SESUG Paper ED159-2019

Complete your IPEDS Completions in 60 minutes or less (using SAS® and Colleague®)

Yolanda N. Ingram, Wake Technical Community College, Raleigh, North Carolina

ABSTRACT

This paper is designed for Institutional Research and higher education professionals who are responsible for completing the federal IPEDS Completions Survey. You will learn how to complete your IPEDS Completions report in 60 minutes or less by using a SAS®, Colleague® and Notepad++.

The output is an electronic file that can be uploaded by the key holder to minimize the need for manual data entry and to improve efficiency and accurate reporting.

INTRODUCTION

The IPEDS (Integrated Postsecondary Education Data System) Completions Survey collects data of the number of students completing a postsecondary education program and the number of awards received from higher education institutions across the country (within a 12 month period). Submission of IPEDS data is mandatory for all institutions that participate in Federal financial assistance programs.

While, there are many ways to enter your completions data into the IPEDS system, this paper will focus on a six step process to turn your Completions output from Colleague® into an electronic file that can be easily uploaded into IPEDS using SAS® code. Once established, this process can greatly reduce the amount of time dedicated to finalize the data as well as decrease human errors.

To accomplish this process, you will need:

Colleague® Completions Output (screen mnemonic ICMP)
Notepad++ (regular Notepad will not work)
BASE SAS®
IPEDS Completions Survey Materials: Import Specifications (file data layout)

STEP 1: RETRIEVE IPEDS COMPLETIONS SURVEY DATA FROM COLLEAGUE®

Listed below are some quick instructions on how to retrieve the completions data from Colleague®.

A – Log into Colleague® and access the IPEDS Completions Survey (ICMP) screen.

B – Enter your university identifier.

C - Enter the Twelve-month Reporting Dates. (July 1st to June 30th)

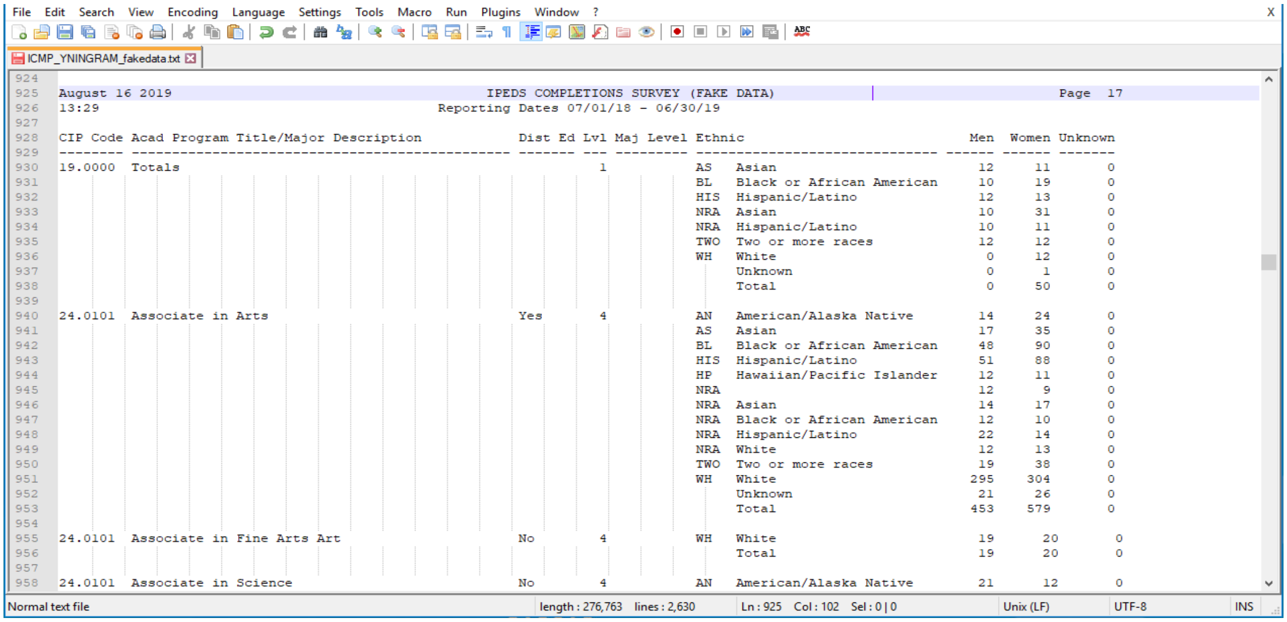
D - Enter degree and/or certificate types awarded by your institution.

E - Enter additional information to be included or excluded on the report: (Academic Program, CIP Codes, etc.)

F – Once the report is generated, right click the download button and save as .txt file.

STEP 2: OPEN .TXT FILE WITH NOTEPAD++

Open the saved ICMP Completers text file in Notepad++. (Please note: Using the regular Notepad text editor will not work as the formatting is not retained.)



```

August 16 2019                                       IPEDS COMPLETIONS SURVEY (FAKE DATA)                                   Page  17
13:29                                          Reporting Dates 07/01/18 - 06/30/19

CIP Code Acad Program Title/Major Description            Dist Ed Lvl Maj Level Ethnic                            Men  Women Unknown
-------- ----------------------------------------------- ------- --- --------- ------------------------------ ------ ------ -------
19.0000  Totals                                                    1           AS   Asian                         12     11       0
                                                                               BL   Black or African American     10     19       0
                                                                               HIS  Hispanic/Latino               12     13       0
                                                                               NRA  Asian                         10     31       0
                                                                               NRA  Hispanic/Latino               10     11       0
                                                                               TWO  Two or more races             12     12       0
                                                                               WH   White                          0     12       0
                                                                                    Unknown                        0      1       0
                                                                                    Total                          0     50       0

24.0101  Associate in Arts                               Yes       4           AN   American/Alaska Native        14     24       0
                                                                               AS   Asian                         17     35       0
                                                                               BL   Black or African American     48     90       0
                                                                               HIS  Hispanic/Latino               51     88       0
                                                                               HP   Hawaiian/Pacific Islander     12     11       0
                                                                               NRA                                12      9       0
                                                                               NRA  Asian                         14     17       0
                                                                               NRA  Black or African American     12     10       0
                                                                               NRA  Hispanic/Latino               22     14       0
                                                                               NRA  White                         12     13       0
                                                                               TWO  Two or more races             19     38       0
                                                                               WH   White                        295    304       0
                                                                                    Unknown                       21     26       0
                                                                                    Total                        453    579       0

24.0101  Associate in Fine Arts Art                      No        4           WH   White                         19      20       0
                                                                                    Total                         19      20       0

24.0101  Associate in Science                            No        4           AN   American/Alaska Native        21      12       0
```

