvencode vtax, mv(9) /\*any program taxonomy not in list above code as 9\*/

tab vtax /\*1 is valid, 9 is not in API\*/

tab fice vtax if vtax==9

tab fice program\_taxonomy if vtax==9

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Valid degree sought\*

gen vdegree=.

replace vdegree=1 if degree==10 | degree==20 | degree==30 | degree==40 | degree==50 | degree==60 /\*valid\*/

replace vdegree=1 if degree== 65|degree== 70| degree==81| degree==85| degree==86| degree==99 /\*valid\*/

replace vdegree=2 if (degree>10&degree<20)|(degree>20&degree<30)|(degree>30&degree<40)|(degree>40&degree<47)/\*out of range\*/

replace vdegree=2 if(degree>47&degree<50)|(degree>50&degree<60)|(degree>60&degree<65)/\*out of range\*/

replace vdegree=2 if(degree>65&degree<70)|(degree>70&degree<81)|(degree>81&degree<85)|(degree>86&degree<99)/\*out of range\*/

replace vdegree=3 if degree ==./\*blank\*/

replace vdegree=4 if degree<10|degree>99/\*out of range\*/

tab vdegree

tab fice if vdegree>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*entry term\*

gen veterm=.

replace veterm=1 if entrm==1| entrm==2| entrm==3| entrm==4 /\*valid\*/

replace veterm=2 if entrm!=1 & entrm!=2& entrm!=3& entrm!=4 & entrm!=. /\*not valid\*/

tab veterm

tab fice if veterm==2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*entry year\*

gen veyear=.

replace veyear=1 if entyr >=1000 /\*valid\*/

replace veyear=2 if entyr<=999 /\*not valid\*/

replace veyear=3 if entyr==. /\*not valid\*/

tab veyear

tab fice veyear if veyear>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*reverse flag\*

/\*This is a student who transfers from a 4-year institution to a 2-year institution.

Only the 2-year institution awarding the associate's degree should flag the student.\*/

encode revtrn, gen (rtrans)

tab rtrans

gen vrtrans=.

replace vrtrans=1 if (rtrans==1 & degree==20) /\*valid Y with Assoc degree\*/

replace vrtrans=2 if rtrans==. /\*valid blank with any degree\*/

replace vrtrans=3 if rtrans!=1&rtrans!=. /\*invalid, data set sent with space rather than blank (or other value)\*/

replace vrtrans=4 if (rtrans==1 & degree!=20) /\*invalid, Y entered with degree other than Assoc\*/

tab vrtrans

tab fice vrtrans if vrtrans>2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid Degree Date\*

\*update sequence each year to align to expected graduation year for collection\*

gen vddate=.

replace vddate=1 if degdate !=. & degdate>=20180701 /\*valid\*/

replace vddate=2 if degdate ==./\*not valid\*/

replace vddate=3 if degdate <=20180700/\*not valid\*/

replace vddate=4 if degdate >=20190700/\*not valid\*/

tab vddate

tab fice vddate if vddate>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*prior degree\*

encode pdflg, gen (pdegree)

tab pdegree

gen vpdegree=.

replace vpdegree=1 if pdegree==1|pdegree==./\*valid\*/

replace vpdegree=2 if pdegree!=1 & pdegree!=./\*not valid\*/

tab vpdegree

tab fice if vpdegree==2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid CGPA\*

/\*Based upon a 4.0 scale as defined or converted to 4.0 by the institution.

Cumulative Grade Point Average is blank when the GPA has not been converted to 4.0 scale

or when the student has none available (e.g. taken only remedial courses, taken

only pass/fail courses, had all incompletes, exclusively auditors or withdrawn from

all courses). If the field is zero-filled, this is interpreted to mean the student

has a zero cumulative GPA (i.e. failed all courses attempted).\*/

gen vcgpa=.

replace vcgpa=1 if cumgpa>=0.00 & cumgpa<=4.99/\*valid\*/

replace vcgpa=2 if cumgpa==./\*valid, report high levels of blanks\*/

replace vcgpa=3 if (cumgpa<0.00 | cumgpa>5.00) & cumgpa!=. /\*not valid\*/

tab vcgpa

tab fice vcgpa if vcgpa>1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Valid Cum Native Credits Earned\*

