PharmaSUG 2016 - Paper DG15

Elevate your Graphics Game: Violin Plots

Spencer Childress, Rho, Inc., Chapel Hill, NC

**[Note2Author**: This paragraph style, with yellow highlighting, shows notes and instructions to authors.

Formatting and editing notes:

1. Check your margins. The paper should print so that both the left and right margins are 1 inch each.
2. Please list each author on a separate line. Keep the company name, city, state, and country on the same line. If multiple authors are from the same company, repeat the information for each author.
3. Please copy the edited abstract that you submitted at <https://www.softconf.com/g/pharmasug2016/>
4. **Paragraphs**: The template has been changed this year. One change is the incorporation of white space as part of the paragraph style. You do not need to add an extra paragraph or press the ENTER key to create extra white space between paragraphs.
5. Figures, displays, outputs, and tables should have captions. See page 2 for instructions about inserting captions. Please use initial capitalization for the captions. The example captions in this sample have initial capitalization.
6. Page 2 of this paper sample has examples of formatted lists, tables, output, and displays that you can copy to use as a starting point. Page 2 also shows some simple instructions for inserting captions, cross-references, and graphics.

**Note**: Delete any text highlighted in yellow and page 2 before submitting the paper.

# Abstract

If you've ever seen a box-and-whisker plot you were probably unimpressed.  It lives up to its name, providing a basic visualization of the distribution of an outcome: the interquartile range (the box), the minimum and maximum (the whiskers), the median, and maybe a few outliers if you’re (un)lucky. Enter the violin plot.  This data visualization technique harnesses density estimates to describe the outcome’s distribution.  In other words the violin plot widens around larger clusters of values (the upper and lower bouts of a violin) and narrows around smaller clusters (the waist of the violin), delivering a nuanced visualization of an outcome. With the power of SAS/GRAPH®, the savvy SAS® programmer can reproduce the statistics of the box-and-whisker plot while offering improved data visualization through the addition of the probability density ‘violin’ curve. This paper covers various SAS techniques required to produce violin plots.

# Introduction

The introduction explains the purpose and scope of your paper and provides readers with any general information they need to understand your paper.

The box-and-whisker plot is a basic data visualization which with a little SAS magic can be improved drastically. The SAS programmer needs a few tools to round those hard corners. All data visualizations begin with the underlying data. Throughout this paper the dataset in reference is SASHELP.CARS, which contains qualitative and quantitative data on a number of vehicles.

The primary purpose of this paper is to illustrate similarities and differences between the box-and-whisker plot and the violin plot. Secondly I will discuss augmentations to the violin plot which provide additional information about the data. Producing a violin plot in SAS requires kernel density estimates, descriptive statistics, some data manipulation, and PROC SGPANEL, and each will be thoroughly explained.

# Box-and-whisker plot

The box-and-whisker plot gives a quick outline of the distribution of continuous data. It’s a visualization of the five-number summary, i.e. the sample minimum, first quartile, median, third quartile, and sample maximum. Figure 1 displays an example of the box-and-whisker plot with horsepower as the outcome, continent of origin as the group comparison, and number of cylinders as the panel comparison.