Display 1. Sample data file using Notepad++ text editor

There are three main components in the Colleague Completions text file. We will use the three portions to complete Parts A-D in the final IPEDS export file layout.

We will read in each component separately during Step 3 and will separate the data into Parts A-D in Step 4.

- **IPEDS COMPLETIONS SURVEY** (used in Part A and Part B)
- **IPEDS COMPLETIONS - ALL COMPLETERS** (used in Part C -normally on the next to the last page of the file)
- **IPEDS COMPLETIONS - BY LEVEL** (Part D -normally on the last page of the file)

STEP 3: IMPORT COMPLETIONS SURVEY DATA FROM .TXT FILE INTO SAS

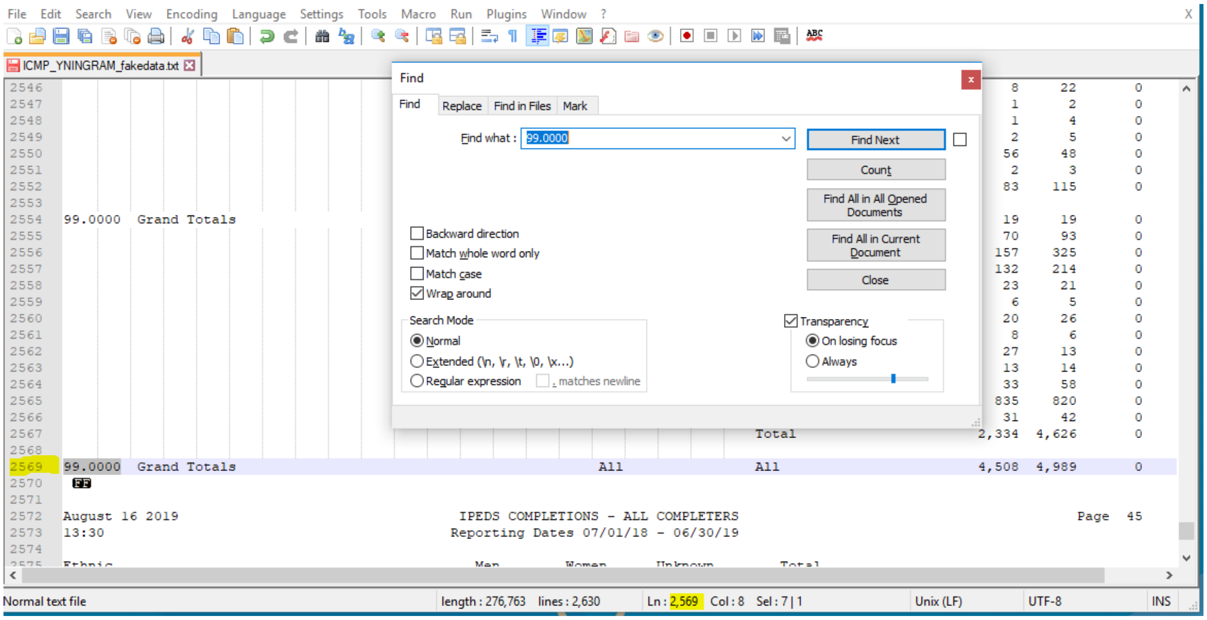
We are going to read in each of the three main components (listed above) into SAS using simple Infile and Input statements.

We will tackle the **IPEDS COMPLETIONS SURVEY** component first.

For the obs= option on the **INFILE** statement, we want to find the line number associated with the last line of the IPEDS COMPLETIONS SURVEY.

This should be on directly before the *IPEDS COMPLETIONS - ALL COMPLETERS* section.

If you cannot find it, you can select CNTL and hit the F key, and search for '99.0000'. Find the last instance of 99.0000. It should have the word 'All' to the right. In our example the line is 2569.



Find what : 99.0000

```
2546                                                      8     22      0
2547                                                      1      2      0
2548                                                      1      4      0
2549                                                      2      5      0
2550                                                     56     48      0
2551                                                      2      3      0
2552                                                     83    115      0
2553
2554 99.0000  Grand Totals                               19     19      0
2555                                                     70     93      0
2556                                                    157    325      0
2557                                                    132    214      0
2558                                                     23     21      0
2559                                                      6      5      0
2560                                                     20     26      0
2561                                                      8      6      0
2562                                                     27     13      0
2563                                                     13     14      0
2564                                                     33     58      0
2565                                                    835    820      0
2566                                                     31     42      0
2567                                   Total          2,334  4,626      0
2568
2569 99.0000  Grand Totals               All          All          4,508  4,989      0
2570  FF
2571
2572 August 16 2019              IPEDS COMPLETIONS - ALL COMPLETERS            Page  45
2573 13:30                       Reporting Dates 07/01/18 - 06/30/19
2574
```