/\*The current number of credit hours completed in credit courses at the reporting

institution (excluding transfer credits) as of the collection period at a grade

level satisfactory for degree requirements.\*/

gen vnatcr=.

replace vnatcr=1 if cncrhre >=3 & cncrhre<=999 /\*valid\*/

replace vnatcr=2 if cncrhre==. /\*not valid\*/

replace vnatcr=3 if cncrhre<3 | cncrhre>999 /\*valid, report high levels of less than 3 more than 999\*/

tab vnatcr

tab fice vnatcr if vnatcr>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Valid Cum Degree Credits Awarded\*

/\*Includes all credit hours awarded by the reporting institution that can be applied

toward the requirements for a formal award. This includes native credit, transfer

credit, credit by examination, and all other forms of credit.\*/

gen vndegcr=.

replace vndegcr=1 if ccrhra>0/\*valid\*/

replace vndegcr=2 if ccrhra==./\*not valid\*/

replace vndegcr=3 if ccrhra==0/\*valid, report high levels of 0\*/

replace vndegcr=4 if ccrhra<0/\*not valid\*/

tab vndegcr

tab fice vndegcr if vndegcr>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Valid Credits Required for Award\*

/\*The number of credit hours required to earn the award as specified in the catalogue

that applies to the student. Do not adjust this figure to account for circumstances

of the specific student (e.g., waiver, transfers).

A blank is allowed only when credits are undetermined for a degree sought is higher

than a Master's degree. This applies typically to medical programs and doctoral/professional

level programs. Credits that can be determined for a master's program should be submitted.\*/

/\*10, -- 12 credit min

 20, -- 60 credit min

 30, -- 12 credit min

 40, -- 120 min

 50, -- 12 credit min

 60, -- any or blank

 65, -- any or blank

 70, -- any or blank

 81, -- any or blank

 85, -- any or blank

 86, -- any or blank

 99.-- 120 min\*/

gen vrqcrhr=.

replace vrqcrhr=1 if (rqcrhr>=60 & degree==20) /\*valid\*/

replace vrqcrhr=5 if (rqcrhr<60|rqcrhr==.) & degree==20 /\*not valid\*/

replace vrqcrhr=2 if (rqcrhr>=120 & degree==40) /\*valid\*/

replace vrqcrhr=6 if (rqcrhr<120|rqcrhr==.) & degree==40 /\*not valid\*/

replace vrqcrhr=3 if rqcrhr>0 & degree>60/\*valid for graduate degrees\*/

replace vrqcrhr=4 if rqcrhr==. & degree>60/\*valid for graduate degrees\*/

replace vrqcrhr=7 if rqcrhr>=12& (degree==10|degree==30|degree==50) /\*valid\*/

replace vrqcrhr=8 if (rqcrhr<12|rqcrhr==.) & (degree==10|degree==30|degree==50) /\* not valid\*/

tab vrqcrhr

tab fice vrqcrhr if vrqcrhr==5|vrqcrhr==6|vrqcrhr==8

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*First Majors\*

\*test all degree=99 have entry in first major link\*

gen vlink=.

replace vlink=1 if (degree==99 & majorlk!=.) /\*99s have first major link\*/

replace vlink=2 if (degree==99 & majorlk==.) /\*not valid - degree = 99 but no code in first major link\*/

replace vlink=3 if (majorlk!=. & degree!=99) /\*not valid - code in first major link but with wrong degree\*/

tab fice if vlink>=2& vlink!=. /\*first major link errors\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*First Majors\*

\*test code has correct degree\*

\*Step 1: disaggregrate firstmajorlink to isolate degree\*

\*Step 2: test to determine if all fmdeg are 40 or 60\*

\*compare counts of 40s and 60s to last year totals\*

\*step 1: convert to string\*

tostring majorlk, gen(fmjrd)

\*step 2: parse code where first number is starting position, and second number is number of characters to keep\*

\*so, start with the first number and keep 2 digits\*

\*this pulls out the degree sought code from the first\_major string\*

gen fmdeg=substr(fmjrd, 1, 2)

destring fmdeg, replace

\*then encode all missing values(9=not applicable) as unit record is not a double

