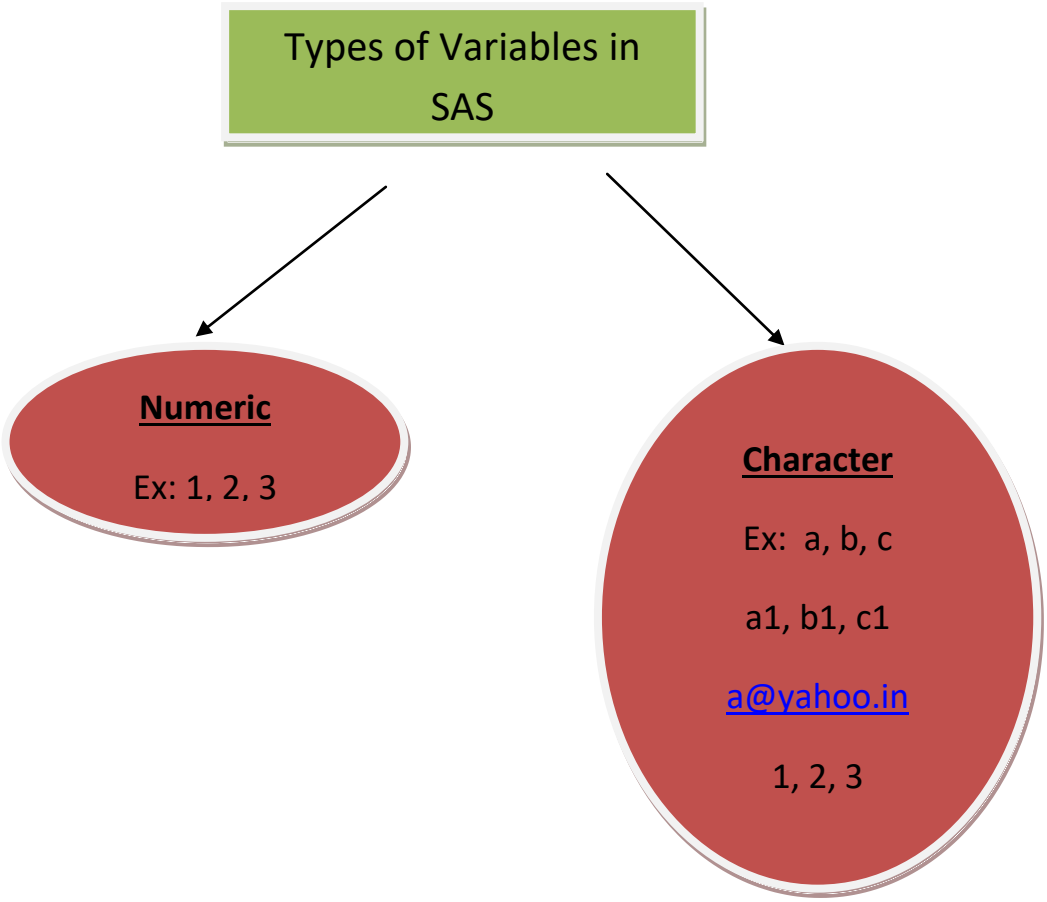
Dated: 24/02/2018

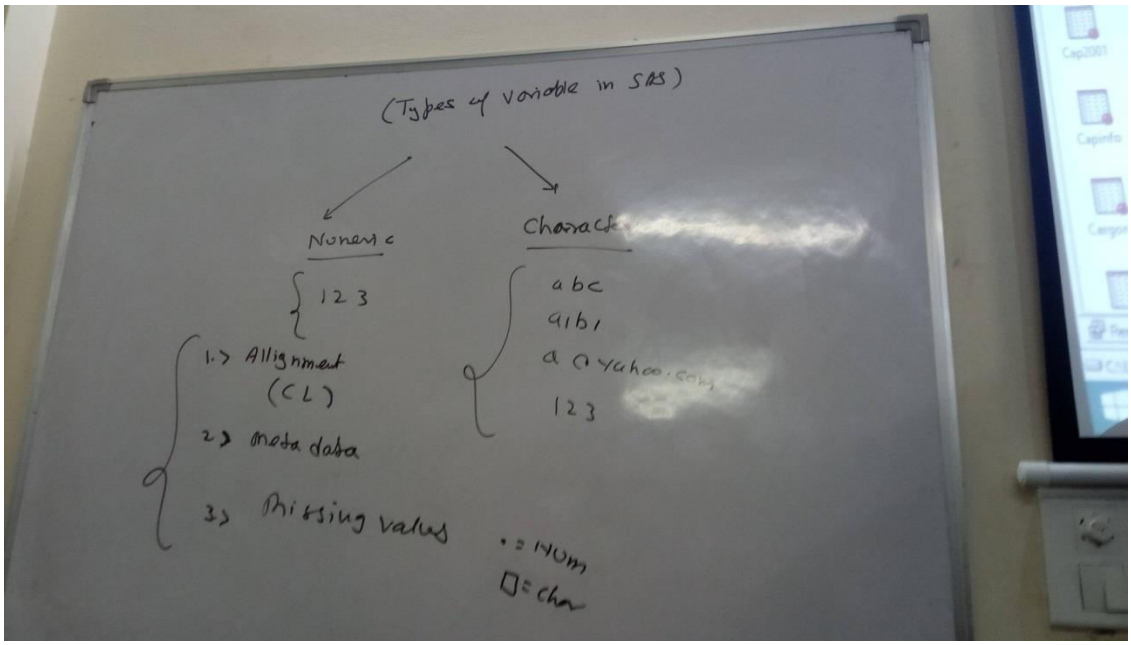
Statistical Analysis System (SAS v9.4): Class 4

……

Types of Variables in SAS :-



- Numeric variable type can only store numbers.
- Character variable type is more commonly used as it can store characters, numbers (stores in string format), alpha-numeric, special characters. Therefore this type is more feasible and has more scope.



There are 3 ways to identify if a variable is character type or numeric :-

