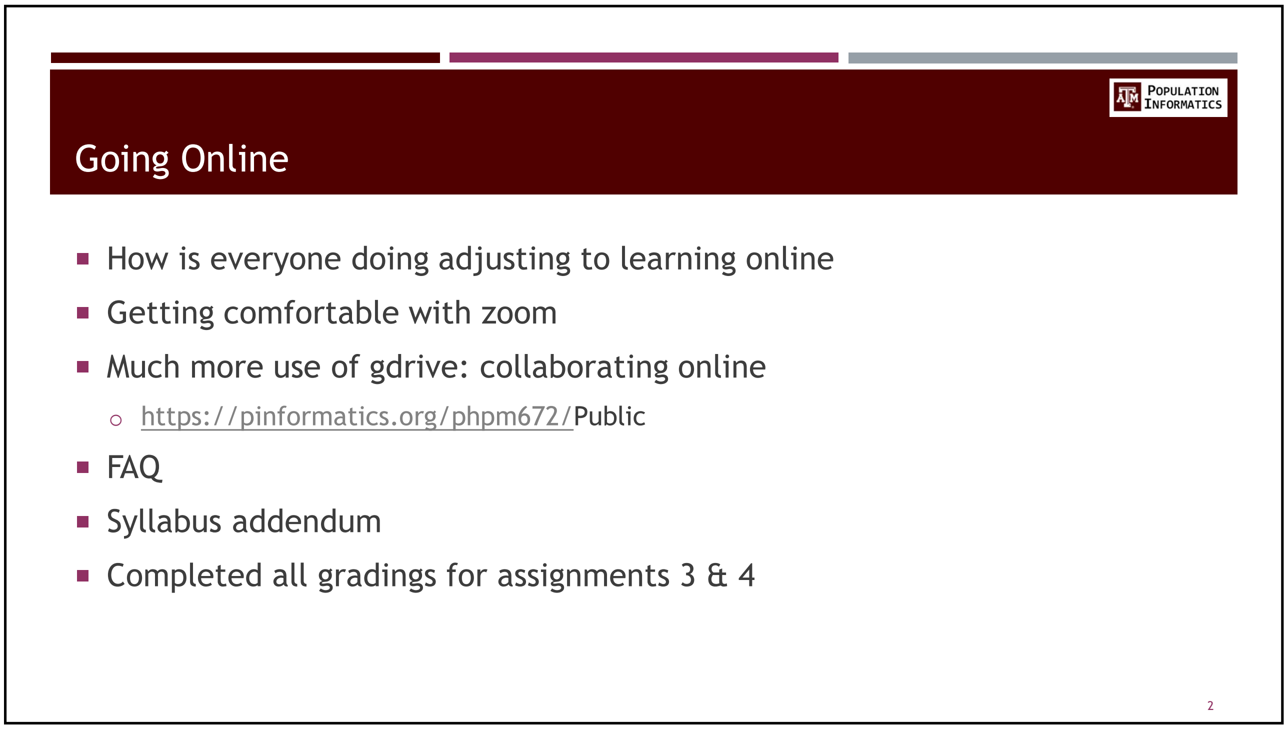
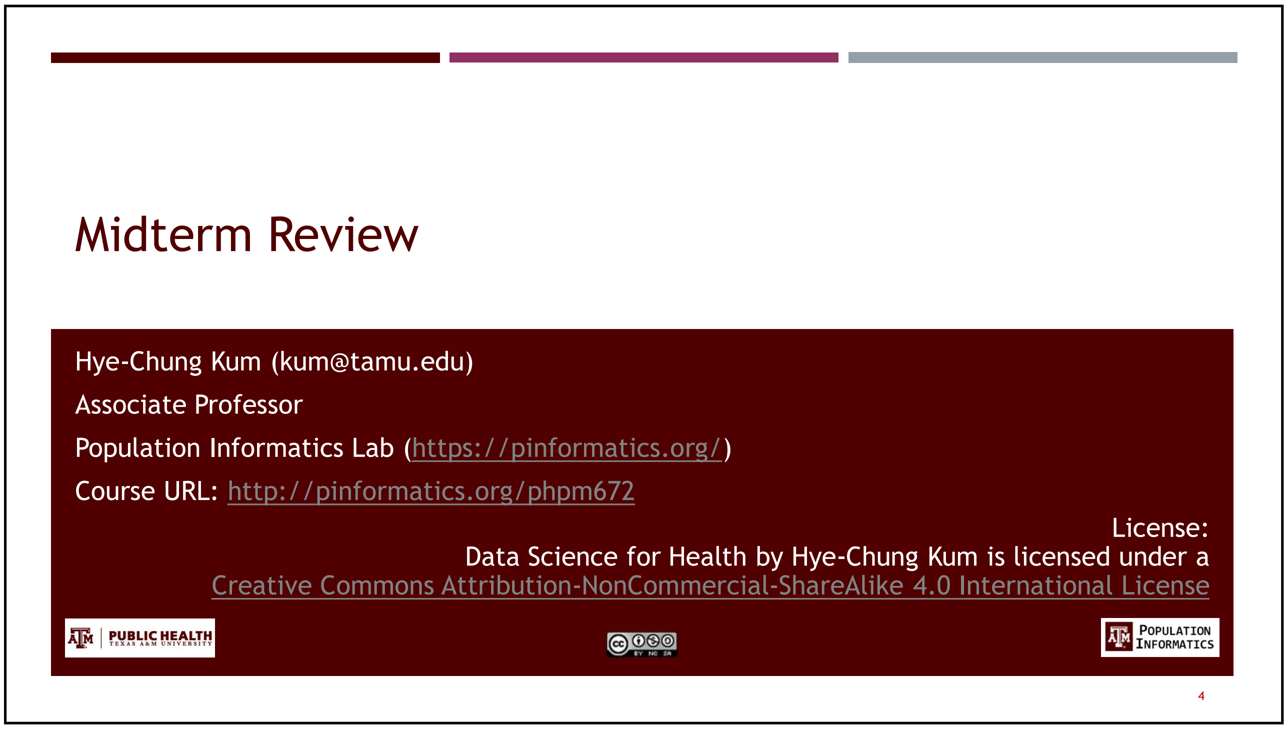