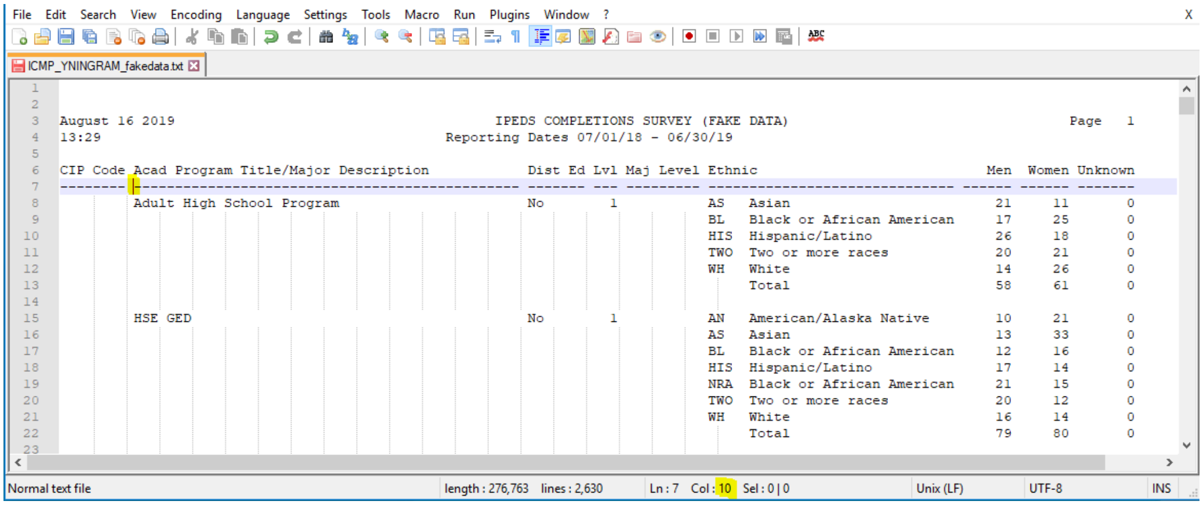
Normal text file   length : 276,763   lines : 2,630   Ln : 2,569   Col : 8   Sel : 7 | 1   Unix (LF)   UTF-8   INS

Display 2. Screenshot to find value for obs= option on INFILE statement

For the **INPUT** Statement: This is where Notepad++ comes in handy as it easily identifies the column and row.

There are 9 variables in the text file, but I decided to break up race code and race description into separate variables, so I have 10 variables on the input statement.

By placing your cursor at the beginning of each variable, you can quickly see that CIP Code starts at column 1, Acad Program Title starts at column 10 and so forth. Write down the starting points of each variable so that you can fill in the input statements starting positions (@x values). Also place the cursor at the end of each variable to determine the length.


File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
ICMP_YNINGRAM_fakedata.txt

```


August 16 2019                                IPEDS COMPLETIONS SURVEY (FAKE DATA)                  Page   1
13:29                                   Reporting Dates 07/01/18 - 06/30/19

CIP Code Acad Program Title/Major Description            Dist Ed Lvl Maj Level Ethnic                              Men  Women Unknown
-------- ----------------------------------------------- ------- --- --------- ------------------------------ ------ ------ -------
         Adult High School Program                       No        1           AS   Asian                          21     11       0
                                                                               BL   Black or African American      17     25       0
                                                                               HIS  Hispanic/Latino                26     18       0
                                                                               TWO  Two or more races              20     21       0
                                                                               WH   White                          14     26       0
                                                                                    Total                          58     61       0

         HSE GED                                         No        1           AN   American/Alaska Native         10     21       0
                                                                               AS   Asian                          13     33       0
                                                                               BL   Black or African American      12     16       0
                                                                               HIS  Hispanic/Latino                17     14       0
                                                                               NRA  Black or African American      21     15       0
                                                                               TWO  Two or more races              20     12       0
                                                                               WH   White                          16     14       0
                                                                                    Total                          79     80       0
```

Normal text file   length : 276,763   lines : 2,630   Ln : 7   Col : 10   Sel : 0 | 0   Unix (LF)   UTF-8   INS

Display 3. Identify variable column starting points on INPUT statement for IPEDS Survey Component

Now let’s pull all of the information together and start coding!

We will start off by creating a few macro variables that will be used throughout the program.

```
%let UnitID  = xxxxxx;                    /* Institution 6 digit IPEDS number */
%let fileloc = C:\Documents\IPEDs Completions;
%let fname   = ICMP_YNINGRAM_fakedata.txt; /*name of Completions text file*/
%let outfile = IPEDS_Completions_2019_2020; /*name of final file to export*/
```

Next, we will read in the IPEDS COMPLETIONS SURVEY portion of the .txt file.

The title lines from each page of the report are still being read in with the actual data, so we needed to find a quick way to remove the junk data.

```
data IPEDS_Survey;

infile "&fileloc.\&fname" obs=2569  ;   /*Fill in last line number of SURVEY section*/


/*Fill in starting position and length of each variable using Notepad++ text editor */
input  @1   CIP                 $8.
       @10  Program_Title       $48.
       @58  Dist_ED             $8.
       @66  awlevel             $3.
       @70  majornum            $9.
       @80  race_code           $3.
       @85  race_des            $25.
       @111 Male_count          $6.
       @118 Female_count        $6.
       @125 Unknown             $6.;


*The title lines from each page of the report are still being read in,
 this is a quick way to get rid of the junk lines that we do not need ;

if dist_ed not in ('No','Yes',' ')  then delete;
run;
```

Now, it’s time to read in the second component of the .txt file**. IPEDS COMPLETIONS - ALL COMPLETERS**.