\*major and so first-major field is correctly blank\*

mvencode fmdeg, mv(9)

gen vfmdeg=.

replace vfmdeg=1 if fmdeg==40|fmdeg==60|fmdeg==9

replace vfmdeg=2 if fmdeg<40& fmdeg!=9

replace vfmdeg=3 if fmdeg>60

replace vfmdeg=4 if (fmdeg>40 & fmdeg<60)

tab vfmdeg

tab fice vfmdeg if vfmdeg>=2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*First Majors\*

\*test code has valid HEGIS\*

\*Step 1: disaggregrate firstmajorlink to isolate its Program Taxonomy (fmpt)

\*Step 2: test to determine if all fmpt are valid program taxonomy codes

\*start with the third digit and keep that digit and 3 additional)

\*this pulls out the program taxonomy code from the first\_major string\*

gen fmpt=substr(fmjrd,3, 4)

\*finally, destring to reinstate numbers for validation and code blanks as 9\*

destring fmpt, replace

mvencode fmpt, mv(9)

\*valid program taxonony code is correct\*

gen vfmpt=.

replace vfmpt=1 if fmpt==101 | fmpt==102 | fmpt==103 | fmpt==104 | fmpt==105 | fmpt==106 | ///

fmpt==107 | fmpt==108 | fmpt==109 | fmpt==110 | fmpt==111 | fmpt==112 | fmpt==113 | ///

fmpt==114 | fmpt==115 | fmpt==116 | fmpt==117 | fmpt==199 | ///

fmpt==201 | fmpt==202 | fmpt==203 | fmpt==204 | fmpt==205 | fmpt==206 | fmpt==299 | ///

fmpt==301 | fmpt==302 | fmpt==303 | fmpt==304 | fmpt==305 | fmpt==306 | fmpt==307 | ///

fmpt==308 | fmpt==309 | fmpt==310 | fmpt==311 | fmpt==312 | fmpt==313 | fmpt==314 | fmpt==399 | ///

fmpt==401 | fmpt==402 | fmpt==403 | fmpt==404 | fmpt==405 | fmpt==406 | fmpt==407 | ///

fmpt==408 | fmpt==409 | fmpt==410 | fmpt==411 | fmpt==412 | fmpt==413 | fmpt==414 | fmpt==415 | ///

fmpt==416 | fmpt==417 | fmpt==418 | fmpt==419 | fmpt==420 | fmpt==421 | fmpt==422 | fmpt==423 | ///

fmpt==424 | fmpt==425 | fmpt==426 | fmpt==427 | fmpt==499 | ///

fmpt==501 | fmpt==502 | fmpt==503 | fmpt==504 | fmpt==505 | fmpt==506 | fmpt==507

replace vfmpt=1 if fmpt==508 | fmpt==509 | fmpt==510 | fmpt==511 | fmpt==512 | fmpt==513 | fmpt==514 | fmpt==515 | ///