Going Online

- How is everyone doing adjusting to learning online
- Getting comfortable with zoom
- Much more use of gdrive: collaborating online
  - https://pinformatics.org/phpm672/Public
- FAQ
- Syllabus addendum
- Completed all gradings for assignments 3 & 4

2

2



Midterm Review

Hye-Chung Kum (kum@tamu.edu)
Associate Professor
Population Informatics Lab (https://pinformatics.org/)

Course URL: http://pinformatics.org/phpm672

4

4

Record Linkage & SAS String functions

- Record linkage
- https://support.sas.com/publishing/pubcat/chaps/59343.pdf
  - SAS 9.1

5

5

Questions: Assignment 5?

6

What you should have learned in 8 weeks

- Learning Objective Questions (do today at the end of class)
  - Do you know how to talk to a computer? (To get it to do what you want)
  - Do you know how to think data?
  - Can you use SAS to manipulate data into a format you need?
    - libname, variables, data steps, labels, formats, arrays, loops, conditionals, boolean expressions, proc summary, proc transpose
- What is left: 6 weeks
  - 1 week: midterm
  - 2 weeks: reusable code (macros)
  - (1)+3 weeks: a project to try this out

7

7

Midterm Schedule

- Midterm take home after class today
  - Submit in one Week midnight Monday next week
- Midterm in class: thur
- Office hours & lab cancelled since midterm goes out today
  - while take home midterm (until next Tues)
- What you need:
  - Part 2: E-campus, SAS on your laptop
  - Part 1: E-campus, camera, sound,

8

8

Midterm format (20%)

- 5 questions (50 points): take home today
  - Open book / open notes / use SAS
  - Programming/debugging questions
  - submit by midnight Monday on E-Campus (one week)
- 25 questions (about 2*25=50 points)
  - On E-Campus
  - multiple choice similar to practice quiz
  - Closed book
  - Thur (in class): 1h 15min (proctored on zoom with video + sound on)

