

VIEWS

International SAS Programmer Community

News, Hints, Tips and Information
for SAS® Users

Issue 67, 1st Quarter 2022

Happy New Year! This newsletter issue will be a graphics special, as it contains an article from our new VIEWS Consultant Robert Allison (The Graph Guy) about cleaning up your graphs, and I've written a solution to creating box plots with geometric means. As usual I have publicised some SAS-related events for the next few months, and I have also added some more SAS formats, options and functions that you may or may not have come across before, but you might find interesting, and even useful.

If you would like to contribute an article, to re-visit and improve an existing article, or just discuss the possibility of doing so, please feel free to send an email to me at views-uk@hollandnumerics.org.uk.

Philip R Holland (Newsletter Editor)

Membership of VIEWS is **free** and you can register for notification of all future VIEWS News issues by registering at blog.hollandnumerics.org.uk, and then emailing views-uk@hollandnumerics.org.uk, or using the blog site menu, to request access to the VIEWS UK forum. Members will receive an email notification whenever a new issue of the free quarterly VIEWS News is published on the web site, please inform us when you change your email address, otherwise your registration may be cancelled.

Did You Know?

If you have a useful hint or tip, please send it to the Editor, and share it with the VIEWS membership.

Polishing Your Graphs

Did your family polish rocks with a rock tumbler back in the 1970s? If you are not old enough to know what that is, do a web search (here is a link, www.youtube.com/watch?v=KOT375x8t4E, to save you some time). As you can see, the rocks went in all plain and ordinary, and then the same rocks came out looking spectacular, they were pleasantly shiny and smooth, and you could see much more of the interior detail and natural beauty. With a little extra effort, you can “polish” your graphs, and make them look better too.

Titles

The first bit of polish I recommend is a title that clearly describes what is in your graph. And if your graph is in an image file (such as PNG), make the title part of the image, rather than separate text in the document. This way, if someone copies the graph, they will also have the title. How many times have you seen graphs shared on social media (Reddit, Facebook, etc.) that do not have a title. It is very frustrating! Here is an example of a portion of a map, showing a title:

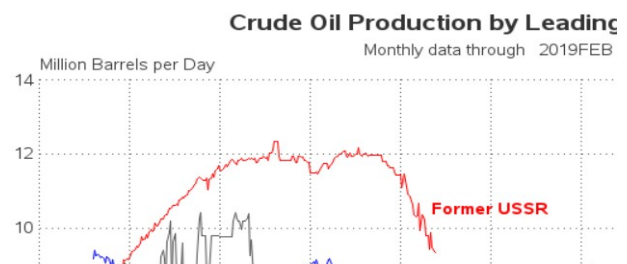
Hurricanes - Category 4 and Higher

That Passed Over North Carolina with 100+ mph Winds



Axis Integer Values

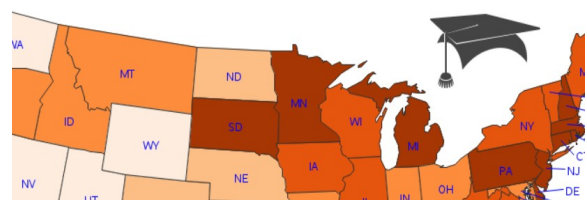
These days it is easy to create a graph, just drop the data into something like Microsoft Excel, and take whatever defaults it gives you. But often, with numeric values, the defaults might show unnecessary decimal places along the axis (for example, rather than a '12' it might show '12.0'). If the values at the axis tick marks are all integers, I recommend not showing a decimal place. Showing the '.0' takes up extra space, and makes the graph more cluttered (which takes more time to read). Here is an example of a piece of a graph showing good (integer) axis values:



Images

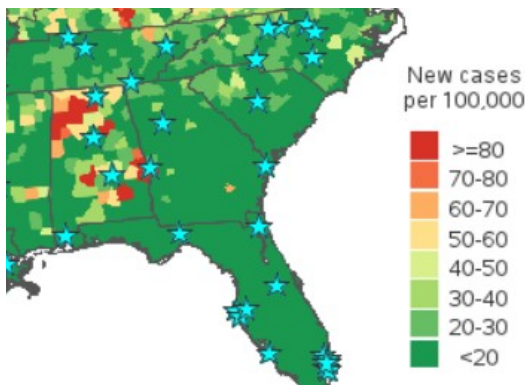
Users might only glance at a graph for a few seconds, to determine whether it is of interest to them. Therefore it is sometimes helpful to add an image/glyph/icon to help the user quickly know what the graph is about (without even having to read the title). These images should be simple, and should blend in well with the graph. Using images with the same colour background as the graph, or transparent background, often helps. For example, here is an example of a graph with an image. Can you quickly tell what general topic the graph is about?

