

SAS/AF Frame in programmer toolkit

INTRODUCTION

The SAS/AF based toolkit presented in this presentation has been develop in Novartis Modeling and Simulation group to provide support in validation process and data review. Some of the utility found in this toolkit will be presented in this paper.

DATA COMPARE

During the validation of a dataset it is helpful to view both the base and the validation datasets at the same time. While it can be done by having two separate viewer open, both are independent and need to have subset apply and the columns ordered. A Frame that displays both datasets and keeps them in synch can be very handy.

The frame display temporary copy of the selected datasets. Displaying copies allows update to be made to either datasets will the datasets remain visible in the frame. Once there is change in the datasets, the frame can be refreshed to display the changes in both datasets without losing the columns orders and subset.

Base

	ZIP	Y	X	ZIP_CLASS	CITY	STATE	STATECODE	STATENAME	COUNTY
1	00501	40.813078	-73.046388	U	Holtsville	36	NY	New York	103
2	00544	40.813223	-73.049288	U	Holtsville	36	NY	New York	103
3	00601	18.165950	-66.723627		Adjuntas	72	PR	Puerto Rico	1
4	00602	18.383005	-67.186553		Aguada	72	PR	Puerto Rico	3
5	00603	18.433236	-67.151954		Aguadilla	72	PR	Puerto Rico	5
6	00604	18.505289	-67.135899	P	Aguadilla	72	PR	Puerto Rico	5
7	00605	18.436149	-67.151346	P	Aguadilla	72	PR	Puerto Rico	5
8	00606	18.185616	-66.977377		Maricao	72	PR	Puerto Rico	93
9	00610	18.285948	-67.144161		Anasco	72	PR	Puerto Rico	11
10	00611	18.287716	-66.797578	P	Angeles	72	PR	Puerto Rico	141
11	00612	18.471326	-66.728149		Arecibo	72	PR	Puerto Rico	13
12	00613	18.472737	-66.719283	P	Arecibo	72	PR	Puerto Rico	13

Compare

	ZIP	Y	X	ZIP_CLASS	CITY	STATE	STATECODE	STATENAME	COUNTY
1	00501	40.813078	-73.046388	U	Holtsville	36	NY	New York	103
2	00544	40.813223	-73.049288	U	Holtsville	36	NY	New York	103
3	00601	18.165950	-66.723627		Adjuntas	72	PR	Puerto Rico	1
4	00602	18.383005	-67.186553		Aguada	72	PR	Puerto Rico	3
5	00603	18.433236	-67.151954		Aguadilla	72	PR	Puerto Rico	5
6	00604	18.505289	-67.135899	P	Aguadilla	72	PR	Puerto Rico	5
7	00605			P					
8	00606	18.185616	-66.977377		Maricao	72	PR	Puerto Rico	93
9	00610	18.285948	-67.144161		Anasco	72	PR	Puerto Rico	11
10	00611	18.287716	-66.797578	P	Angeles	72	PR	Puerto Rico	141
11	00612	18.471326	-66.728149		Arecibo	72	PR	Puerto Rico	13
12	00613	18.472737	-66.719283	P	Arecibo	72	PR	Puerto Rico	13

Find OBS: Find Where: Apply Hide/Unhide Compare Refresh OK

SYNCHRONIZATION BETWEEN VIEWERS

The top view is considered the base one and the bottom one the compare one.

In addition to defining both viewer and models we will need two scrollbars for horizontal and vertical scrolling.

PhUSE 2010

The number of observation scroll on page down or page up need to be defined. To ensure all observations will be visible it should be set to the number of observation displayed. This should be set in the INIT section of the frame.

```
*****;  
* Define large scroll amount  
*****;  
scrollbarVertical.largeScrollAmount = 12 ;
```

When either of the selected table is changed there is a need to change scrollbars property.

```
scrollbarHorizontal.value = 1 ;  
scrollbarHorizontal.minimum = 1 ;  
scrollbarHorizontal.maximum = max maximum number of columns in both viewers ;  
  
scrollbarVertical.value = 1 ;  
scrollbarVertical.minimum = 1 ;  
scrollbarVertical.maximum = maximum number of rows in both viewers ;
```