9

Take home: Due Monday midnight

- Write SAS code to (8*5=40pts)
  - Data Step 1
    - Q1.1 read in datasets X1..Xn and make new dataset Y
    - Q1.2 keep, rename, label variables v1-vn
    - Q1.3 code variable c1
    - Q1.4 use arrays and loops to recode variable c2
  - Proc Steps
    - Q2.5 convert dataset Y to dataset Z
    - Q2.6 Find and show descriptive (avg/max/median) (Must use SAS code)
  - Data Step 2
    - Q2.7 link in dataset L to dataset Y
    - Q2.8 Print observations meeting condition (Must use SAS code)
  - Typically few lines of code per question
  - Submit code/log/output
- Debug the following code (10pts)
  - Fix the program to run properly
  - Submit code/log/output
- Extra Credit (10pts)

10

```
******** Section 1: Frist Data Step *******;
code

* Q1.1;
code

* Q1.2;
code

******** Section 2: Proc Steps *******;
* Q2.7;
code

******** Section 3: Second Data Step *******;
```

11

Things to watch out for

- Make your code readable to people
  - Indent your code
  - Use newline
  - Use reason variable names
- Be efficient
  - No unnecessary data steps

12

Extra Credit (10pts=1+2+7)

- Part 2.2
- Part 3.1: Extra Credit
  - READ your assignment 2 (this is the first real program you submitted in class) that you submitted, and make is more elegant code now that you know more about coding.
  - Submit FOUR files, the regular sas (the more elegant code you wrote)/log/lst AND the code annotated with the changes you made and why (you can do this in word so that you can use formatting, such as bold/color, to annotate.
- Part 3.2: Try to solve a computational problem

13

Midterm: Responsible materials

- Readings from the Little SAS Book
  - All sections in chapter 1
  - All sections in chapter 2
  - All sections in chapter 3
  - Sections 4.1 to 4.10 in chapter 4
  - All sections in chapter 6
  - Note that some of the materials were not covered in class or assignment, but you are responsible for anything covered in the required reading from the book
- Other materials
  - All class notes up to this class (slides on the class website).
  - None of the articles are part of the midterm (except to the extent covered in class on the notes)

14

SAS Basics

- program/log/output (lst or html)
- libname
- ;
- setting up work environment
  - How you will use the software
  - How you will organize your files

15

SAS

- keywords
  - data, set, merge, obs, where, if, do, end, keep, drop, rename, label, in
  - array
  - proc
    - sort, print, summary, transpose, freq
- functions
  - put ()
  - compress ()
  - lowcase () / upcase ()

16

Boolean expression evaluation

- X | | (Y & X)

X	Y		X||Y&X

17

Questions: midterm? (gdocs)

If you are feeling sick, email Dr. Kum

18

Assignment 1

- Setup work environment
- Use the SAS software
- SAS programming basics
  - data step & proc step
  - Libname (where is the folder with the data?)
  - Writing code & Reading logs

19

Assignment 2

- Understand variables (names, types, labels)
- To write conditional logic codes
- Subset columns (variables) from a table
- Subset rows (observations) from a table
- Recode, rename variables and calculate new variables
- Label variables and values

20

Assignment 3

- use for loops (iterative loops)
- use while loops (conditional loops)
- SAS: use one dimensional arrays