Average Student Loan Debt per Borrower (2019)



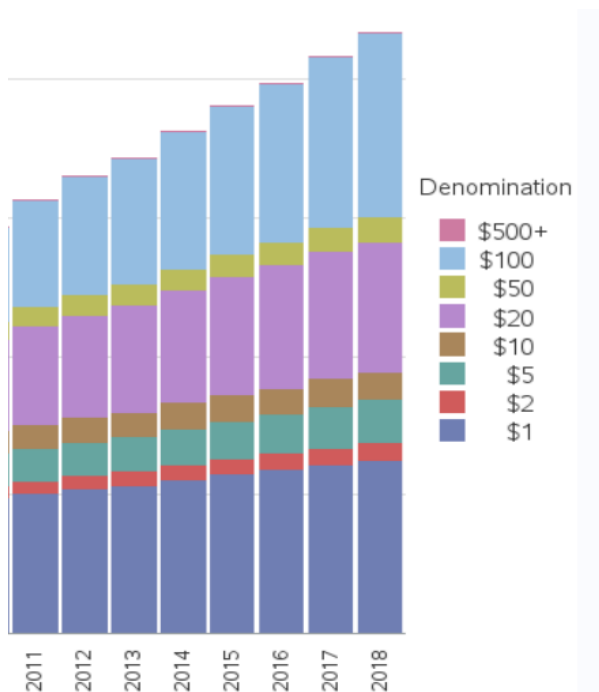
Legend Numeric Order

Colour legends usually show the data values sorted from the minimum to the maximum. If it is a horizontal legend, the smallest values are on the left and the largest are on the right, and that works well. But if the legend is arranged vertically, the smallest values are usually at the top (by default), and the largest at the bottom. I think the vertical legend is more intuitive with the smaller values at the bottom, like this:



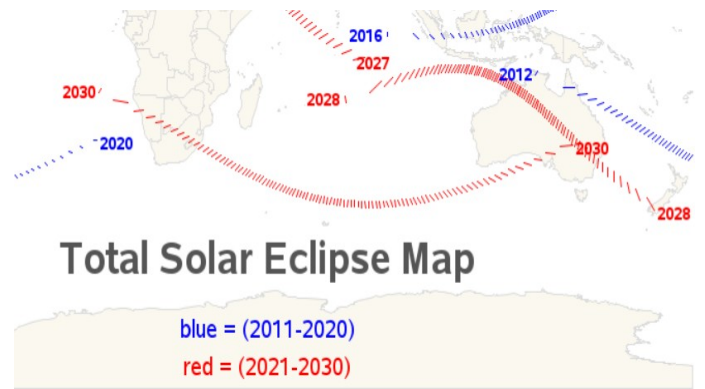
Stacked Bar Legend

With a stacked bar chart, I recommend ordering the legend items so they are “stacked” in the same order as the bar segments. This makes it easier to relate the legend to the chart. Here is a piece of a graph that demonstrates this technique:



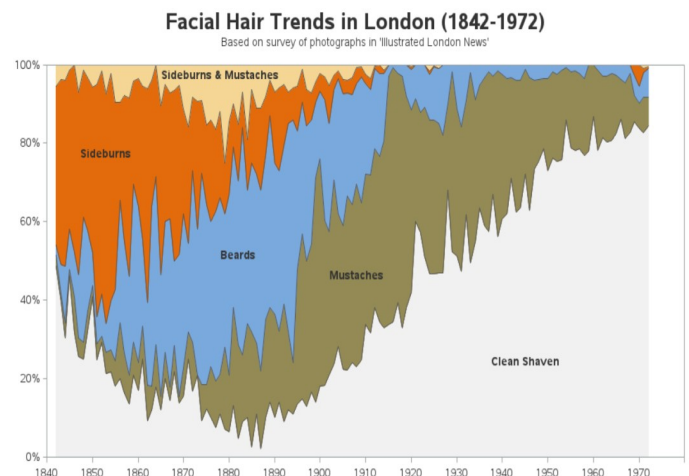
Utilise White Space

In some graphs and maps, there is often a lot of “white space” wasted on the page. You can often place the title and legend in that space, rather than above or below the graph. It is more efficient, and allows the graph to utilise more of the available space for the important/graphical part. Here is a piece of a map with the title in a space that would have otherwise been vacant/wasted:



Labelling Graph Directly

If there is enough room in the graph to label the areas or lines directly, rather than using a colour legend, I recommend doing that. This saves the user some time (looking between the graph and the legend), and also makes it more accessible for colour-blind users. Here is an example:



Using these techniques, hopefully you will be able to “polish” your graphs, so they shine like fine gemstones.

Robert Allison, The Graph Guy

Ask The Consultant

This part of VIEWS News is where you can get your technical questions answered. Send your questions to the Editor.

Box Plots with Geometric Means

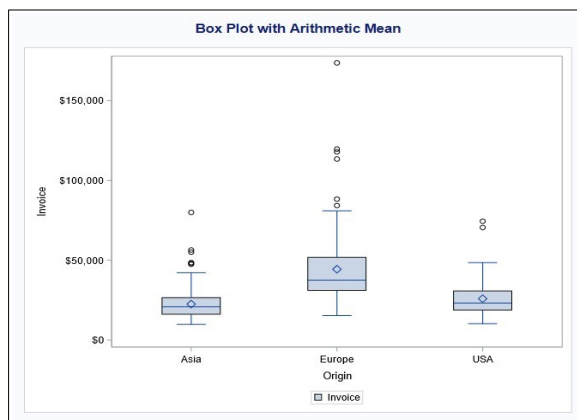
Q: How do you create a box plot with a geometric, instead of an arithmetic, mean?

A:[SAS 9] Here is some SAS code to generate a box plot using `sashelp.cars`, using the `&var` and `&byvar` macro variables to specify the variables for the plot. The `KEYLEGEND` statement will be used later to label the features.

```
%LET var = invoice;
%LET byvar = origin;

DATA cars;
    SET sashelp.cars;
RUN;
```

```
PROC SGPLOT DATA = cars;
  TITLE "Box Plot with Arithmetic Mean";
  VBOX &var. / CATEGORY = &byvar. name = 'vbox';
  KEYLEGEND 'vbox';
RUN;
```



An arithmetic mean is the average of the values over all records, but a geometric mean is the exponential of the average of the log values over all records, and is not directly available in the VBOX statement. As such the geometric mean can be calculated in SAS in a number of ways:

```
DATA cars;
  SET sashelp.cars;
  log&var. = LOG(&var.);
RUN;

** PROC UNIVARIATE **;
PROC UNIVARIATE DATA = cars ALPHA = 0.05 ALL;
  CLASS &byvar.;
  VAR &var.;
  OUTPUT OUT = univsumm GEOMEAN = &var._geomean
    MEAN = &var._mean P5 = &var._p5
    P95 = &var._p95;
  ODS OUTPUT BasicIntervals = univint;
RUN;

** PROC SQL **;
PROC SQL;
  CREATE TABLE sqlsumm AS
  SELECT
    &byvar.,
    MEAN(&var.) AS &var._mean,
    EXP(MEAN(LOG(&var.))) AS &var._geomean
  FROM cars
  GROUP BY &byvar.
  ORDER BY &byvar.;
QUIT;

** PROC SUMMARY **;
PROC SUMMARY DATA = cars NWAY;
  CLASS &byvar.;
  VAR &var. log&var.;
  OUTPUT OUT = summ1
    MEAN(&var. log&var.) = &var._mean
    log&var._mean
    P5(&var. log&var.) = &var._p5 log&var._p5
    P95(&var. log&var.) = &var._p95 log&var._p95
  ;
RUN;
```

All 3 output data sets contain the same values, so any of them can now be combined with the original cars data set to create the