fmpt==516 | fmpt==517 | fmpt==599 | ///

fmpt==601 | fmpt==602 | fmpt==603 | fmpt==604 | fmpt==605 | fmpt==699 | ///

fmpt==701 | fmpt==702 | fmpt==703 | fmpt==704 | fmpt==705 | fmpt==799 | ///

fmpt==801 | fmpt==802 | fmpt==803 | fmpt==804 | ///

fmpt==805 | fmpt==806 | fmpt==807 | fmpt==808 | ///

fmpt==809 | fmpt==810 | fmpt==811 | ///

fmpt==812 | fmpt==813 | fmpt==814 | fmpt==815 | ///

fmpt==816 | fmpt==817 | fmpt==818 | fmpt==819 | ///

fmpt==820 | fmpt==821 | fmpt==822 | fmpt==823 | fmpt==824 | ///

fmpt==825 | fmpt==826 | fmpt==827 | fmpt==828 | ///

fmpt==829 | fmpt==830 | fmpt==831 | fmpt==832 | ///

fmpt==833 | fmpt==834 | fmpt==835 | fmpt==836 | ///

fmpt==837 | fmpt==838 | fmpt==839 | fmpt==899

replace vfmpt=1 if fmpt==901 | fmpt==902 | ///

fmpt==903 | fmpt==904 | fmpt==905 | fmpt==906 | ///

fmpt==907 | fmpt==908 | fmpt==909 | fmpt==910 | ///

fmpt==911 | fmpt==912 | fmpt==913 | fmpt==914 | ///

fmpt==915 | fmpt==916 | fmpt==917 | fmpt==918 | ///

fmpt==919 | fmpt==920 | fmpt==921 | fmpt==922 | ///

fmpt==923 | fmpt==924 | fmpt==925 | fmpt==999 | ///

fmpt==1001 | fmpt==1002 | fmpt==1003 | fmpt==1004 | ///

fmpt==1005 | fmpt==1006 | fmpt==1007 | fmpt==1008 | ///

fmpt==1009 | fmpt==1010 | fmpt==1011 | fmpt==1099

replace vfmpt=1 if fmpt==1101 | fmpt==1102 | ///

fmpt==1103 | fmpt==1104 | fmpt==1105 | fmpt==1106 | ///

fmpt==1107 | fmpt==1108 | fmpt==1109 | fmpt==1110 | ///

fmpt==1111 | fmpt==1112 | fmpt==1113 | ///

fmpt==1114 | fmpt==1115 | fmpt==1116 | fmpt==1199 | fmpt==1201 | ///

fmpt==1202 | fmpt==1203 | fmpt==1204 | fmpt==1205 | fmpt==1206 | ///

fmpt==1207 | fmpt==1208 | fmpt==1209 | fmpt==1210 | ///

fmpt==1211 | fmpt==1212 | fmpt==1213 | fmpt==1214 | ///

fmpt==1215 | fmpt==1216 | fmpt==1217 | fmpt==1218 | ///

fmpt==1219 | fmpt==1220 | fmpt==1221 | fmpt==1222 | ///

fmpt==1223 | fmpt==1224 | fmpt==1225 | fmpt==1299 | ///

fmpt==1301 | fmpt==1302 | fmpt==1303 | fmpt==1304 | ///

fmpt==1305 | fmpt==1306 | fmpt==1307 | fmpt==1399

replace vfmpt=1 if fmpt==1401 | fmpt==1499 | ///

fmpt==1501 | fmpt==1502 | fmpt==1503 | ///

fmpt==1504 | fmpt==1505 | fmpt==1506 | fmpt==1507 | ///

fmpt==1508 | fmpt==1509 | fmpt==1510 | fmpt==1599 | ///

fmpt==1601 | fmpt==1699 | ///

fmpt==1701 | fmpt==1702 | fmpt==1703 | fmpt==1799 | ///

fmpt==1801 | fmpt==1802 | fmpt==1803 | fmpt==1899

replace vfmpt=1 if fmpt==1901 | fmpt==1902 | fmpt==1903 | fmpt==1904 | fmpt==1905 | ///

fmpt==1906 | fmpt==1907 | fmpt==1908 | fmpt==1909 | ///

fmpt==1910 | fmpt==1911 | fmpt==1912 | fmpt==1913 | ///

fmpt==1914 | fmpt==1915 | fmpt==1916 | fmpt==1917 | ///

fmpt==1918 | fmpt==1919 | fmpt==1920 | fmpt==1999 | ///

fmpt==2001 | fmpt==2002 | fmpt==2003 | ///

fmpt==2004 | fmpt==2005 | fmpt==2006 | fmpt==2007 | ///

fmpt==2008 | fmpt==2009 | fmpt==2010 | fmpt==2099 | ///

fmpt==2101 | fmpt==2102 | fmpt==2103 | ///

fmpt==2104 | fmpt==2105 | fmpt==2106 | fmpt==2199

replace vfmpt=1 if fmpt==2201 | fmpt==2202 | fmpt==2203 | fmpt==2204 | ///

fmpt==2205 | fmpt==2206 | fmpt==2207 | fmpt==2208 | ///

fmpt==2209 | fmpt==2210 | fmpt==2211 | fmpt==2212 | ///

fmpt==2213 | fmpt==2214 | fmpt==2215 | fmpt==2299 | fmpt==2301 | ///

fmpt==2302 | fmpt==2303 | fmpt==2304 | fmpt==2399

replace vfmpt=1 if fmpt==4901 | fmpt==4902 | fmpt==4903 | fmpt==4904 | fmpt==4910 | ///