1) **Alignment**: Character always will be left aligned ,
Numeric always will be Right aligned
Ex: If id = 2548 is numeric defined, it will be right aligned.
Ex: If id = 2548; is character defined, it will be left aligned.

2) **Metadata:** Data about data is called metadata. To go to metadata
Right Click then View Column on the dataset.

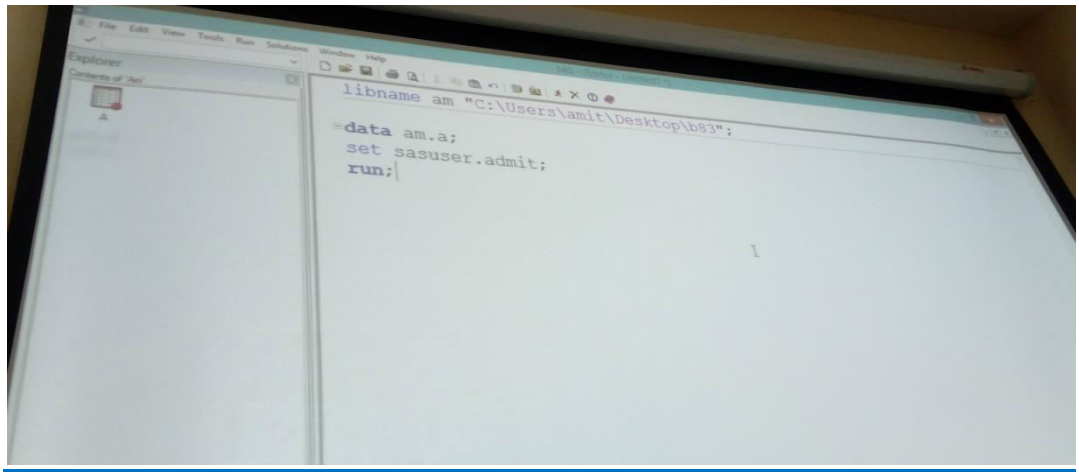
3) **Missing Value:**
" . " implies Numeric variable type ,
□ Blank cell represent Character variable type.