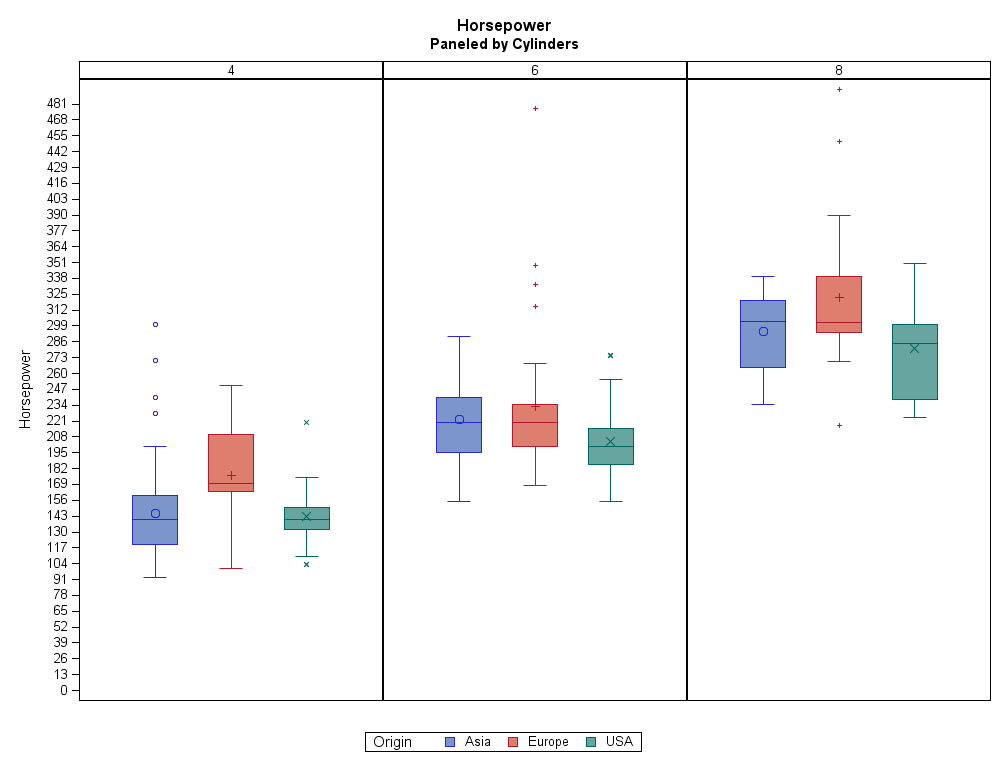


Figure 1. A Box-and-Whisker Plot

This plot visualizes the distribution of horsepower by number of cylinders and continent of origin. Each solid box encompasses all points between the first and third quartiles, otherwise known as the interquartile range. The “whiskers” encompass all points inside 1.5x the interquartile range. Points outside the interquartile range are considered outliers for the purposes of this plot. The following code will produce this plot:

**proc** **sort**

data = sashelp.cars (where = (cylinders in (**4** **6** **8**)))

out = cars;

by Cylinders Origin Horsepower;

**run**;

**proc** **sgpanel**

data = cars;

panelby Cylinders / novarname

rows = **1**;

vbox Horsepower /

group = Origin;

**run**;

It’s pretty basic and I know we can do better.

# Violin Plot

The violin plot is a box plot with a kernel density plot instead of a box. You might call it an outside-the-box plot. “In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable” (“Kernel Density Estimation,” 2016, para. 1). Don’t let that scare you away.

## Subhead A Level <heading 2>

This heading level is for a subtopic of a main topic. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

Table 1 is a sample table:

| **SAS Variable Format** | **DB2 Data Type** |
| --- | --- |
| $*w*.  $CHAR*w*. | CHARACTER |
| any date format | DATE |

Table 1. DBLOAD Procedure: Default DB2 Data Types for SAS Variable Formats

Use captions for tables.

# Second Main Topic <heading 1>

This is a main topic in the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

If you need to include a numbered or an ordered list:

1. This is a sample numbered or ordered list item. This is list item text. This is list item text. This is list item text. This is list item text. This is list item text.
2. This is a sample numbered or ordered list item. This is list item text.

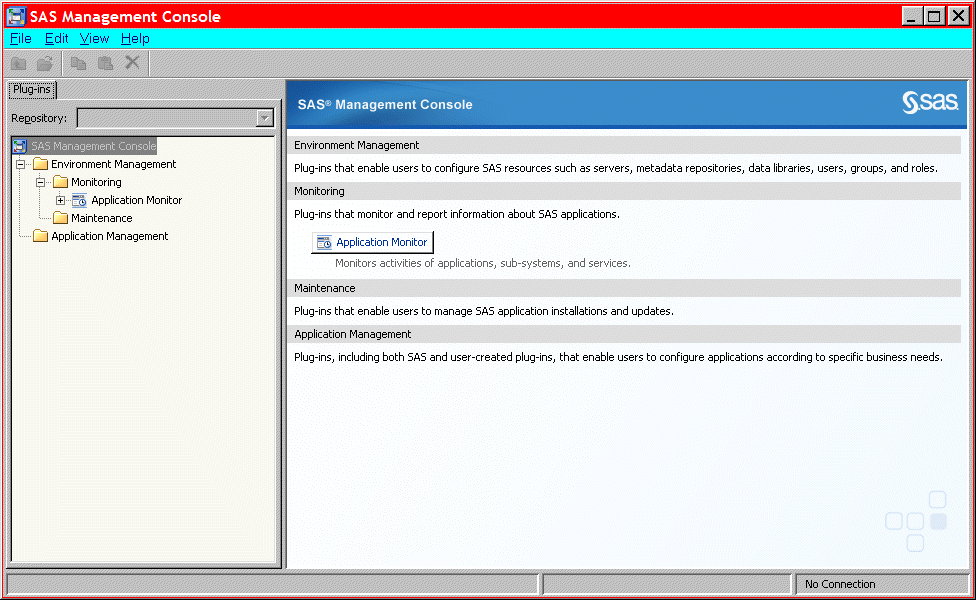
This is the paper body. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph. This is another sample paragraph.

If you need to include a bulleted or an unordered list:

* This is a sample bulleted list item. This is list item text. This is list item text. This is list item text. This is list item text. This is list item text. This is list item text. This is list item text. This is list item text.
* This is a sample bulleted list item. This is list item text.

Continuation of paper body—after an unordered list. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

Display 1 is an example of a display or screen capture.



Display 1. Former Main Interface for SAS Management Console

Use captions for displays.