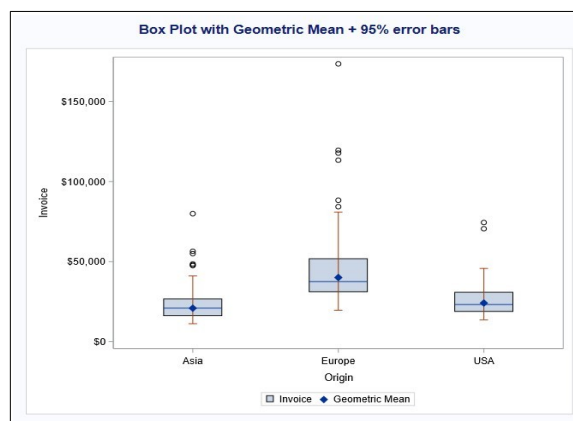
plotdata data set for PROC SGPLOT, but the output from PROC SUMMARY (summ1) has been used below. The geometric mean is calculated using the exponential of the log-mean, along with the 5th and 95th percentiles:

```
DATA summ2;
  SET summ1;
  &var._geomean = EXP(log&var._mean);
  &var._geop5 = EXP(log&var._p5);
  &var._geop95 = EXP(log&var._p95);
  LABEL &var._geomean = 'Geometric Mean';
RUN;

DATA plotdata (KEEP = &byvar. &var.);
  SET cars summ2;
RUN;
```

The code for the following box plot hides the arithmetic mean and whiskers, and replaces them with the geometric mean and error bars using a SCATTER statement:

```
PROC SGPLOT DATA = plotdata;
  TITLE "Box Plot with Geometric Mean "
    "+ 95% error bars";
  VBOX &var. /
    CATEGORY = &byvar. NOEAN
    WHISKERATTRS = (THICKNESS = 0)
    name = 'vbox';
  SCATTER Y = &var._geomean X = &byvar. /
    YERRORUPPER = &var._geop95
    YERRORLOWER = &var._geop5
    MARKERATTRS = (SYMBOL = DIAMONDFILLED)
    name = 'scatter';
  KEYLEGEND 'vbox' 'scatter';
RUN;
```



Notice that the positions of the whiskers and outliers look remarkably similar to the original box plot, although the error bars and original whiskers have different caps! This is due to the 5th and 95th percentiles being identical for arithmetic and geometric means, so the error bars are not really needed.

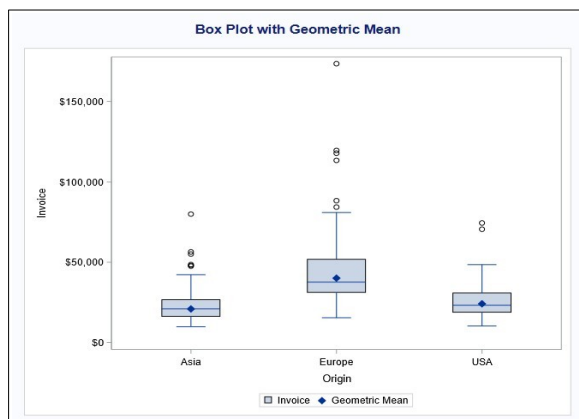
Mean, Geometric Mean and 95% error bars						
Obs	Origin	invoice_mean	invoice_p5	invoice_p95	invoice_geomean	invoice_geop5
1	Asia	\$22602	\$11209	\$41100	\$20934	\$11209
2	Europe	\$44395	\$19638	\$80939	\$40109	\$19638
3	USA	\$25949	\$13697	\$45815	\$24162	\$13697

Not hiding the whiskers and removing the error bars will return the look-and-feel of the box plot to that of the original image:

```

PROC SGPLOT DATA = plotdata;
  TITLE "Box Plot with Geometric Mean";
  VBOX &var. /
    CATEGORY = &byvar. NOMEAN name = 'vbox';
  SCATTER Y = &var._geomean X = &byvar. /
    MARKERATTRS = (SYMBOL = DIAMONDFILLED)
    name = 'scatter'
  ;
  KEYLEGEND 'vbox' 'scatter';
RUN;