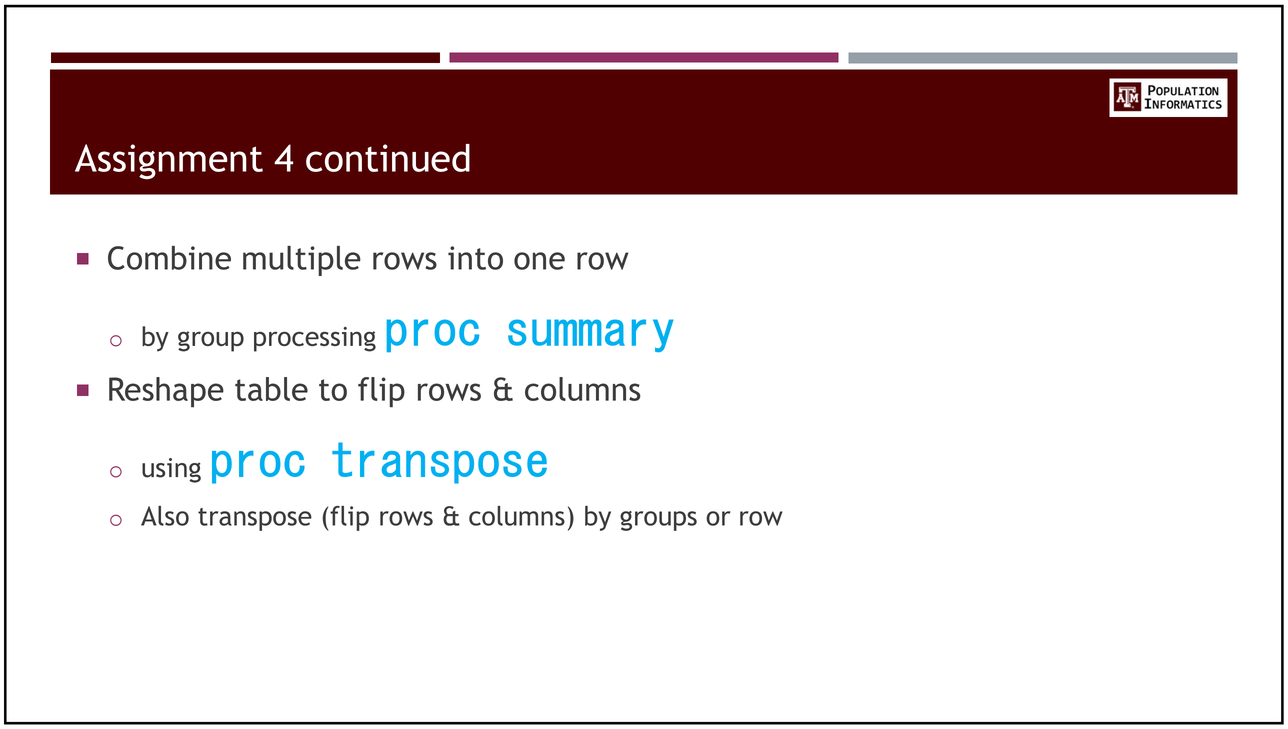
21

Assignment 4

- **Concatenate multiple tables (more rows)**
  - **stack tables on top of each other to increase the number of rows**
  - using `set`
  - Be sure to understand the different behavior given different situations (i.e. what happens to shared variables? What happens to not shared variables?)
- **Link up multiple tables using a shared key (more columns)**
  - **align the rows using the shared key, and link multiple tables to increase the number of variables in the tables**
  - using `merge`
  - Be sure to understand the different behavior given different situations (i.e. what happens to shared vars? What happens to not shared vars?)
  - What is a 1-to-1 link
  - What is a 1-to-N link
  - What is a N-to-N link (you will not be doing this, but need to understand what this is. This must be done with proc sql in SAS)
- **New keyword** `in=`

22

22



Assignment 4 continued

- Combine multiple rows into one row
  - by group processing `proc summary`
- Reshape table to flip rows & columns
  - using `proc transpose`
  - Also transpose (flip rows & columns) by groups or row