fmpt==4920 | fmpt==4930 | fmpt==4940 | fmpt==4950 | ///

fmpt==4960 | fmpt==4970 | fmpt==4980 | fmpt==4999 | ///

fmpt==5001 | fmpt==5002 | fmpt==5003 | ///

fmpt==5004 | fmpt==5005 | fmpt==5006 | fmpt==5007 | fmpt==5008 | ///

fmpt==5009 | fmpt==5010 | fmpt==5011 | fmpt==5012 | fmpt==5099

replace vfmpt=1 if fmpt==5101 | fmpt==5102 | ///

fmpt==5103 | fmpt==5104 | fmpt==5105 | fmpt==5199 | ///

fmpt==5201 | fmpt==5202 | ///

fmpt==5203 | fmpt==5204 | fmpt==5205 | fmpt==5206 | ///

fmpt==5207 | fmpt==5208 | fmpt==5209 | fmpt==5210 | ///

fmpt==5211 | fmpt==5212 | fmpt==5213 | fmpt==5214 | fmpt==5215 | ///

fmpt==5216 | fmpt==5217 | fmpt==5218 | fmpt==5219 | fmpt==5299

replace vfmpt=1 if fmpt==5301 | fmpt==5302 | fmpt==5303 | ///

fmpt==5304 | fmpt==5305 | fmpt==5306 | fmpt==5307 | fmpt==5308 | ///

fmpt==5309 | fmpt==5310 | fmpt==5311 | fmpt==5312 | fmpt==5313 | ///

fmpt==5314 | fmpt==5315 | fmpt==5316 | fmpt==5317 | fmpt==5399 | ///

fmpt==5401 | fmpt==5402 | fmpt==5403 | fmpt==5404 | ///

fmpt==5405 | fmpt==5406 | fmpt==5407 | fmpt==5408 | fmpt==5499 | ///

fmpt==5501 | fmpt==5502 | fmpt==5503 | ///

fmpt==5504 | fmpt==5505 | fmpt==5506 | fmpt==5507 | fmpt==5508 | fmpt==5599 | fmpt==5601 | fmpt==9099

replace vfmpt=2 if fmpt==9 /\*no code as record is not a first major link\*/

mvencode vfmpt, mv(9) /\*any program taxonomy not in list above code as 9\*/

tab vfmpt /\*1 is valid, 9 is not in API, 2 is not a first majorlink record\*/

tab fice vfmpt if vfmpt==9 /\*not valid due to blank or number not in API\*/

\*this file reviews frequency distributions for comparison to prior year files\*

\*identify variations of 10% and review for errors\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

table citizen

table sex

table fice if degree==10|degree==30

table fice if degree==20

table fice if degree==40

table fice if degree==50

table fice if degree==60

table fice degree if degree>60&degree<99

table fice degree if fmdeg==40

table fice degree if fmdeg==60

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*separate population to ugrad and grad\*

gen ugradgrad=.

replace ugradgrad=1 if degree==10|degree==20|degree==30|degree==40

replace ugradgrad=2 if degree>=50 & degree!=.

replace ugradgrad=3 if degree==47|degree==0

label define ugradgrad 1 "ugrad" 2 "grad" 3 "other"

label values ugradgrad ugradgrad

tab ugradgrad

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*recode for multi race\*

gen rmulti=.

replace rmulti=1 if (rwhite==1 & rblack==2 & rasian==3 & raian==4 & rnhpi==5 & hislat!=2)

replace rmulti=1 if (rwhite==1 & rblack==2 & hislat!=2)

replace rmulti=1 if (rwhite==1 & rasian==3 & hislat!=2)

replace rmulti=1 if (rwhite==1 & raian==4 & hislat!=2)

replace rmulti=1 if (rwhite==1 & rnhpi==5 & hislat!=2)

replace rmulti=1 if (rblack==2 & rasian==3 & hislat!=2)

replace rmulti=1 if (rblack==2 & raian==4 & hislat!=2)

replace rmulti=1 if (rblack==2 & rnhpi==5 & hislat!=2)

replace rmulti=1 if (rasian==3 & raian==4 & hislat!=2)

replace rmulti=1 if (rasian==3 & rnhpi==5 & hislat!=2)

replace rmulti=1 if (raian==4 & rnhpi==5 & hislat!=2)

tab rmulti

\*recode for unknown race\*

gen runknown=.