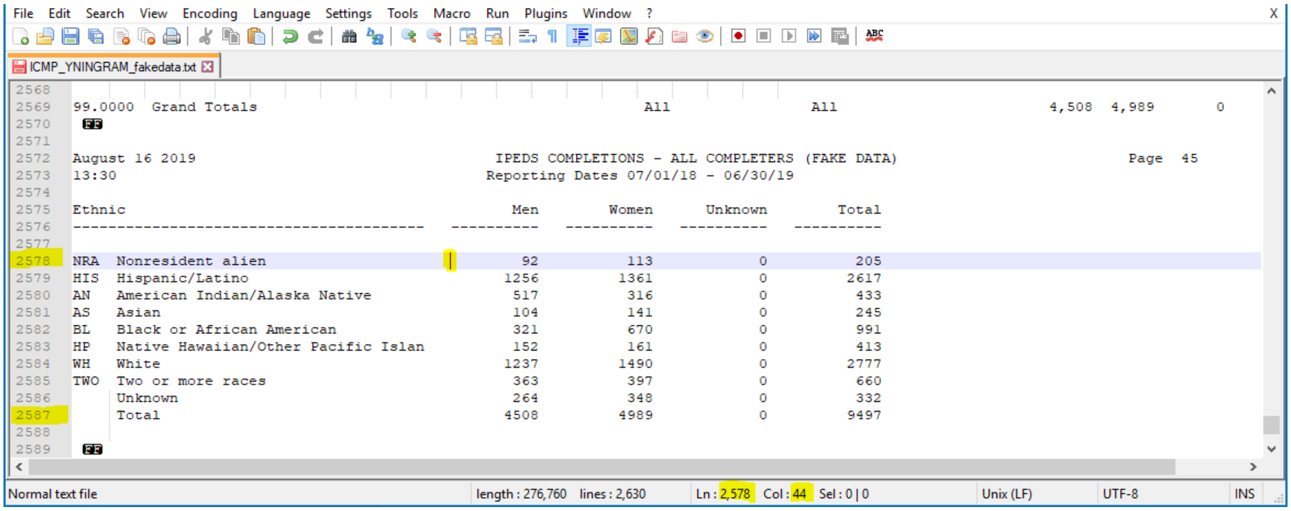
We will follow the same steps as we did to read in the IPEDS COMPLETIONS SURVEY portion, except this time we will use the FIRSTOBS= and OBS= options.

Select CNTL and hit the F key, and search for ‘ALL COMPLETERS’.   Since this is a summary portion, we want to only concentrate on the actual data. Locate the first line and last line of actual data.

In our example, it is 2578 and 2587. The values you find will be placed in the FIRSTOBS= and OBS= values.

Again, place your cursor at the beginning of each variable and write down the starting points of each variable so that you can fill in the input statements starting positions (@x values).

In this section there are five variables (Ethnic, Men, Women, Unknown and Total). For the purposes of this paper, divide Ethnic is separated into two separate variables for a total of six. (race_code and race_des).



```
2568
2569 99.0000  Grand Totals                                           All               All                        4,508  4,989       0
2570  FF
2571
2572 August 16 2019                                 IPEDS COMPLETIONS - ALL COMPLETERS (FAKE DATA)                         Page  45
2573 13:30                                         Reporting Dates 07/01/18 - 06/30/19
2574
2575 Ethnic                                            Men        Women      Unknown        Total
2576 -----------------------------------------  ----------   ----------   ----------   ----------
2577
2578 NRA  Nonresident alien                             92          113            0          205
2579 HIS  Hispanic/Latino                             1256         1361            0         2617
2580 AN   American Indian/Alaska Native                517          316            0          433
2581 AS   Asian                                        104          141            0          245
2582 BL   Black or African American                    321          670            0          991
2583 HP   Native Hawaiian/Other Pacific Islan          152          161            0          413
2584 WH   White                                       1237         1490            0         2777
2585 TWO  Two or more races                            363          397            0          660
2586      Unknown                                      264          348            0          332
2587      Total                                       4508         4989            0         9497
2588
2589  FF
```

Normal text file | length : 276,760 lines : 2,630 | Ln : 2,578 Col : 44 Sel : 0 | 0 | Unix (LF) | UTF-8 | INS

Display 4. **Identify variable column starting points for IPEDS ALL COMPLETERS Component**

```
data IPEDS_ALL_COMPLETERS;

 infile "&fileloc.\&fname" firstobs=2578  obs=2587 ;


/*Fill in starting position and length of each variable using Notepad++ text editor */
input
      @1   race_code        $3.
      @6   race_des         $25.
      @44 Male_count        $10.
      @57 Female_count      $10.
      @70 Unknown           $10.
      @83 Total             $10.   ;

run;
```

Now to read in the final component of the .txt file**:** **IPEDS COMPLETIONS – BY LEVEL**.

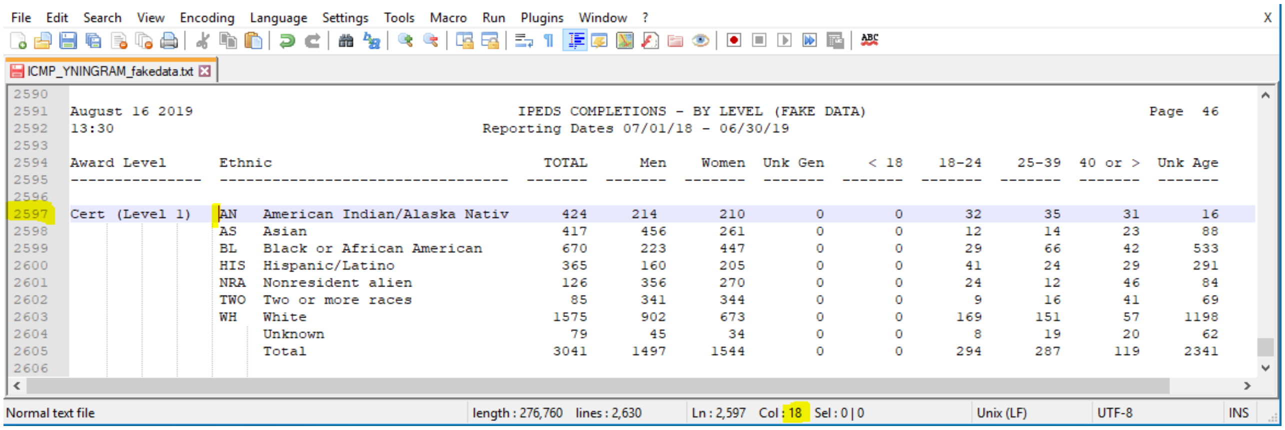
Repeat the process again to find the line numbers for the FIRSTOBS= and OBS= options.