The synchronization of the viewer is relatively easy to implement as shown below

```
*****;  
* Scroll between columns  
*****;  
scrollbarHorizontal:  
  * Retrieve scrolled value ;  
  sclCol = scrollbarHorizontal.value ;  
  
  * Ensure we do not go before first column ;  
  if sclCol lt 1 then  
    sclCol = 1 ;  
  
  rc = clearlist(dispColList) ;  
  
  * Ensure we do not go after last column ;  
  if sclCol le maxColBase then  
    rc = insertn(dispColList, sclCol , -1) ;  
  else  
    rc = insertn(dispColList, maxColBase, -1) ;  
  
  dataViewerBase._setLeftColumn(dispColList) ;  
  
  rc = clearlist(dispColList) ;  
  
  * If number of columns are not the same ensure we do not go over ;  
  * the limit for comp dataset ;  
  if sclCol le maxColComp then  
    rc = insertn(dispColList, sclCol , -1) ;  
  else  
    rc = insertn(dispColList, maxColComp, -1) ;  
  
  dataViewerComp._setLeftColumn(dispColList) ;  
return ; * scrollbarHorizontal ;  
  
*****;  
* Scroll between observations  
*****;  
scrollbarVertical:  
  * Retrieve scrolled value ;  
  sclRow = scrollbarVertical.value ;  
  
  * Define the new top row to display ;  
  rc = clearlist(topRowlist) ;  
  rc = insertn(topRowlist, sclRow) ;  
  
  * Ensure we do not go after last column ;  
  if sclRow le maxRowBase then  
    dataViewerBase._setToprow(topRowlist) ;  
  
  * If number of columns are not the same ensure we do not go over ;
```

PhUSE 2010

```
* the limit for comp dataset ;
if scrlRow le maxRowComp then
    dataViewerComp._setToprow(topRowlist) ;
return ; * scrollbarVertical;
```

ADDITIONAL FEATURES

Once the basic of synchronizing both viewers is implemented other useful features can be added

- Display the number of observation in the dataset.
- For dataset in source control, allowing display of previous version of a dataset.
- Scrolling both viewer to a selected observation number or applying a where clause.
- Selection of the columns to display and their order.
- Traffic lightning on difference between both datasets when performing a proc compare
- Preferences on default library and columns selection

DATA VIEWER

The principle used above on using temporary copy of dataset can be applied with only one dataset viewer. The advantage is again that column order and subset can be keep while changes are made to the dataset. While it is not mandatory to select two datasets in data compare, the number of rows display is more limited due to the second viewer.