replace runknown=2 if rwhite==. & rblack==. & rasian==. & raian==. & rnhpi==. & hislat!=2

tab runknown

\*\*distribution of population by race\*\*

tab rwhite if hislat!=2 & rmulti!=1

tab rblack if hislat!=2 & rmulti!=1

tab rasian if hislat!=2 & rmulti!=1

tab raian if hislat!=2 & rmulti!=1

tab rnhpi if hislat!=2 & rmulti!=1

tab hispanic

tab hislat

tab runknown

tab rmulti

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

gen eyear=.

replace eyear = 1 if entyr <=1999 /\*review large distribtions of entry years prior to 1999\*/

replace eyear = 2 if entyr >=2000 & entyr <=2010

replace eyear = 3 if entyr >=2011

tab eyear

tab fice eyear

tab entrm

tab pdflg

tab degree if pdflg=="Y"

tab revtrn

tab degree if revtrn=="Y"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

tab idt

tab majorlk if majorlk!=.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*age groupings\*

\*step 1: convert to string\*

tostring birthdt, gen(birthyear)

\*step 2: start with the first number and keep 4 digits\*

\*this pulls out the year from the birth date string\*

gen byear=substr(birthyear, 1, 4)

destring byear, replace

gen age=(colyr-byear)

gen age\_grp=.

replace age\_grp=1 if age<25

replace age\_grp=2 if age>=25 & age <=45

replace age\_grp=3 if age>45 & age <=65

replace age\_grp=4 if age>65

tab age\_grp

tab fice age\_grp

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*time to degree\*

\*is entry year aligned to graduation date?\*/

gen time=(colyr-entyr)

gen time2=.

replace time2=1 if (time>=0&time<=2) & degree==20 /\*100% of timeframe\*/

replace time2=2 if (time>2&time<=4) & degree==20 /\*200% of timeframe\*/

replace time2=3 if (time>4&time<=8) & degree==20 /\*400% of timeframe\*/

replace time2=4 if time>8 & degree==20 /\*400+% of timeframe\* these may be bad entry years\*/

replace time2=5 if (time>=0&time<=4) & degree==40 /\*100% of timeframe\*/

replace time2=6 if (time>4&time<=8) & degree==40 /\*200% of timeframe\*/

replace time2=7 if (time>8&time<=12) & degree==40 /\*300% of timeframe\*/

replace time2=8 if time>12 & degree==40 /\*300+% of timeframe\* these may be bad entry years\*/

mvdecode time2, mv(9) /\*certs and graduate degrees\*/

tab time2

tab fice if time2==4 /\*400+% of timeframe\*these may be bad entry years\*/

tab fice if time2==8 /\*300+% of timeframe\*these may be bad entry years\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Cum GPAs\*

gen cgpadist=.

replace cgpadist=0 if cumgpa==0

replace cgpadist=1 if (cumgpa>0 & cumgpa<2)

replace cgpadist=2 if (cumgpa>=2 & cumgpa<3)

replace cgpadist=3 if (cumgpa>=3 & cumgpa<4)

replace cgpadist=4 if cumgpa==4

replace cgpadist=5 if cumgpa>4

replace cgpadist=9 if cumgpa==.

tab cgpadist

tab fice cgpadist if cgpadist<=1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Cum Native Credits Earned\*

gen cearn=.

replace cearn=1 if cncrhre==.

replace cearn=2 if cncrhre==0

replace cearn=3 if (cncrhre>=.001&cncrhre<=30.99)

replace cearn=4 if (cncrhre>=31&cncrhre<=60.99)

replace cearn=5 if (cncrhre>=61&cncrhre<=90.99)

replace cearn=6 if (cncrhre>=91&cncrhre<=119.99)

replace cearn=7 if cncrhre>=120

replace cearn=8 if cncrhre<0

tab cearn

tab cearn if degree==20

tab cearn if degree==40

tab fice cearn if cearn<=2&degree==20

tab fice cearn if cearn<=4&degree==40

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*Cum Credits Awarded\*

gen caward=.