Search for the phrase ‘BY LEVEL’ to find the correct section. Locate the first line and last line of actual data.

In our example, it is 2597 and 2628. The values you find will be placed in the FIRSTOBS= and OBS= values.

Again, place your cursor at the beginning of each variable and write down the starting points of each variable so that you can fill in the input statements starting positions (@x values).

In this section there are eleven variables in the .txt file. However, in the SAS code, Ethnic is separated into two separate variables for a total of twelve.


File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
X
ICMP_YNINGRAM_fakedata.txt
Normal text file
length : 276,760 lines : 2,630
Ln : 2,597 Col : 18 Sel : 0 | 0
Unix (LF)
UTF-8
INS

```
2590
2591  August 16 2019                                     IPEDS COMPLETIONS - BY LEVEL (FAKE DATA)                                Page  46
2592  13:30                                          Reporting Dates 07/01/18 - 06/30/19
2593
2594  Award Level      Ethnic                               TOTAL      Men    Women  Unk Gen     < 18    18-24    25-39  40 or >  Unk Age
2595  ---------------  ---------------------------------  -------  -------  -------  -------  -------  -------  -------  -------  -------
2596
2597  Cert  (Level 1)  AN   American Indian/Alaska Nativ      424      214      210        0        0       32       35       31       16
2598                   AS   Asian                             417      456      261        0        0       12       14       23       88
2599                   BL   Black or African American         670      223      447        0        0       29       66       42      533
2600                   HIS  Hispanic/Latino                   365      160      205        0        0       41       24       29      291
2601                   NRA  Nonresident alien                 126      356      270        0        0       24       12       46       84
2602                   TWO  Two or more races                  85      341      344        0        0        9       16       41       69
2603                   WH   White                            1575      902      673        0        0      169      151       57     1198
2604                        Unknown                            79       45       34        0        0        8       19       20       62
2605                        Total                            3041     1497     1544        0        0      294      287      119     2341
2606
```

Display 5. **Identify variable column starting points for IPEDS COMPLETIONS – BY LEVEL Component.**

```
data IPEDS_BY_LEVEL;

infile "&fileloc.\&fname" firstobs=2597  obs=2628 ;


/*Fill in starting position and length of each variable using Notepad++ text editor */

input
      @1     ct_level        $16.
      @18    race_code        $3.
      @23    race_des        $25.
      @53    total            $7.
      @62    Male_count       $7.
      @71    Female_count     $7.
      @80    Unknown          $7.
      @89    _LT_18           $7.
      @98    _18_24           $7.
      @107   _25_39           $7.
      @116   _40_GT           $7.
      @125    _UKN_AGE        $7.;
run;
```

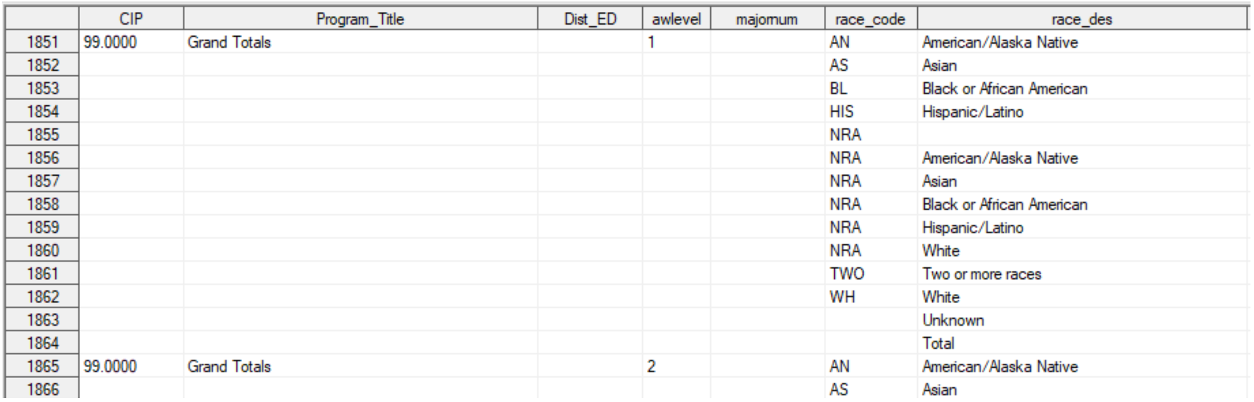
STEP 4: GET DATA IN SHAPE BASED ON IPEDS DATA SPECIFICATIONS

We now have our three components read into SAS and should have three datasets: IPEDS_SURVEY, IPEDS_ALL and IPEDS_BY_LEVEL.

The next step in the process is to do a little house cleaning. We need to clean each dataset to get it in shape for its final destination.

IPEDS_SURVEY:

Right now, the IPEDS_Survey SAS dataset is in pretty good shape, however currently, the CIP code, Program title and other variables are only populated on the first record for each program.