PL Tools Dataset Viewer Prototype Version 1 Build 01b

Libname: Sashelp ZIPCODE Number of obs: 41761 Use version: Version

	ZIP	Y	X	ZIP_CLASS	CITY	STATE	STATECODE	STATENAME	COUNTY	COU
1	00501	40.813078	-73.046388	U	Holtsville	36	NY	New York	103	Suffo
2	00544	40.813223	-73.049288	U	Holtsville	36	NY	New York	103	Suffo
3	00601	18.165950	-66.723627		Adjuntas	72	PR	Puerto Rico	1	Adju
4	00602	18.383005	-67.186553		Aguada	72	PR	Puerto Rico	3	Agua
5	00603	18.433236	-67.151954		Aguadilla	72	PR	Puerto Rico	5	Agua
6	00604	18.505289	-67.135899	P	Aguadilla	72	PR	Puerto Rico	5	Agua
7	00605	18.436149	-67.151346	P	Aguadilla	72	PR	Puerto Rico	5	Agua
8	00606	18.195616	-66.977377		Maricao	72	PR	Puerto Rico	93	Maric
9	00610	18.285948	-67.144161		Anasco	72	PR	Puerto Rico	11	Anas
10	00611	18.287716	-66.797578	P	Angeles	72	PR	Puerto Rico	141	Utua
11	00612	18.471326	-66.728149		Arecibo	72	PR	Puerto Rico	13	Areci
12	00613	18.472737	-66.719283	P	Arecibo	72	PR	Puerto Rico	13	Areci
13	00614	18.456904	-66.735892	P	Arecibo	72	PR	Puerto Rico	13	Areci
14	00616	18.426456	-66.673779		Bajadero	72	PR	Puerto Rico	13	Areci
15	00617	18.458792	-66.537234		Barceloneta	72	PR	Puerto Rico	17	Barc
16	00622	18.017976	-67.170627		Boqueron	72	PR	Puerto Rico	23	Cabo
17	00623	18.086721	-67.151780		Cabo Rojo	72	PR	Puerto Rico	23	Cabo
18	00624	18.063481	-66.723612		Penuelas	72	PR	Puerto Rico	111	Penu
19	00627	18.485866	-66.844057		Camuy	72	PR	Puerto Rico	27	Cam
20	00631	18.180272	-66.832917	P	Castaner	72	PR	Puerto Rico	81	Lares
21	00636	18.162126	-67.077928	P	Rosario	72	PR	Puerto Rico	125	San
22	00637	18.081674	-66.968022		Sabana Grande	72	PR	Puerto Rico	121	Saba
23	00638	18.338774	-66.466418		Ciales	72	PR	Puerto Rico	39	Ciale
24	00641	18.268891	-66.704425		Utua	72	PR	Puerto Rico	141	Utua
25	00646	18.466563	-66.270521		Dorado	72	PR	Puerto Rico	51	Dora
26	00647	17.967309	-66.939935		Ensenada	72	PR	Puerto Rico	55	Guar
27	00650	18.362641	-66.565388		Florida	72	PR	Puerto Rico	54	Florid
28	00652	18.466248	-66.569371		Garrochales	72	PR	Puerto Rico	13	Areci
29	00653	17.974015	-66.908699		Guanica	72	PR	Puerto Rico	55	Guar
30	00656	18.025970	-66.787588		Guayanilla	72	PR	Puerto Rico	59	Guay

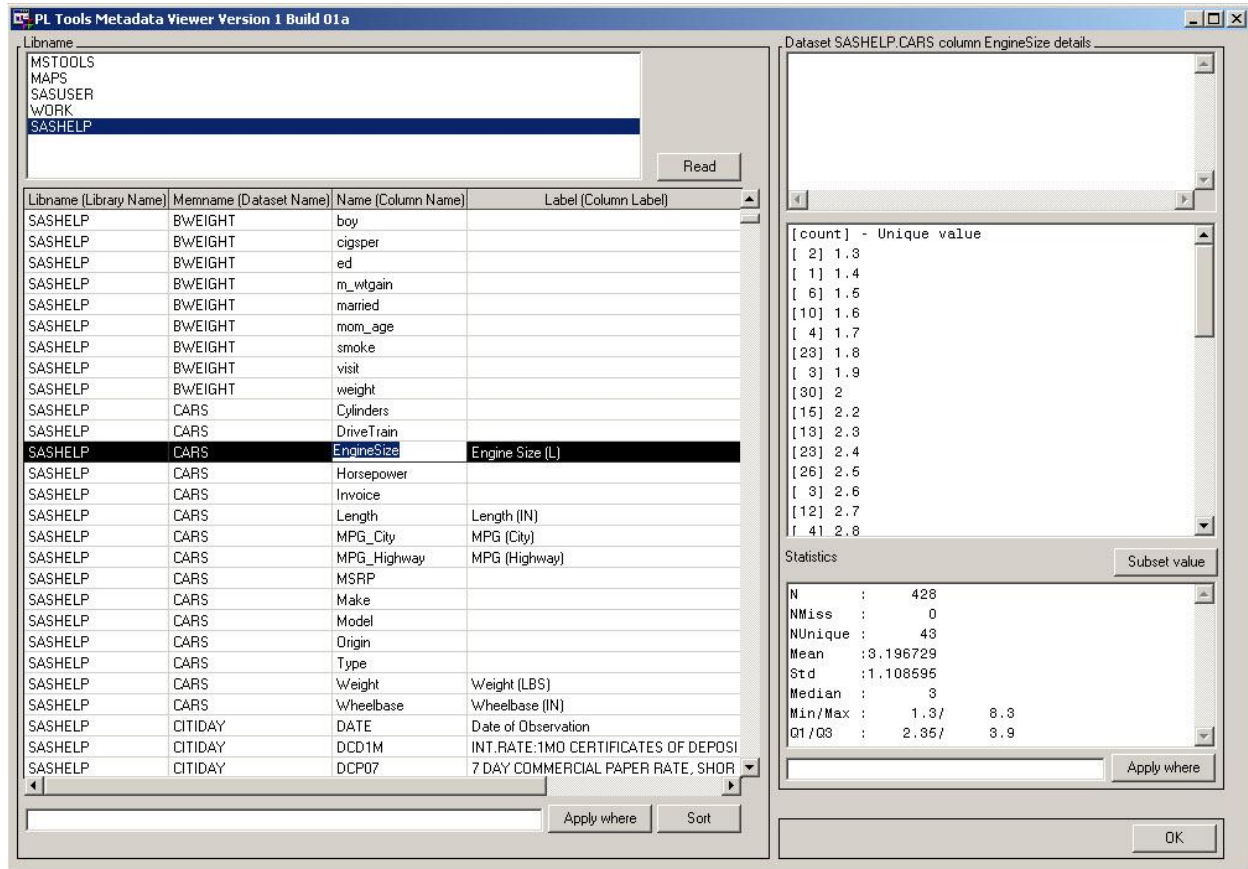
Find OBS: Find Where: Apply Hide/Unhide Refresh OK