replace caward=1 if ccrhra==.

replace caward=2 if ccrhra==0

replace caward=3 if (ccrhra>=.001&ccrhra<=30.99)

replace caward=4 if (ccrhra>=31&ccrhra<=60.99)

replace caward=5 if (ccrhra>=61&ccrhra<=90.99)

replace caward=6 if (ccrhra>=91&ccrhra<=119.99)

replace caward=7 if ccrhra>=120

replace caward=8 if ccrhra<0

tab caward

tab caward if degree==20

tab caward if degree==40

tab fice caward if caward<=2&degree==20

tab fice caward if caward<=4&degree==40

\*this file includes logic checks across data elements\*

\*test distribution of NON\_SSN by FICE\*

\*identify FICE with high frequency of non-SSN\*

table fice if idt==2

table fice idt if idt==2 & citizen==1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*is entry year no more than 400% of normal timeframe\*

\*400% allows for normal progression for part-time enrollment\*

\*high frequency of 400% may indicated bad entry year\*

gen yrlogic=(colyr-entyr)

gen yrlogic2=.

replace yrlogic2=1 if (yrlogic>=0& yrlogic<=16) & degree==40 /\*400% of timeframe\*/

replace yrlogic2=2 if (yrlogic>=0& yrlogic<=8) & degree==20 /\*400% of timeframe\*/

replace yrlogic2=3 if (yrlogic>=0& yrlogic<=6) & (degree==10|degree==30) /\*400% of timeframe\*/

replace yrlogic2=4 if degree>=50 /\*out of scope\*/

mvencode yrlogic2, mv(9) /\*longer than 400%\*/

tab yrlogic2

tab fice degree if yrlogic2<4

tab fice degree if yrlogic2==9

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*GPA at least 2.0 for ugrads and at least 3.0 for grads

gen mingpa=.

replace mingpa=1 if (cumgpa>=2&degree <=40)

replace mingpa=2 if (cumgpa>=3&degree >=50)

replace mingpa=3 if (cumgpa<2&degree<=40)|(cumgpa==.&degree<=40)

replace mingpa=4 if ((cumgpa<3&degree>=50)|(cumgpa==.&degree>=50))&degree!=99

tab mingpa

tab fice degree if mingpa==3

tab fice degree if mingpa==4

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*credits awarded (ccrhra) are equal to or greater than credits required for degree - rqcrhr\*

/\*Includes all credit hours awarded by the reporting institution that can be applied toward

\*the requirements for a formal award. This includes native credit, transfer credit, credit by

\*examination, and all other forms of credit.\*/

gen vearnreq=.

replace vearnreq=1 if rqcrhr==ccrhra /\*valid\*/

replace vearnreq=2 if (ccrhra>rqcrhr & rqcrhr!=.) & degree<=40 /\*valid\*/

replace vearnreq=3 if (ccrhra<rqcrhr)&degree<=40 /\*not valid\*/

replace vearnreq=4 if rqcrhr!=ccrhra & degree>=50 /\*valid for grad degrees\*/

tab vearnreq

tab fice degree if vearnreq==3

tab fice if vearnreq==3

\*credits awarded (ccrhra) are equal to or greater than credits earned for degree - cncrhre\*

/\*awarded includes all credit hours awarded by the reporting institution that can be applied toward

\*the requirements for a formal award. This includes native credit, transfer credit, credit by

\*examination, and all other forms of credit. earned is credit earned at the institution\*/

/\*awarded cannot be less than earned as awarded includes all credits earned\*/

gen vnataw=.

replace vnataw=1 if ccrhra>=cncrhre & degree<=40 /\*valid\*/

replace vnataw=2 if ccrhra<cncrhre & degree<=40/\*not valid\*/

replace vnataw=3 if degree>=50 /\*valid for grad degrees\*/

tab vnataw

tab fice vnataw if vnataw==2

\*test for excess credits\*

\*review records where there is more than 45 credits awarded beyond what is required\*

gen excess=ccrhra-rqcrhr

tab fice if excess>45 & degree!=99

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*review all students with degree from community college\*

\*determine if any have record in EIS from four-year institution in period prior to degree award\*