	CIP	Program_Title	Dist_ED	awlevel	majornum	race_code	race_des
1851	99.0000	Grand Totals		1		AN	American/Alaska Native
1852						AS	Asian
1853						BL	Black or African American
1854						HIS	Hispanic/Latino
1855						NRA	
1856						NRA	American/Alaska Native
1857						NRA	Asian
1858						NRA	Black or African American
1859						NRA	Hispanic/Latino
1860						NRA	White
1861						TWO	Two or more races
1862						WH	White
1863							Unknown
1864							Total
1865	99.0000	Grand Totals		2		AN	American/Alaska Native
1866						AS	Asian

Use the following macro to populate the variables on each record:

```
%macro fill_in(variable);
retain fill_&variable;
if not missing(&variable) then fill_&variable = &variable;

/* fills the new variable with non-missing value */
&variable = fill_&variable;

%mend;
```

Now for a little more cleaning and to call the macro above to fill in each row. We also need to define a race variable as defined in the IPEDS Survey Materials: Import Specifications.

Race/ethnicity table
1 - Nonresident alien
2 - Hispanic/Latino
3 - American Indian or Alaska Native
4 - Asian
5 - Black or African American
6 - Native Hawaiian or Other Pacific Islander
7 - White
8 - Two or more races
9 - Race and ethnicity unknown

```
data IPEDS_Survey_Clean;
set IPEDS_Survey;

*Add IPEDS race values;
if       race_code    = 'NRA'  then race=1;
else if race_code    = 'HIS'  then race=2;
else if race_code    = 'AN'   then race=3;
else if race_code    = 'AS'   then race=4;
else if race_code    = 'BL'   then race=5;
else if race_code    = 'HP'   then race=6;
else if race_code    = 'WH'   then race=7;
else if race_code    = 'TWO'  then race=8;
else race=9;  /*All others fall as Unknown */


/* This will write over where race des is total, so it doesn't count as unknown*/
if race_des = 'Total' then race=10;
if race_des = 'All'  then race=999;


*Run macro so that CIP and other var are on every observation;
%fill_in(CIP)
%fill_in(Program_Title)
%fill_in(Dist_ED)
%fill_in(awlevel)
%fill_in(majornum)

*remove commas from number counts;
male_count=compress(male_count,",");
female_count=compress(female_count,",");
unknown=compress(unknown,",");


drop fill_:; /* drop the retained variables */
run;
```

```
data IPEDS_ALL_CLEAN;
set IPEDS_ALL;

*Add IPEDS race values;
if       race_code    = 'NRA'  then race=1;
else if race_code    = 'HIS'  then race=2;
else if race_code    = 'AN'   then race=3;
else if race_code    = 'AS'   then race=4;
else if race_code    = 'BL'   then race=5;
else if race_code    = 'HP'   then race=6;
else if race_code    = 'WH'   then race=7;
else if race_code    = 'TWO'  then race=8;
else race=9;  /*All others fall as Unknown */
;

/* This will write over where race des is total, so it doesn't count as unknown*/
if race_des = 'Total' then race=10;
if race_des = 'Unknown'  then race=9;
run;
```

```
data IPEDS_BY_LEVEL_CLEAN;
set IPEDS_BY_LEVEL;

*Add IPEDS race values;
if      race_code    = 'NRA'  then race=1;
else if race_code    = 'HIS'  then race=2;
else if race_code    = 'AN'   then race=3;
else if race_code    = 'AS'   then race=4;
else if race_code    = 'BL'   then race=5;
else if race_code    = 'HP'   then race=6;
else if race_code    = 'WH'   then race=7;
else if race_code    = 'TWO'  then race=8;
else race=9;  /*All others fall as Unknown */
;
if race_des = 'Total' then do; race=10; race_code='TOT'; end;  /* This will
write over where race des is total, so it doesn't count as unknown*/
if race_des = 'Unknown'  then do; race=9; race_code = 'UKN'; end;

*Run macro so that CT_LEVEL and other var are on every observation;
%fill_in(ct_level);
drop fill_:; /* drop the retained variables */

run;
```

STEP 5: SPLIT INTO PARTS A-D

According to the IPEDS Survey Materials: Import Specifications, the final file needs to be split into four parts.

The cleaned IPEDS_SURVEY dataset creates Part A and Part B of the final file. IPEDS_ALL creates Part C and IPEDS_BY_LEVEL creates Part D.

We will need to sort, summarize and create a final file for each part.

PART A: COMPLETIONS – CIP DATA

```
proc sort data=IPEDS_Survey_Clean out=Part_A;
by CIP awlevel majornum race ;
run;


data Part_A_summary ;
  retain sum_male sum_female sum_unknown ;

/*Do not include Totals (CIP that end in .0000) in the counts */
  set Part_A (where=(Program_Title ne 'Totals'));
  by CIP awlevel majornum race ;

if first.race then do ;
         sum_male=Male_count*1 ;
         sum_female=female_count*1;
         sum_unknown=unknown*1 ;
end;
else do;
        sum_male=sum(sum_male,male_count)*1 ;
        sum_female=sum(sum_female,female_count)*1 ;
        sum_unknown=sum(sum_unknown,unknown)*1 ;
 end;
  if last.race then output ;
  drop program_title male_count female_count unknown;
run ;


*Now we need to have gender count on a separate line;
proc sort data=Part_A_summary out=Part_A_sort;
by CIP   awlevel majornum race race_code race_des;
run;


proc transpose data= Part_A_sort out=Part_A_long name=gender;
by CIP  awlevel majornum race race_code race_des;
var sum_male sum_female sum_unknown;
run;
```