## Subhead A Level <heading 2>

This is subtopic for the above. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

If you need to include SAS output, this is an example of how to present it:

Output 1 shows an example of how to present output.

CREATE TABLE ALLACCTX(SourceSystem varchar(4),

cctnum numeric(18,5) CONSTRAINT "ALLACCT\_PK" PRIMARY KEY,

ccttype numeric(18,5),balance numeric(18,5),clientid numeric(18,5),

losedate date,opendate date,primary\_cd numeric(18,5),status varchar(1))

Output 1. Output from a CREATE TABLE Statement

Use captions for output. Note that output is the same font as source code, but it is in a box. (Not a Text box.)

Continuation of paper body—after output.

## Subhead A Level <heading 2>

This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

### Subhead B Level <Heading 3>

This heading level is for a subtopic of a secondary topic. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

#### Subhead C Level <Heading 4>

This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

# Conclusion <heading 1>

The conclusion summarizes your paper and ties together any loose ends. You can use the conclusion to make any final points such as recommendations, predictions, or judgments.

This is the text for the paper’s conclusion. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

# References <heading 1>

This section is not required. Use a bulleted list if you have more than one reference. The references below are examples and do not cover the spectrum of examples that might be included. The important thing is to be consistent in the formatting and organization of your references. If you prefer, you can follow a bibliographic approach such as the American Psychological Association (<http://www.apastyle.org/manual/index.aspx> ) or the American Statistical Association (<http://www.amstat.org/publications/jse/jse_author_info.htm> )

This is the text for the references.

Book <Author name: last name, first name>. <Copyright date>. <*Book title*>. <page numbers>. <City, State of Publisher>: <Publisher name>

Journal article <Author name: last name, first name>. <Copyright date>. “<Article title>.” <*Journal title*>. <page numbers>. <City, State of Publisher>: <Publisher name>

Article in conference proceeding <Author name: last name, first name>. <Copyright date>. “<Article title>.” <*Title of proceedings such as Proceedings of the SAS Global 2010 Conference*>. <City, State of Publisher>: <Publisher name> Optional: You can add a URL to access available online copies. For example: Available at <http://support.sas.com/resources/papers/proceedings09/TOC.html>.

Web site <Author name: last name, first name>. “<Title>.” <*Source*>. <Date>. Available at <URL>.

# Acknowledgments <heading 1>

This section is not required.

This is the text for the acknowledgments. This is the paper body. This is the paper body. This is the paper body. This is the paper body. This is the paper body.

# Recommended Reading <heading 1>

This section is not required. Use a bulleted list if you have more than one reference. This is the format for recommended reading.

* Base SAS® Procedures Guide
* SAS® For Dummies®

# Contact Information <heading 1>

In case a reader wants to get in touch with you, please provide your contact information.

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

Name:

Enterprise:

Address:

City, State ZIP:

Work Phone:

Fax:

E-mail:

Web:

Twitter:

The next two paragraphs are **required** and need to remain in the paper.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Page of Formatted Samples

This page has samples that you can copy into the body of your paper and adapt as necessary for your content.

**Note**: Delete this page before submitting your paper.

**Source Code Sample**

data one;

set two;

if mix(var1, var2) > 0 then do;

**List: Numbered or Ordered**

1. numbered list item
2. numbered list item
3. numbered list item

**List: Bulleted or Unordered**

* This is a sample bulleted list item.
* This is a sample bulleted list item.

**Output Sample**

CREATE TABLE ALLACCTX(SourceSystem varchar(4),

cctnum numeric(18,5) CONSTRAINT "ALLACCT\_PK" PRIMARY KEY,

ccttype numeric(18,5),balance numeric(18,5),clientid numeric(18,5),

losedate date,opendate date,primary\_cd numeric(18,5),status varchar(1))

Output 2. Output from a CREATE TABLE Statement

**Table Sample**

| **Heading for Column 1** | **Heading for Column 2** | **Heading for Column 3** | **Heading for Column 4** |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table 2. Sample Table

Basic Instructions to Insert Captions, Cross-References, and Graphics

These instructions are written for MS Word 2007 and 2010. The steps are similar for MS Word 2003.

To insert a caption:

1. Click **References** on the main Word menu.
2. Click **Insert Caption**.
3. Select the **Label** type you want.
4. Click **OK**.

To insert a cross-reference:

1. Click **References** on the main Word menu.
2. Click **Cross-reference**.
3. In the **Reference type** list box, select Figure, Table, Display, or Output.
4. In the **For which caption** list, select the caption you want.
5. From the **Insert reference to** list, select **Only label and number**.

To insert a graphic from a file:

1. Click **Insert** on the main Word menu.
2. Click **Picture**.
3. In the Insert Picture dialog box, navigate to the file you want to insert.
4. When the name of the file you want to insert is displayed in the **File name** box, click **Insert**.