

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!**"

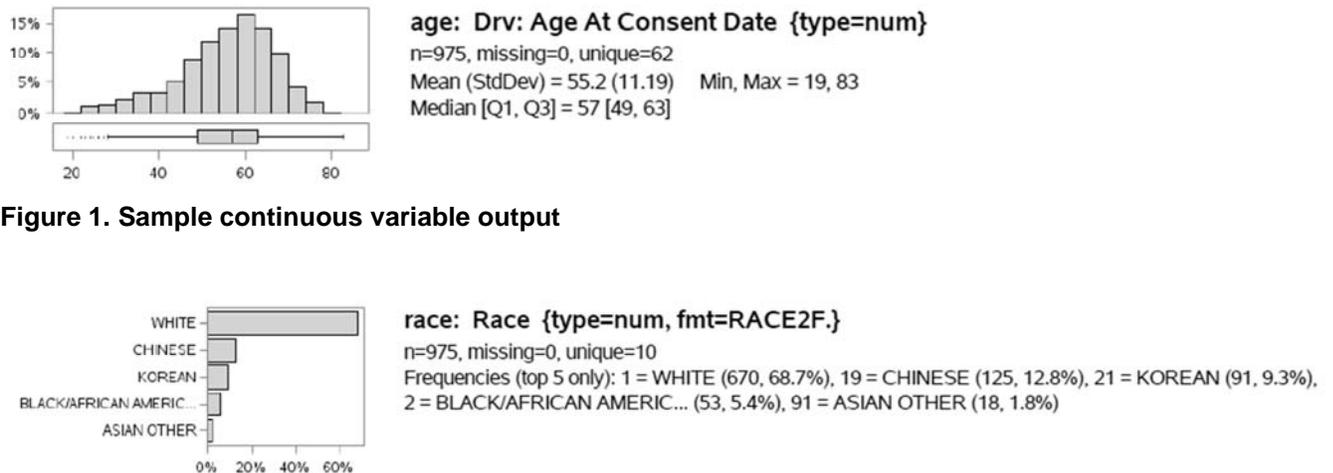


Figure 1. Sample continuous variable output

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.