We can now create the final dataset for Part A to mimic the sample file record:

```
DATA Part_A_final ;
LENGTH UNITID $13
 SURVSECT $13
 PART $8
 MAJORNUM $12
 CIPCODE $17
 AW_LEVEL $12
 RACE $9
SEX $7
COUNT $17 ;

SET part_A_long(where=(gender ne 'sum_unknown')
                rename=(race=race1));

*ASSIGN GENDER VALUES;
IF GENDER = 'sum_male' THEN SEX1=1;
IF GENDER = 'sum_female' THEN SEX1=2;

UNITID="UNITID=&UNITID";
SURVSECT=",SURVSECT=COM";
PART=",PART=A";
MAJORNUM=",MAJORNUM=1";
CIPCODE=CATS(',CIPCODE=',CIP);
AW_LEVEL=CATS(',AWLEVEL=',AWLEVEL);
RACE=CATS(',RACE=',RACE1);
SEX=CATS(',SEX=',SEX1);
count=CATS(',COUNT=',COL1);

KEEP UNITID--COUNT;

if awlevel='All' then delete;

*may need to delete if race=10;
RUN;
```

```
PART A: Sample record
UNITID=nnnnnn,SURVSECT=COM,PART=A,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=1,RACE=1,SEX=1,COUNT=nnnnn
UNITID=nnnnnn,SURVSECT=COM,PART=A,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=1,RACE=1,SEX=2,COUNT=nnnnn
UNITID=nnnnnn,SURVSECT=COM,PART=A,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=2,RACE=2,SEX=1,COUNT=nnnnn
UNITID=nnnnnn,SURVSECT=COM,PART=A,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=3,RACE=3,SEX=2,COUNT=nnnnn
```

Display 6. PART A: Sample record (from IPEDS Survey Materials: Import Specifications)

PART B: COMPLETIONS – DISTANCE EDUCATION

We will follow a similar pattern as we did in creating the final files for Part A. We first need to sort the data and get the data into the required format.

```
proc sort data=IPEDS_Survey_Clean out=Part_B;
by CIP awlevel majornum  dist_ed  ;
run;


data Part_B_summary ;
    /*Do not include Totals (CIP that end in .0000) in the counts */
  set Part_B (where=(Program_Title ne 'Totals'));
  by CIP  awlevel majornum dist_ed  ;

  if last.dist_ed;

  if dist_ed='Yes' then DE=1;
  else if dist_ed='No' then DE = 2;

  keep CIP  majornum awlevel dist_ed  DE;
run ;
```

```
DATA Part_B_final ;
LENGTH      UNITID      $13
            SURVSECT    $13
            PART        $8
            MAJORNUM    $12
            CIPCODE     $17
            AW_LEVEL    $12
            DistanceED  $14;

SET part_B_Summary;


UNITID="UNITID=&UNITID";
SURVSECT    =",SURVSECT=COM";
PART        =",PART=B";
MAJORNUM    =",MAJORNUM=1";
CIPCODE     =CATS(',CIPCODE=',CIP);
AW_LEVEL    =CATS(',AWLEVEL=',AWLEVEL);
DistanceED  =CATS(',DistanceED=',DE);

if awlevel ='All' then delete;

keep UNITID --DistanceED;
RUN;
```

```
PART B: Sample record
UNITID=nnnnnn,SURVSECT=COM,PART=B,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=3,DistanceED=1
UNITID=nnnnnn,SURVSECT=COM,PART=B,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=5,DistanceED=2
UNITID=nnnnnn,SURVSECT=COM,PART=B,MAJORNUM=1,CIPCODE=01.0101,AWLEVEL=i,DistanceED=n
```

Display 7. PART B: Sample record (from IPEDS Survey Materials: Import Specifications)

PART C: ALL COMPLETIONS

```
*Now we need to have gender count on a separate line;
proc sort data= IPEDS_ALL_CLEAN out=Part_C_sort;
by race;
run;

proc transpose data= Part_C_sort out=Part_C_long name=gender;
by race race_code race_des;
var male_count female_count unknown total;
run;

*Put all together;
data part_c_summary;
set part_c_long;
if gender = 'Male_count'   then sex1=1;
if gender = 'Female_count' then sex1=2;
if gender = 'Unknown'      then sex1=0;
if gender = 'Total'        then sex1=9;
run;
```

```
DATA Part_C_final ;
LENGTH      UNITID      $13
            SURVSECT    $13
            PART        $8
            RACE        $9
            SEX         $7
            COUNT       $17;
SET part_C_Summary(RENAME=(RACE=RACE1)WHERE=(SEX1 IN (1,2)));


UNITID="UNITID=&UNITID";
SURVSECT=",SURVSECT=COM";
PART=",PART=C";
RACE=CATS(',RACE=',RACE1);
SEX=CATS(',SEX=',SEX1);
COUNT=CATS(',COUNT=',COL1);

*Instructions state not to include 10 in import file;
IF RACE1=10 THEN DELETE;

KEEP UNITID--COUNT;
RUN;
```

```
PART C: Sample record
UNITID=nnnnnn,SURVSECT=COM,PART=C,RACE=2,SEX=1,COUNT=nnnnn
UNITID=nnnnnn,SURVSECT=COM,PART=C,RACE=2,SEX=2,COUNT=nnnnn
UNITID=nnnnnn,SURVSECT=COM,PART=C,RACE=j,SEX=k,COUNT=nnnnn
```

Display 8. PART C: Sample record (from IPEDS Survey Materials: Import Specifications)

