#### PharmaSUG 2018 - Paper DV-12

# Using ODS LAYOUT, GTL, and ODSTEXT to Generate a Compact Graphical Codebook

Shane Rosanbalm, Rho, Inc.

## ABSTRACT

In this paper we present a SAS® macro that will produce a compact graphical PDF summary of all of the variables in a SAS dataset. The report is considered compact because each variable summary is confined to a 1" tall strip. The report is considered graphical because, in addition to presenting metadata and statistics for each variable, the report also includes a small graph of the distribution for each variable: a histogram for continuous variables and a bar chart for categorical variables. Not to be mistaken for a data dictionary, this compact graphical report is designed to allow the user to quickly gain familiarity with the variables in a dataset.

## **INTRODUCTION**

A codebook is a report which provides information about the variables in a dataset. The information presented varies from one codebook package to another, but codebooks typically contain information such as: the variable type, label, number of cases, each level of categorical variables, means and other statistics for continuous variables, etc.

There are already many very good SAS®-based codebook programs out in the world, so why create yet another codebook program? The short answer is, "**pictures!**"



age: Drv: Age At Consent Date {type=num} n=975, missing=0, unique=62 Mean (StdDev) = 55.2 (11.19) Min, Max = 19, 83 Median [Q1, Q3] = 57 [49, 63]

Figure 1. Sample continuous variable output



race: Race {type=num, fmt=RACE2F.} n=975, missing=0, unique=10 Frequencies (top 5 only): 1 = WHITE (670, 68.7%), 19 = CHINESE (125, 12.8%), 21 = KOREAN (91, 9.3%), 2 = BLACK/AFRICAN AMERIC... (53, 5.4%), 91 = ASIAN OTHER (18, 1.8%)

Figure 2. Sample categorical variable output

# **MACRO BASICS**

The simplest application of the codebook macro involves only one parameter: data=.

%codebook\_generic(data=examples.cars);

This macro call will generate a file named examples.cars.pdf. The dataset being summarized is nothing more than a local copy of the SASHELP.CARS dataset.

Toyota	- 0	make: no label {type=char}
Chevrolet	- 0	n=428. missing=0. unique=38
Aercedes-Benz	- · · ·	Frequencies (top 5 only): Toyota (28, 6.5%), Chevrolet (27, 6.3%), Mercedes-Benz (26, 6.1%), Ford (23, 5.4%),
Ford	- 0	BMW (20, 4.7%)
BMA		
Dates		
	036 234 436 634	
		model: no label {type=char}
	Too many unique values.	n=428, missing=0, unique=425
	No graphic will be produced.	First 5: 3.5 RL 4dr, 300M 4dr, 325Ci 2dr, 325i 4dr, 325xi 4dr
		Last 5: Z4 convertible 3.0i, Z4 convertible 2.5i, Wrangler Sahara con, Viper SRT-10 conver, Tundra Regular
		Cab
Sedan -	o	type: no label {type=char}
SUV -	0	n=428 missing=0 unique=6
Sports -	0	Frequencies (top 5 only): Sedan (262, 61,2%), SUV (60, 14,0%), Sports (49, 11,4%), Wagon (30, 7,0%), Truck (24,
Vanon - e		5.6%)
Test		
111CK - 0		
0%	20% 40% 50%	
		origin: no label {type=char}
Asia -	0	n=428. missing=0. unique=3
USA -	0	Frequencies: Asia (158, 36,9%), USA (147, 34,3%), Europe (123, 28,7%)
urope	0	
0%	10% 20% 30%	
		drivetrain: no label {type=char}
ront -	0	n=428. missing=0. unique=3
Rear -	0	Erequencies: Front (226, 52,8%), Rear (110, 25,7%), All (92, 21,5%)
		readmentations react features and starts rate ( 1104 more rate of the reaction rate
All -	0	
0% 1	0% 20% 30% 40% 50%	
014		marry na label (hma-num fut-DOLLADS)
0%		msrp: no label {type=num, fmt=DOLLAR8.}
0% -		n=428, missing=0, unique=410
0%-		Mean (StdDev) = \$32,775 (19431.72) Min, Max = \$10,280, \$192,465
0%		Median [Q1, Q3] = \$27,635 [\$20,330, \$39,215]
\$0 \$	50.000 \$100.000 \$150.000 \$200.000	
0%-	1	invoice: no label {type=num, fmt=DOLLAR8.}
0%-		
V70 -		11-420, 1115511y-0, u11que-423

Figure 3. Screen shot of examples.cars.pdf

# PARAMETERS

The next simplest application of the codebook macro involves only one parameter: library=.