\*also verify if degree at 2 year, no degree at 4 year in same file\*

sort tssn segment

egen unique=tag(tssn) /\*this preferences community college\*/

tab unique

gen reverse=.

replace reverse=1 if revtrn=="Y"

replace reverse=2 if revtrn!="Y"

tab fice

tab fice reverse

tab unique reverse

tab unique reverse if segment!=1

\*save data set as RTF\*

keep if segment==1

drop sic subcamp idt campusid sasid program\_orig-entyr rqcrhr-source\_file majords-vrqcrhr

tab fice

tab fice revtrn

sort fice tssn

egen dis\_unique=tag(tssn) /\*unique to file\*/

tab dis\_unique

gsort fice tssn -degree

egen disfice\_unique=tag(fice tssn)/\*unique to college\*/

tab disfice\_unique

tab disfice\_unique reverse

drop if (disfice\_unique==0 & reverse==2)

\*save\*

\*open pooled EIS data set\*

destring, replace

\*construct sector and segment\*

\*step 1: convert to string\*

gen sic2=sic

tostring sic2, replace

\*step 2: sector - parse code to keep only first digit\*

gen sector=substr(sic2, 1, length(sic2)-5)

destring sector, replace

\*step 4: segment - parse code to keep only second digit\*

gen segment=substr(sic2, 2, length(sic2)-5)

destring segment, replace

\*step 3: drop sic2\*

drop sic2

drop if segment==1

drop if colyr>=XXX /\*change to remove all years/terms later than DIS grad dates\*/

drop if coltm==xxx /\*change to remove all years/terms later than DIS grad dates\*/

drop sic subcamp idt-program tcrhra-satebr16

sort tssn

egen eis\_unique=tag(tssn)

tab eis\_unique

drop if eis\_unique==0

\*save\*

\*merge using many (DIS) to one (EIS) on TSSN\*

drop if \_merge==2 /\*keep \_merge==1 no match in EIS to record in DIS and \_merge==3 matched between DIS and EIS\*/

tab reverse \_merge

tab fice \_merge if reverse==1 /\*coded in DIS as reverse\*/

tab fice \_merge if reverse==2 /\*not coded in DIS as reverse\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*from main data file\*

\*multi degree only\*

tab degree

drop tssn2 idt campusid sasid teachcan-source\_file dups-vsasid2 vtax-vlink vfmdeg vfmpt eyear-vnataw vfmpt-vrqcrhr

drop if segment==1

drop if degree<40

drop if (degree>60&degree<99)

drop if degree==50

tab degree

sort fice idn degree /\*sorts lowest degree first so 99 is always second\*/

egen coll\_unique=tag(fice idn)

tab coll\_unique

\*test that all 99's have a 40 or 60 record\*

tab coll\_unique

tab degree coll\_unique /\*all 99s should be 0\*/

tab degree if degree==99

tab fice if coll\_unique==1&degree==99 /\*missing second record for 40 or 60\*/

\*construct var to replace 99 with second major degree code\*

tab fmdeg

gen degree2=degree

replace degree2=40 if fmdeg==40

replace degree2=60 if fmdeg==60

tab degree2

\*test if degree in first major = degree in first major link\*

\*n=0 or n=. if coded is same across both records\*

gsort fice idn -degree

bysort fice idn: generate degchange = degree2 - degree2[\_n-1]

tab degchange

tab fice if (degchange!=0|degchange!=.)& coll\_unique==0 & degree==99 /\*total should equal number of 99s\*/

\*test two: HEGIS in first major link field for 99 is HEGIS in program field for 40 or 60\*

\*construct major from first major link\*

\*first two digits are degree\*

\*next six digits are HEGIS\*

\*step 1: convert to string\*

tostring majorlk, gen(major)

\*step 2: Major, last six digits\*

gen major2=substr(major, 2,.)

destring major2, replace

replace major2=program\_orig if major2==.

\*test if HEGIS in first major = HEGIS in first major link\*

\*n=0 or n=. if coded is same across both records\*

gsort fice idn -degree

bysort fice idn: generate hegchange = major2 - major2[\_n-1]

tab hegchange

tab fice if (hegchange!=0|hegchange!=.)& coll\_unique==0 & degree==99 /\*total should equal number of 99s\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*pending logic checks on prior degree flag\*