23

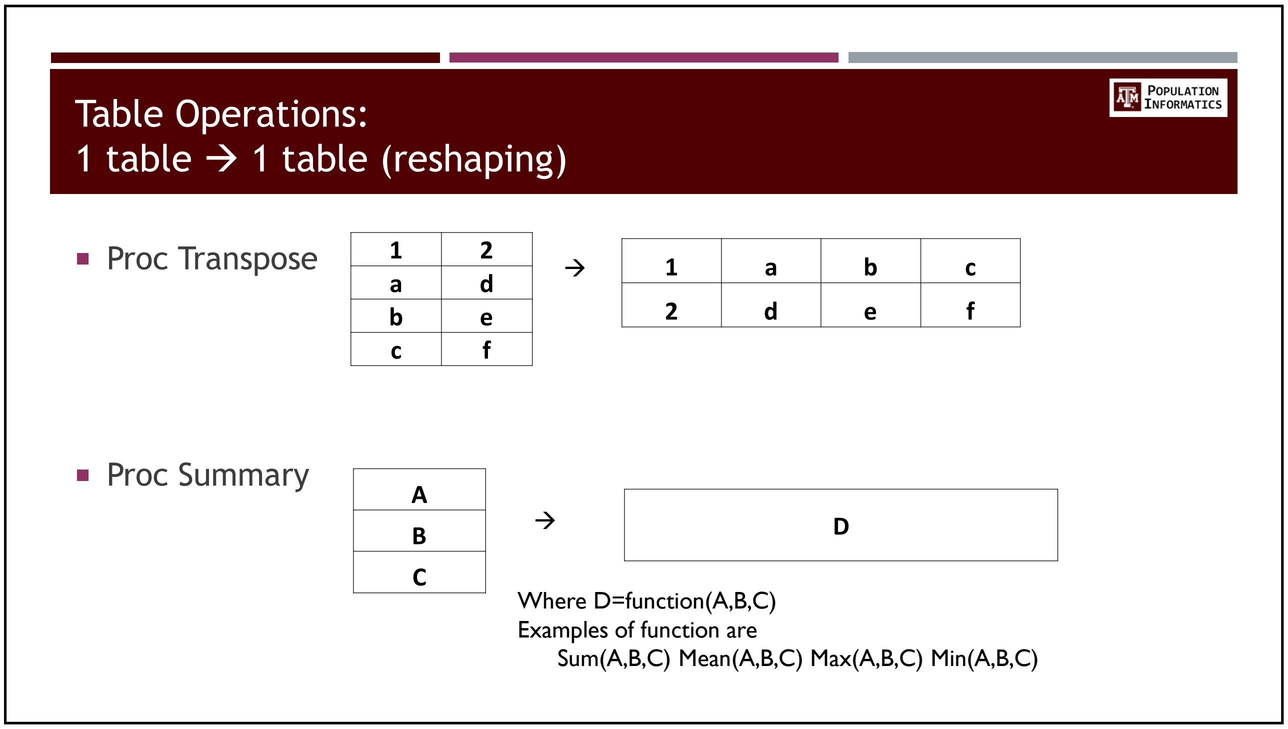


Table Operations: 1 table → 1 table (reshaping)

- Proc Transpose

1	2
a	d
b	e
c	f

→

1	a	b	c
2	d	e	f

- Proc Summary

A
B
C

→

D

Where D=function(A,B,C)
Examples of function are
Sum(A,B,C) Mean(A,B,C) Max(A,B,C) Min(A,B,C)