METADATA VIEWER

Being able to easily find in what dataset information is very important when one need to performs analysis. When it has to be done with unfamiliar data structure it can become time consuming to actually find the location of the required information across multiple datasets and library.

The dictionary table dictionary.columns or the equivalent sas view sashelp.vcolumn are very useful in this situation. The metadata viewer frame does provide a GUI on the dictionary table. It has the ability to quickly perform basic data statistics any column.

PhUSE 2010

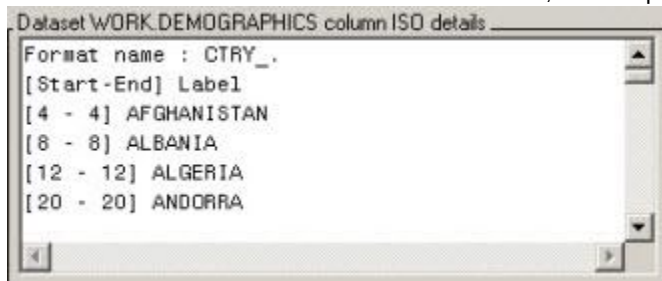


READ METADATA ON SELECTED LIBRARY

Once a user has selected one or more library, the available metadata is read from the dictionary.columns table and displayed in the viewer. Sort of the of columns can be changed and a subset apply to restrict the column displayed.

DATA REVIEW

When a column is selected the unique values for the column, number of occurrence for each value are display. If there is a user defined format attach to a column, the complete format contents is shown.



BASIC STATISTICS

Depending on the selected column data type, the displayed statistics will be different.

For character columns statistics are number of value, number of missing values and number of unique value are displayed.

For numeric columns are number of value, number of missing values and number of unique value, Mean, Standard deviation, Median, Min/Max, Q1 and Q3 are derived.

CONCLUSION

SAS/AF module is a powerful application development tool and can be used to develop useful tools. Relatively simple application can improve data accessibility to people who are not regular SAS users as shown with the metadata viewer. Other tools like the data compare have there usefulness in validation process.

PhUSE 2010

ACKNOWLEDGMENTS

The author would like to thank Georges Ette, programmer in M&S programming group for his thoughtful comments, continual support and encouragement in the development of the toolkit.

RECOMMENDED READING

Recommended reading lists go after your acknowledgments. This section is not required.

CONTACT INFORMATION

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

Patrick Lupien
Novartis Pharma AG
Postfach
CH-4002 Basel
Email: patrick.lupien@novartis.com

Brand and product names are trademarks of their respective companies.