Creating your own Library :-

- Keyword is always fixed, i.e " libname " to be used in the syntax below
- Library name is anything that you decide for your created library.
- Create a folder where your library gets stored
- Open the folder, copy the folder path (from the address bar), this becomes folder path in the syntax below (to be written in between " ")

Syntax: Keyword Library name "Folder path"

Ex: libname am " C:\users\amit\desktop\B83"



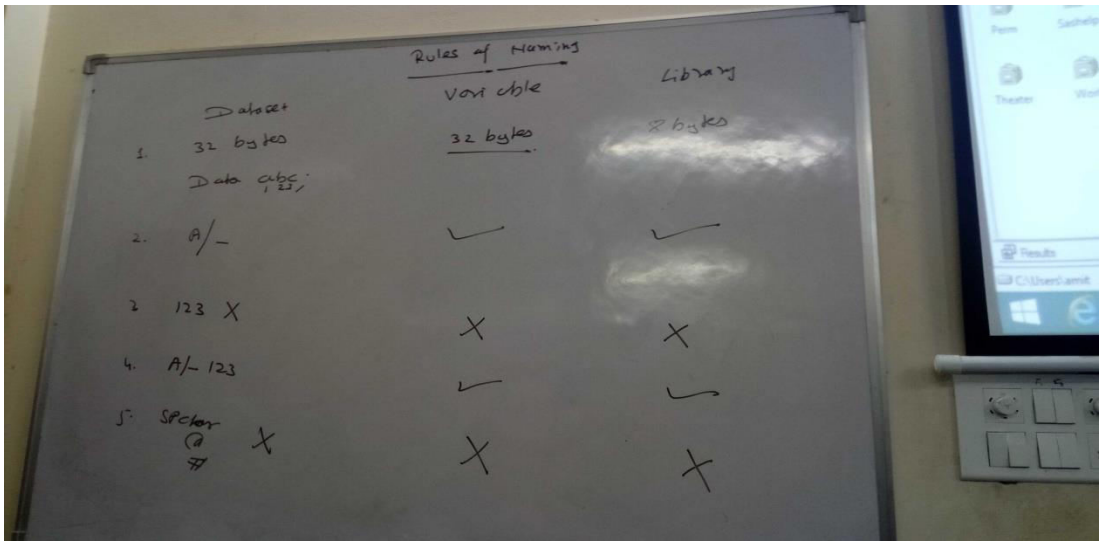
Example: libname am “C:\Users\amit\Desktop\b83”

```
data am.a;
set sasuser.admit;
run;
```

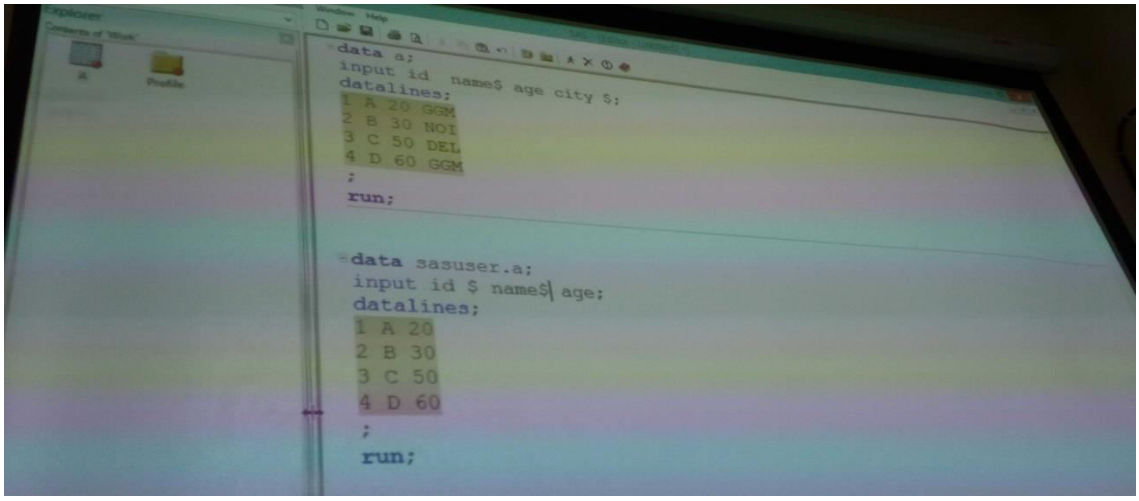
Result : This code creates a dataset “ a ” in the permanent library “ am ”where it copies dataset “admit” from permanent library “sasuser” into dataset “ a ”.