24

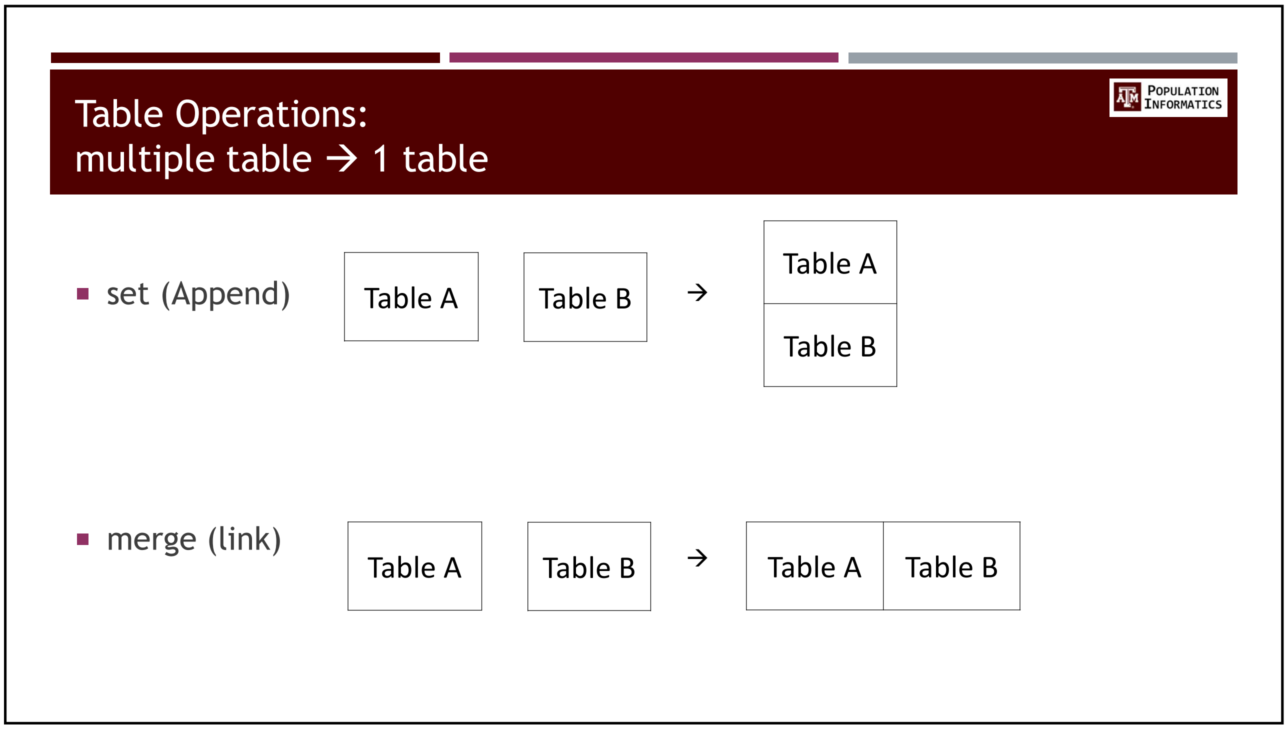


Table Operations: multiple table → 1 table

- set (Append)

Table A | Table B →

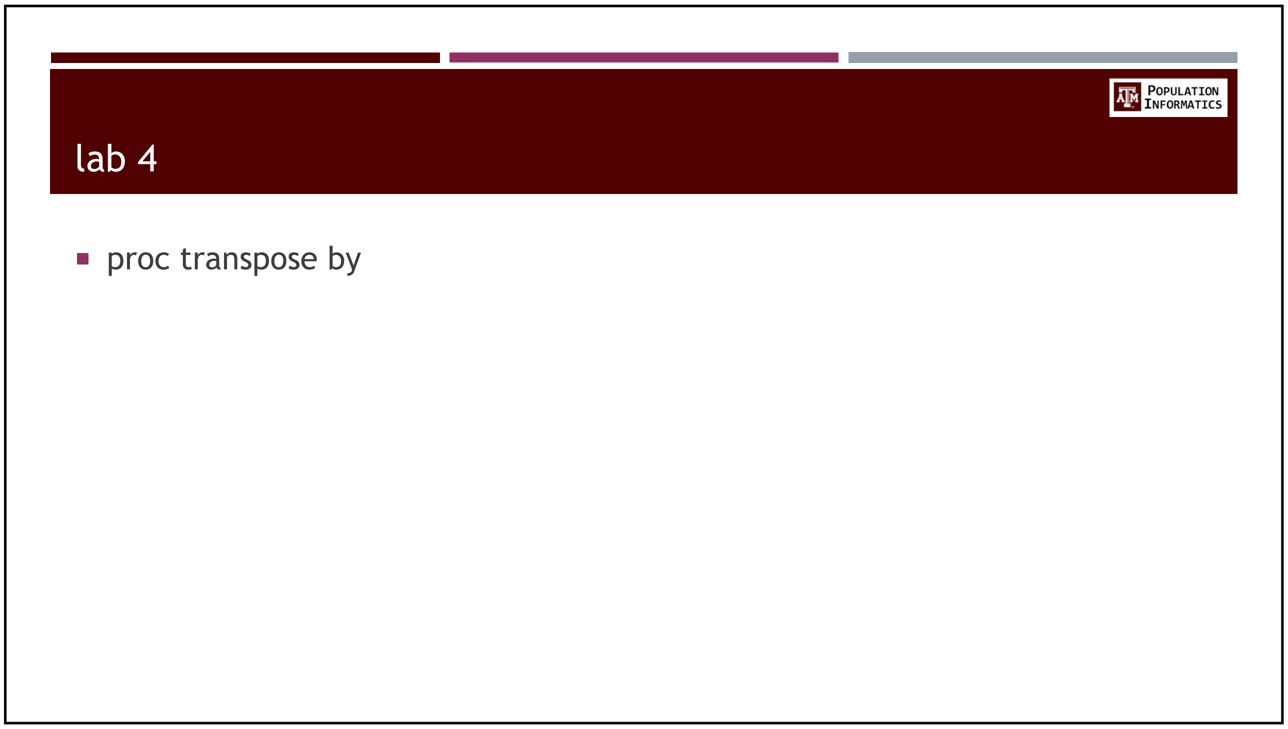
Table A
Table B

- merge (link)