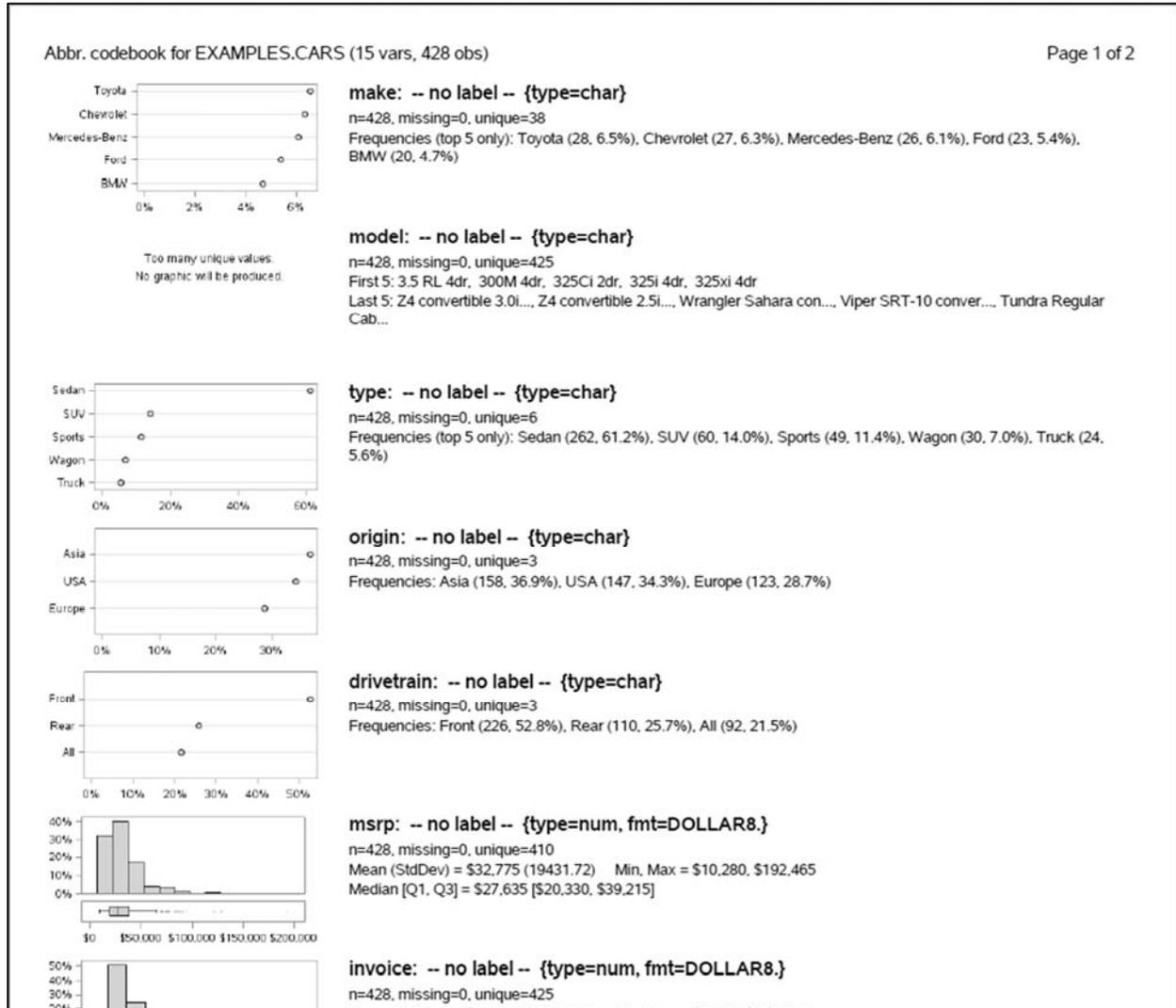


Figure 3. Screen shot of `examples.cars.pdf`

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:

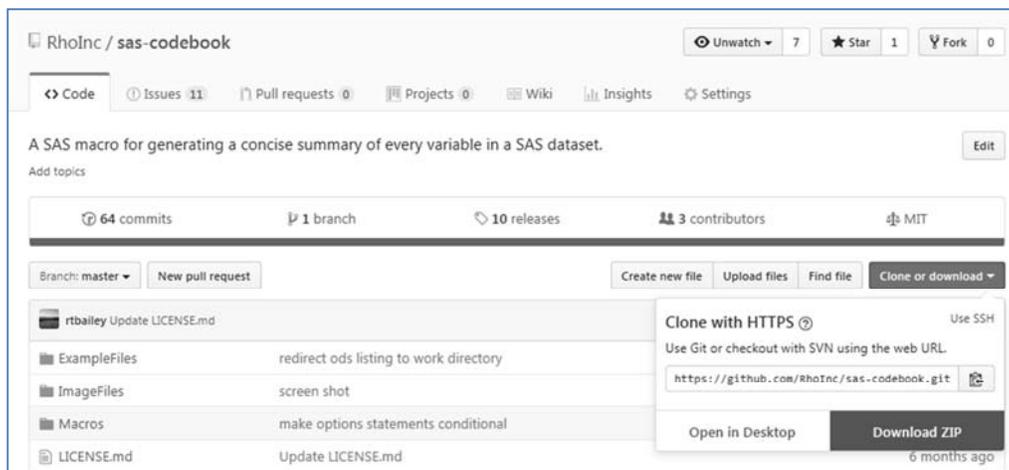
- `pdfpath=` The folder in which to save the PDF file.
 - In case you do not want the PDF in the same directory as the dataset.
- `catplot=` Type of categorical plot (dot vs. hbar).
 - In case you prefer (inferior) horizontal bar charts to (superior) dot plots.
- `maxfreqs=` Maximum number of categories to show for frequencies.
 - 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.



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 compact 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