%codebook\_generic(library=examples);

This macro call will generate one PDF file for every dataset in the EXAMPLES library. In this example, the EXAMPLES library is nothing more than a local copy of the SASHELP library.

- examples.cars.pdf
- examples.class.pdf
- etc.

The macro has roughly a dozen optional parameters that can be used to modify the output. Some of the parameters that you're more likely to take advantage of are:

- pdfpath= The folder in which to save the PDF file.
  - In case you do not want the PDF in the same directory as the dataset.
- catplot= Type of categorical plot (dot vs. hbar).
  - o In case you prefer (inferior) horizontal bar charts to (superior) dot plots.
- maxfreqs= Maximum number of categories to show for frequencies.
  - o By default only the 5 most commonly-occurring categories are displayed.
- plotheight= Height of plot in inches.
  - If you increase maxfreqs=, then you might need to make the plot taller to avoid losing tick marks on the yaxis.
- appendix= Include an appendix of all categorical values (yes vs. no).
  - In case you want to see more categories than maxfreqs= is willing to show you.

# SOURCE CODE

The macro source code, along with several examples, is available on Github.

#### https://github.com/RhoInc/sas-codebook

Once you have arrived on the <>Code tab, select the [Clone or download] button at right and choose the [Download ZIP] option.

RhoInc	/ sas-codeboo	k				⊙ Unwatch →	7	★ Star	1	¥ Fork	0
<> Code	() Issues (11)	1) Pull requests (0)	III Projects 0	🗐 Wiki	<u>ति Insights</u>	Settings					
SAS macr	o for generating	a concise summary o	of every variable	in a SAS da	taset.						Edit
						11 3 contributors			di M	MIT	
0.0	4 commits	P I branch	~	10 releases		a s contributors			eTe IM	1.1	
Branch: mast	er • New pull rec	quest	~	10 releases	Create	new file Upload file	s Fin	d file	Clone	or downlo	ad <del>-</del>
Branch: mast	er • New pull red	quest		10 releases	Create	new file Upload file	rs Fin	d file	Clone	or downlo. Use	ad 🕶 SSH
Branch: mast	er   New pull red Update LICENSE.md Files	quest redirect ods lis	ting to work directo	10 releases	Create	new file Upload file Clone with HTTF Use Git or checkout	PS ③ with SVI	d file	Clone the web	or downlo. Use o URL.	ad <del>-</del> SSH
Branch: mast	Ver view pull red Update LICENSE.md Files	redirect ods lis screen shot	ting to work directe	10 releases	Create	new file Upload file Clone with HTTF Use Git or checkout https://github.co	PS ③ with SVI	d file N using ti nc/sas-c	Clone Clone he web	or downlo Use URL. ok.git	ad <del>-</del> SSH
Branch: mast Ttbailey Example MageFil Macros	Files	redirect ods lis screen shot make options s	ting to work directo	ory nal	Create	New file Upload file Clone with HTTF Use Git or checkout https://github.co	rs Fin PS ③ with SVI om/RhoI	d file N using th nc/sas-c	Clone of the web odeboo	or downlo Use URL. ok.git	ad + SSH

The complete set of optional parameters is detailed on the GitHub wiki.

### https://github.com/RhoInc/sas-codebook/wiki/Generic-Datasets

The wiki provides a self-contained example program that should flatten out the learning curve.

### https://github.com/RhoInc/sas-codebook/wiki/Generic-Example

## CONCLUSION

The macro presented in this paper produces compact graphical PDF summaries of the variables in a SAS dataset. Each variable is summarized in a compart 1" tall strip that includes a graph of the variable distribution, metadata, and summary statistics. If you're looking for a data dictionary, this is not the macro for you. On the other hand, if you just want to quickly get to know a dataset (before later moving on to a data dictionary), then maybe this macro is for you.

## **CONTACT INFORMATION**

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

Shane Rosanbalm Rho, Inc. srosanba@gmail.com