Table A | Table B →

Table A	Table B

25



lab 4

- proc transpose by

26

Formats

- Create using proc format
- Use Case 1: Labeling values
  - Assign using format statement (permanent, temporary)
  - Only used interpret the value (ie. printing, display)
- Use Case 2: Can be used to recode variables (know how different)
  - put(var, format)
  - new variable type? Value?

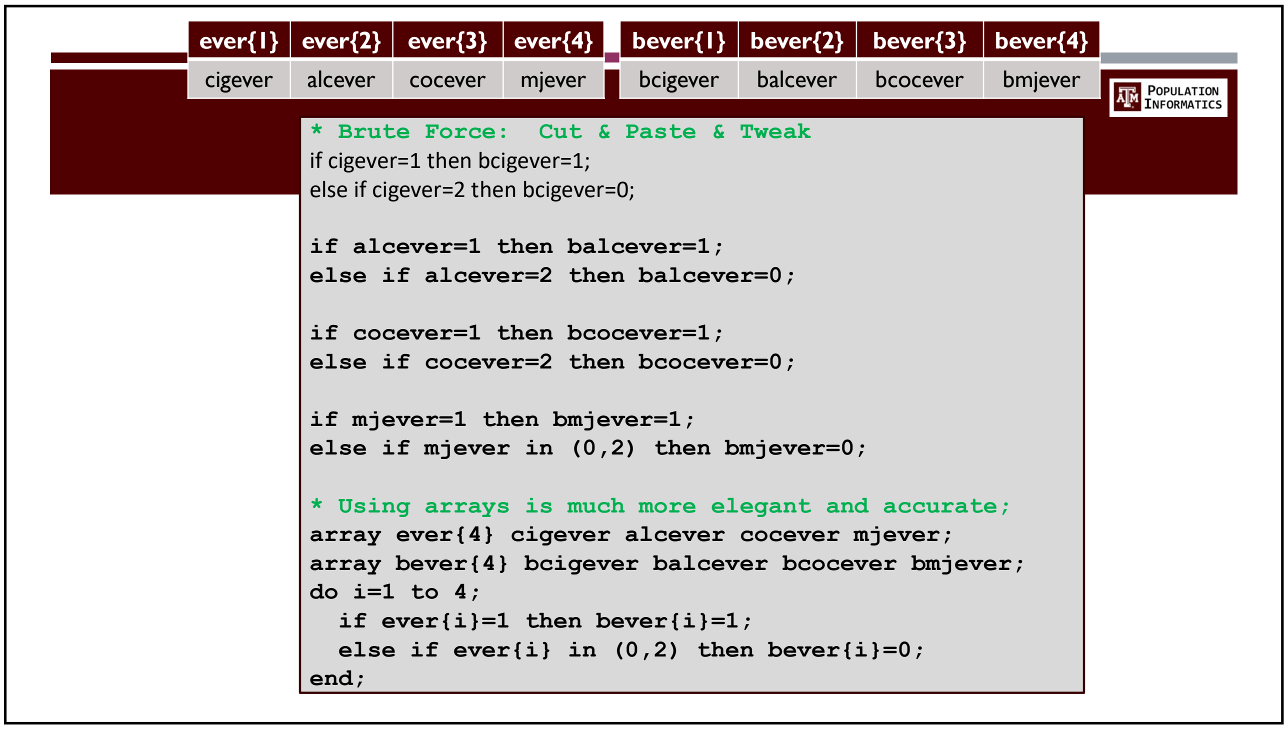
```
proc format;
  value gender
  1= ‘Male’
  2= ‘Female’
  other= ‘Missing’ ;
* In data step;
data outfn;
set infn;

csex $7.;
csex=put(sex, gender.);
```

27

Arrays

- Array n{*} n9-n23;
- Array a{*} $7. a11-a23;
- Name? n and a
- How many elements? N=15 a=13
- Type? N=number, a=string of length 7
- n15 index? 7

