

Lab: Vars

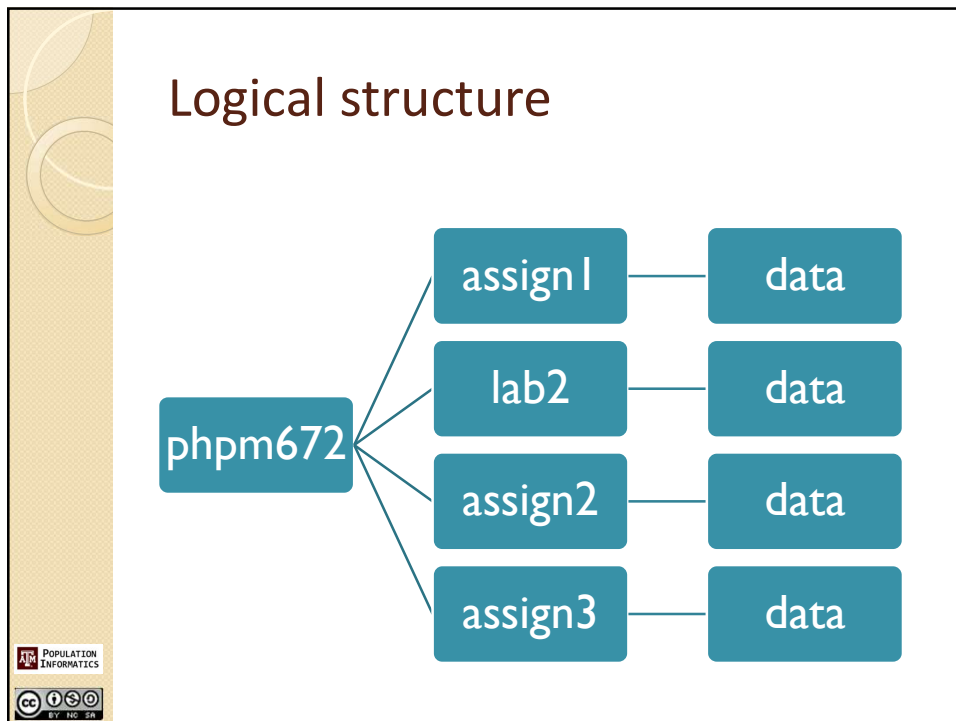
Hye-Chung Kum
 Population Informatics Research Group
<http://pinformatics.org/>

License:
 Data Science in the Health Domain by Hye-Chung Kum is licensed under a
[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

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

POPULATION INFORMATICS
 CC BY NC SA

1



2

Fri 4-6

- Make sure to understand lab 2
 - Try to finish by Thur. So you can spend next week doing assignment 2
 - BUT submit with assignment 2 the week after.
 - You MUST submit programs, logs, and output along with assignment 2
 - This is how you will LEARN
 - Most IMPORTANT part of class
- Dataset(s) you want to use through out the class
 - Flu dataset
 - Texas Inpatient Public Use Data File (PUDF)
 - <http://www.dshs.state.tx.us/thcic/hospitals/Inpatientpdf.shtm>



3

Lab 2 & Assignment 2: Objective

- 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



4

Recommended Reading

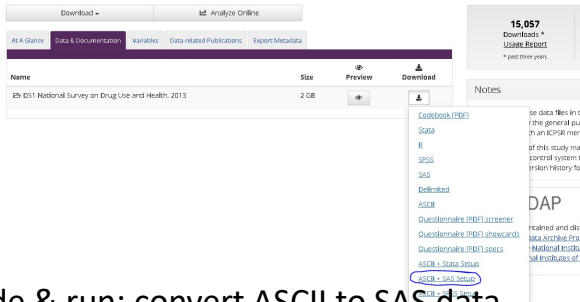
- Carefully read each of the modules below. Each has very good explanations of exactly how to do certain things.
 - <https://stats.idre.ucla.edu/sas/modules/creating-and-recoding-variables-in-sas/>
 - <https://stats.idre.ucla.edu/sas/modules/using-sas-functions-for-making-and-recoding-variables/>
 - <https://stats.idre.ucla.edu/sas/modules/ubsetting-data-in-sas/>
 - <https://stats.idre.ucla.edu/sas/modules/labeling/>
- Little SAS book
 - Sections in Chapter 3

AIM POPULATION INFORMATICS
CC BY-NC-SA

5

Getting data

- Log in
- Read agreement
- **Download**



The screenshot shows the DAP interface with a table of datasets. The selected dataset is '2011 National Survey on Drug Use and Health 2013' with a size of 2 GB. A dropdown menu is open, showing various download options. The 'ASCII - SAS Setup' option is highlighted with a red circle.



- Modify code & run: convert ASCII to SAS data
- Do the lab (subset, label, rename etc)

AIM POPULATION INFORMATICS
CC BY-NC-SA

6

Data Step

```
libname data "D:\HPM-Users\kum\phpm672\lab2\data";  
data outfn;  
  set infn;  
  ...code...  
  
data mynsduh;  
  set data.nsdh;  
  ...code...
```



 

7

Subset columns (variables)

- SAS
 - Three places possible
 - Reading in, writing out, during datastep
 - **keep**, **drop**

```
data mynsduh;  
  set data.nsdh (keep=var);  
  
data mynsduh;  
  set data.nsdh (drop=var);
```

8

Subset rows (observations)

- SAS
 - `where cond ;`
 - `if cond ;`

 - `Data data.mydata;`
 - `Set data.indata;`
 - `Where age>=18;`

 - `Data data.mydata;`
 - `Set data.indata;`
 - `if age>=18;`



9

Calculate new variable (assignment)

- SAS (in data step)
 - `var1 = 1 ; * assignment;`
 - `num1=.; * numeric missing value;`
 - `str1= " " ; * string missing value;`

 - `Gender=1;`
 - `Gender=F;`
 - `Gender= "F" ;`
 - `Gender= 'F' ;`



10