PART D: COMPLETERS BY LEVEL

For Part D, only one record is requested, so all data needs to be collapsed to one line. This involves several steps of gathering the total summary counts, total counts by race and gender, then merging all records together.

```
*Now we need to have to combine race totals on one line;
proc sort data=IPEDS_BY_LEVEL_clean out=Part_D_race(keep= ct_level race
race_code race_des total) ;
by  ct_level race_code;
run;


proc transpose data= Part_D_race out=Part_D_middle ;
by ct_level race_code;
var total;
run;
*Need to proc transpose twice to get in the correct form;
proc transpose data=Part_D_middle out=Part_D_race2(drop=_NAME_) delim=_;
  by ct_level ;
  id race_code _name_ ;
  var col1;
run;

*Now I need to rename according to IPEDS values;

data Part_D_race_final2(keep= ct_level CRACE:);
set Part_D_race2;
rename  NRA_Total = CRACE17_
        HIS_Total = CRACE41_
        AN_Total =  CRACE42_
        AS_Total =  CRACE43_
        BL_Total =  CRACE44_
        HP_Total =  CRACE45_
        WH_Total =  CRACE46_
        TWO_Total = CRACE47_
        UKN_Total = CRACE23_
        TOT_Total = CRACE48_ ;

IF CT_LEVEL= 'All' THEN DELETE;
run;
```

```
*Now for age;

data Part_D_age(keep= ct_level AGE:);
set IPEDS_BY_LEVEL_clean(where=(race=10));

 rename  _LT_18     = AGE1_
          _18_24    = AGE2_
          _25_39    = AGE3_
          _40_GT    = AGE4_
          _UKN_AGE  = AGE5_
          Total     = AGEtotal_;   /*Do not include total in import file */

run;

*Now, need to merge all together;

proc sort data = part_d_total; by ct_level; run;
proc sort data = part_d_race_FINAL2; by ct_level; run;
proc sort data = part_d_age; by ct_level; run;

data part_D_summary;
merge part_d_total (in=a)
      part_d_race_FINAL2
      part_d_age;
by ct_level;
if a;
run;
```

```
DATA Part_D_final ;
LENGTH      UNITID      $13
            SURVSECT    $13
            PART        $8
            CTLEVEL     $11
            CRACE15     $15
            CRACE16     $15
            CRACE17     $15
            CRACE41     $15
            CRACE42     $15
            CRACE43     $15
            CRACE44     $15
            CRACE45     $15
            CRACE46     $15
            CRACE47     $15
            CRACE23     $15
            AGE1        $15
            AGE2        $15
            AGE3        $15
            AGE4        $15
            AGE5        $15;

SET part_D_Summary;

 /*continued on next page */
```

```
IF CT_LEVEL = 'Cert (Level 1)'  THEN CT=1;
IF CT_LEVEL = 'Cert (Lvls 2,4)' THEN CT=2;
IF CT_LEVEL = 'Assc (Level 3)'  THEN CT=3;


UNITID="UNITID=&UNITID";
SURVSECT=",SURVSECT=COM";
PART=",PART=D";
CTLEVEL=CATS(',CTLEVEL=',CT);
CRACE15=CATS(',CRACE15=',CRACE15_);
CRACE16=CATS(',CRACE16=',CRACE16_);
CRACE17=CATS(',CRACE17=',CRACE17_);
CRACE41=CATS(',CRACE41=',CRACE41_);
CRACE42=CATS(',CRACE42=',CRACE42_);
CRACE43=CATS(',CRACE43=',CRACE43_);
CRACE44=CATS(',CRACE44=',CRACE44_);
CRACE45=CATS(',CRACE45=',CRACE45_);
CRACE46=CATS(',CRACE46=',CRACE46_);
CRACE47=CATS(',CRACE47=',CRACE47_);
CRACE23=CATS(',CRACE23=',CRACE23_);

AGE1=CATS(',AGE1=',AGE1_);
AGE2=CATS(',AGE2=',AGE2_);
AGE3=CATS(',AGE3=',AGE3_);
AGE4=CATS(',AGE4=',AGE4_);
AGE5=CATS(',AGE5=',AGE5_);




KEEP UNITID--AGE5;
RUN;
PROC SORT DATA=Part_D_final ; BY CTLEVEL; RUN;
```

STEP 6: CREATE FINAL IPEDS FILE TO EXPORT

The last step in this process is to combine Parts A, B, C and D into one file to export.

```
/* Create base file */
*Removing race=10 (total counts and 99.0000 - Grand total CIP counts);
data _null_;
  set Part_A_final(where=(race ne ',RACE=10' and CIPCODE ne ',CIPCODE=99.0000'));
  file "'&fileloc.\&outfile._FINAL.txt'";
  put (_all_) (+0);

run;

/* Append information to the existing file using MOD option */

data _null_;
set Part_B_final(where=(CIPCODE ne ',CIPCODE=99.0000'));
  file "'&fileloc.\&outfile._FINAL.txt'" mod ;
 put (_all_) (+0);

run;

data _null_;
set Part_C_final;
  file "'&fileloc.\&outfile._FINAL.txt'" mod ;
 put (_all_) (+0);

run;

data _null_;
set Part_D_final;
  file "'&fileloc.\&outfile._FINAL.txt'" mod ;
 put (_all_) (+0);

run;
```

CONCLUSION

By following the steps outlined in this paper, you can automate your IPEDS completions in 60 minutes or less. Once the program is complete, it is very easy to update the following year. Download the IPEDS Import specifications each year to see if anything in the outgoing file has changed.

REFERENCES

National Center for Education Statistics. 2019. “IPEDS 2019-20 Data Collection System Survey Materials Import Specifications.” Accessed August 15, 2019. https://surveys.nces.ed.gov/ipeds/VisImpSpecView.aspx?id=27&show=all&instid=466.

National Center for Education Statistics. 2019. “IPEDS Survey Components.” Accessed August 15, 2019. https://nces.ed.gov/ipeds/use-the-data/survey-components/7/completions.

Notepad++. 2019. “Download Notepad++. Accessed August 15, 2019 https://notepad-plus-plus.org/

Hume, James. 2012. “Uploading Your IPEDS Data Electronically Using SAS®.” *SAS Global Forum 2012 Paper*, 074-2012. Bowling Green, Kentucky: Western Kentucky University.

Available at http://support.sas.com/resources/papers/proceedings12/074-2012.pdf .

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Yolanda N. Ingram
Research Specialist
Wake Technical Community College
(919) 866-6135
yningram@waketech.edu