28

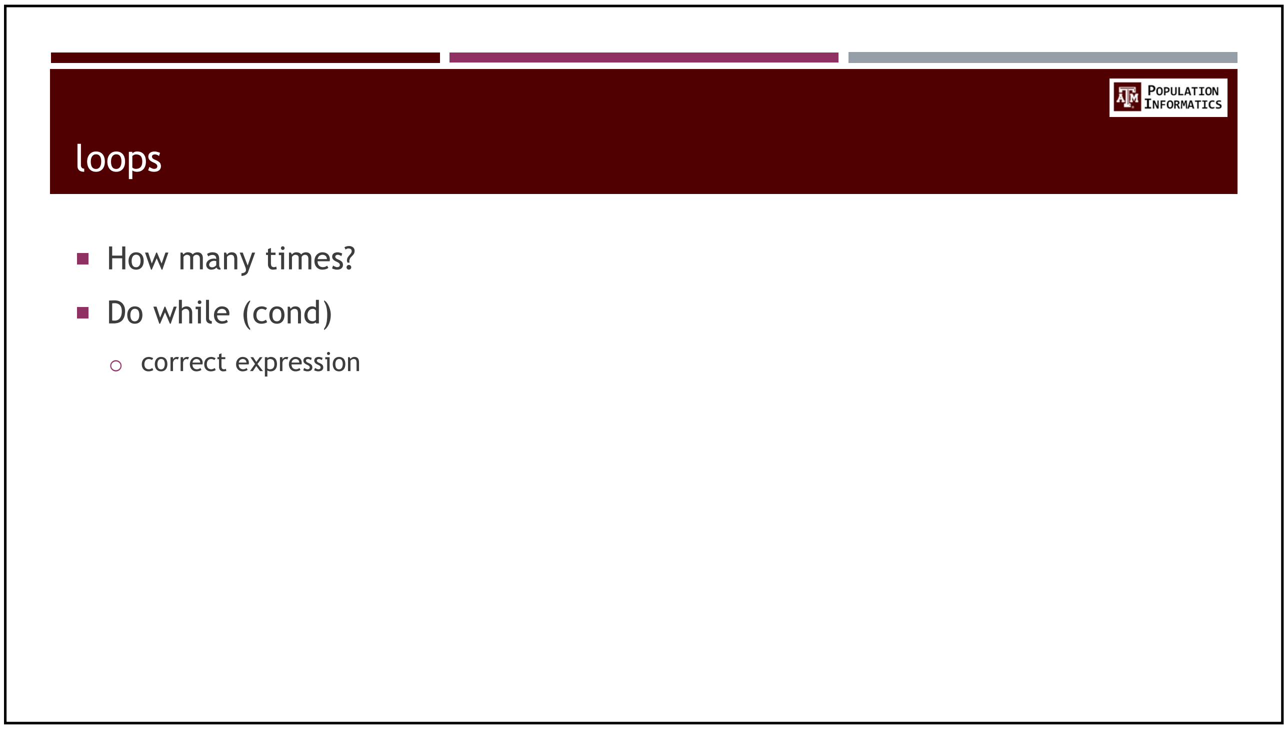


ever{1}	ever{2}	ever{3}	ever{4}	bever{1}	bever{2}	bever{3}	bever{4}
cigever	alcever	cocever	mjever	bcigever	balcever	bcocever	bmjever

POPULATION INFORMATICS

```
* Brute Force:  Cut & Paste & Tweak
if cigever=1 then bcigever=1;
else if cigever=2 then bcigever=0;

if alcever=1 then balcever=1;
else if alcever=2 then balcever=0;

if cocever=1 then bcocever=1;
else if cocever=2 then bcocever=0;

if mjever=1 then bmjever=1;
else if mjever in (0,2) then bmjever=0;

* Using arrays is much more elegant and accurate;
array ever{4} cigever alcever cocever mjever;
array bever{4} bcigever balcever bcocever bmjever;
do i=1 to 4;
  if ever{i}=1 then bever{i}=1;
  else if ever{i} in (0,2) then bever{i}=0;
end;
```

29



POPULATION INFORMATICS

loops

- How many times?
- Do while (cond)
  - correct expression

30