Rules of naming (for dataset, variable, library):-

Dataset	Variable	Library
• can only be 32 bytes. ex: data abc; (has 3 bytes) • can never start with numeric. ex: 123 is wrong • Can have numeric but only if can start with a character or underscore (A / _) • name starts with a character. ex: A123 / _123 • can never have special character (@, #, etc.) in the name.	• can only be 32 bytes. ex: data abc; (has 3 bytes) • can never start with numeric. ex: 123 is wrong • Can have numeric but only if can start with a character or underscore (A / _) • name starts with a character. ex: A123 / _123 • can never have special character (@, #, etc.) in the name.	• can only be 8 bytes. ex: data abcdefgh; (has 8 bytes, it is the maximum length) • can never start with numeric. ex: 123 is wrong • Can have numeric but only if can start with a character or underscore (A / _) • name starts with a character. ex: A123 / _123 • can never have special character (@, #, etc.) in the name.



Creating your own data:-



- **Example 1:**

```
data a;                          \\ statement 1
input id name$ age city $;       \\statement 2
datalines;                       \\statement 3
1   A  20  GGM                   \\statement 4
2   B  30  NOI                        …….5
3   C  50  DEL                        …….6
4   D  60  GGM                        …….7

;                                     …….8

run;                                  …….9
```

Statement 1: creates dataset “a” in “work” library.

Statement 2:

- Input: Input is a keyword and defines the variables in the dataset.
- If a variable defined is character type or numeric type can be differentiated if it has “ $ ” sign after it and likewise for all variables.
  Example: here, id is numeric since no “$” sign after it.
  And name is character since there is a “$” sign after it.

Statement 3: datalines indicates incoming data line follows as per the variables created and is read into the variables accordingly.

Statement 4-7: incoming data is read and is fed into the defined variables in statement 2.

Statement 8: terminate the incoming data flow.

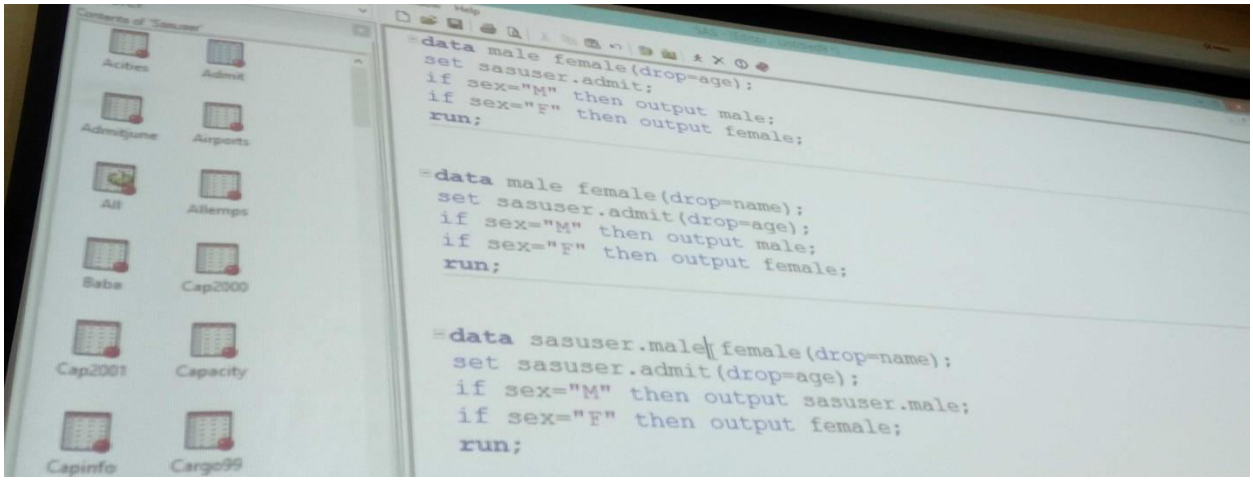
Statement 9: terminates the dataset.

Result: This creates a dataset “a” in work library with 4 variables and 4 observations.

Interview Question: Difference between Datalines and Cards.

Answer: There is no difference between both; it just tells user that data that follows is created with respect to variables defined.

Splitting a dataset into 2 or more dataset; (Slicing & Dicing):-



Example 1:

Data male female (drop = age);

Set sasuser.admit;

If sex = “M” then output male;

If sex = “F” then output female;

Run;

Result: Creates 2 dataset male & female, from original “admit” dataset of “sasuser” library (21 Observations & 9 Variables)

male dataset has 10 observations and 9 variables

female dataset has 11 observations and 8 variables, does not have age variable, since dropped)

Example 2:

Data male female (drop = name);

Set sasuser.admit (drop = age);

If sex = “M” then output male;

If sex = “F” then output female;

Run;

Result: Creates 2 dataset male & female, from original “admit” dataset of “sasuser” library (21 Observations & 9 Variables)

Note: Here, age is already dropped used with Set statement applies to both male & female dataset.

Now, male dataset has 10 observations and 8 variables

female dataset has 11 observations and 7 variables, does not have “name” variable, since dropped along with “age” variable which is dropped as used with Set statement)

Example 3:

Data sasuser.male female (drop = name);

Set sasuser.admit (drop = age);

If sex = “M” then output sasuser.male;

If sex = “F” then output female;

Run;

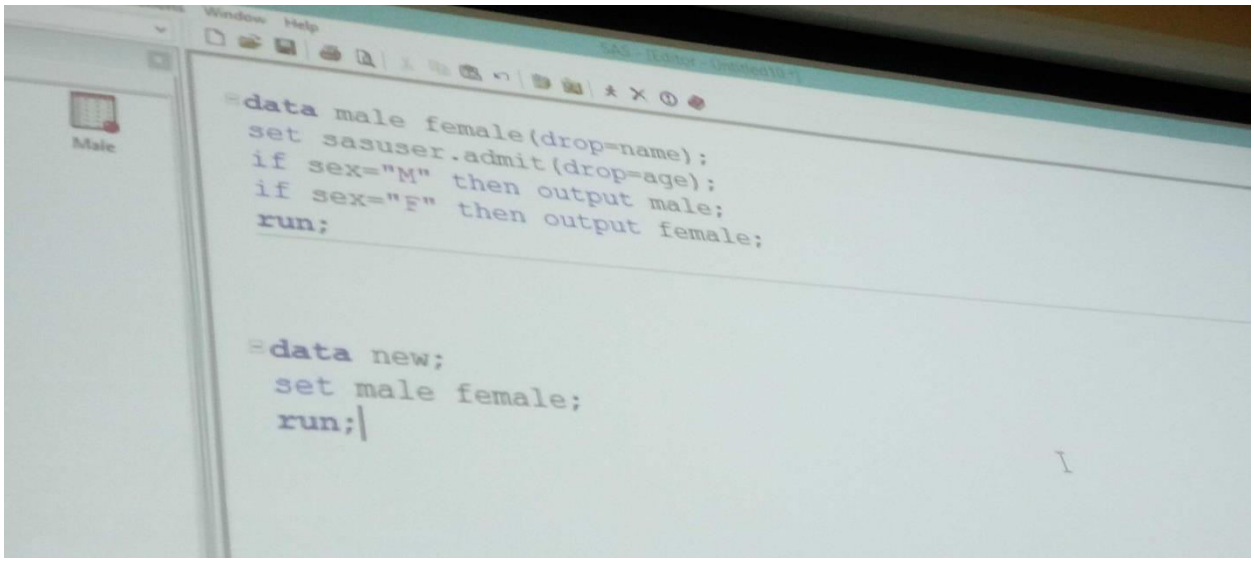
Result: Creates 2 dataset “male” in permanent “sasuser” library & “female” in temporary “work” library, from original “admit” dataset of “sasuser” library (21 Observations & 9 Variables)

Note: Here, age is already dropped used with Set statement applies to both male & female dataset.

Now, male dataset has 10 observations and 8 variables

female dataset has 11 observations and 7 variables, does not have “name” variable, since dropped along with “age” variable which is dropped as used with Set statement)

Note: The only difference to be noted here in this example compared to example 2 above is of “sasuser.male”, usage of the dataset name as declared has to be similar elsewhere in the entire datastep.



Example 4:

Data new;

Set male female;

Run;

Result: This creates a dataset “new” in “work” library which combines male & female dataset into 1 dataset with the combined observations and variables of both the datasets.

Question: Create 3 different datasets r12, r34, r56 in work library from a dataset “cargorev” of “Sasuser” library where r12 has Route variable as Route1 or Route2, r34 has Route variable as Route3 or Route4, r56 has Route variable as Route5 or Route6.

Answer:

```
Data r12 r34 r56;

Set sasuser.cargorev;

If route IN (“Route1” “Route2”) then output r12;

If route IN (“Route3” “Route4”) then output r34;

If route IN (“Route5” “Route6”) then output r56;

Run;
```