```

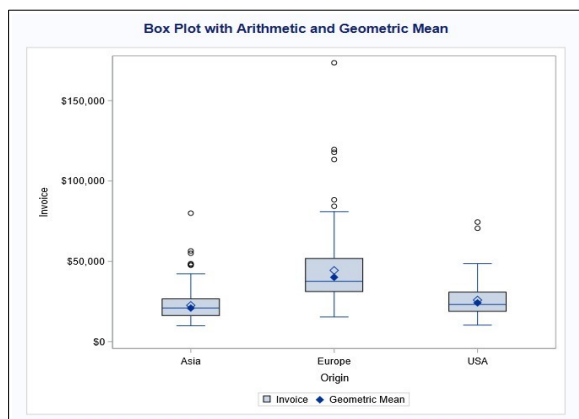


As adding the geometric mean is so easy, it can easily be added alongside the arithmetic mean to act as a comparison:

```

PROC SGPLOT DATA = plotdata;
  TITLE "Box Plot with Arithmetic and "
    "Geometric Mean";
  VBOX &var. / CATEGORY = &byvar. name = 'vbox';
  SCATTER Y = &var._geomean X = &byvar. /
    MARKERATTRS = (SYMBOL = DIAMONDFILLED)
    name = 'scatter'
  ;
  KEYLEGEND 'vbox' 'scatter';
RUN;

```



Philip R Holland (Newsletter Editor)

Formats, Options, and Functions

This section is devoted to the description of useful, or unusual, SAS Formats, Options, and Functions. Please read the SAS documentation to find out more about them.

[SAS] The SAS formats **\$HEXw.** and **HEXw.** convert character and numeric values to their corresponding hexadecimal (base 16) representations, respectively. The SAS informat **\$HEXw.** converts hexadecimal representations to character values.

[SAS] The PROC COMPARE options **METHOD** and **CRITERION** specify how numeric variable values are compared. **METHOD** can equal **ABSOLUTE** or **RELATIVE**, and **CRITERION** can be set to maximum absolute difference or relative ratio difference between the values compared that will be accepted as “equal”, respectively.

[SAS] The SAS functions **UPCASE()** and **LOWCASE()** convert character variables to upper-case and lower-case, respectively. The corresponding macro function **%UPCASE()** performs the same function within macro code. However, **%LOWCASE()** is not actually a macro function, but a SAS-supplied macro, which makes use of **LOWCASE()** inside.

In Brief

- Back issues of *VIEWS News* are available from the VIEWS UK forum at blog.hollandnumerics.org.uk. Please register for free membership on the blog, and then request free access to the VIEWS UK forum.
- The SAS Professional Forum group on LinkedIn, which currently has over 37,300 members. Everyone is welcome to request to join, but discussions here are restricted to SAS-related topics: www.linkedin.com/groups?gid=70702.
- The SAS Opportunities group on LinkedIn, which currently has over 14,700 members. Everyone is welcome to request to join, but discussions here are also restricted to SAS-related topics: www.linkedin.com/groups?gid=805217.
- In the UK and Ireland the SUGUKI group had been holding regular meetings in London, and also occasional meetings in Bristol, Dublin, Edinburgh, Leeds and Manchester. Unfortunately lockdown has necessitated running webinars instead. For more details and to register for future events please go to meetup.com/SUGUKI.
- Fancy joining the SAS EMEA Analytics Explorers, then please send your **country, name, email address** and **company** with your permission to pass these details on to SAS to views-uk@hollandnumerics.org.uk. Other regions include Asia Pacific, Canada, Latin America and US, but to join these communities based on where you live, you will now need to provide this information to a current member of these communities, instead of me.
- WPS Community Edition provides users with free access to WPS Analytics software including a SAS language compiler with coding and workflow tools on unlimited data volumes. A Community Edition licence key will be issued that will activate the software for up to six months. At the end of this term, users may be invited to renew their licence and to upgrade to the latest available version of the software. No technical support service is offered with the Community Edition. Download from www.worldprogramming.com/try-or-buy/editions. Note that the software is made available without warranty or indemnity, and is not made available to US residents.

Diary

Are you organising an event that would be of interest to the VIEWS readership? Let us know as we are interested in all non-profit making events related to SAS.

March 2022

31 **SAS Assemble**, Virtual

A virtual event for SAS Nordics and SAS UK & Ireland. See www.sas.com/sas/events/sas-assemble.html for more details.

April 2022

10-13 **SAS Innovate Americas**, San Diego, CA, USA

A face-to-face conference for Business. For more details please go to the SAS Events web site at www.sas.com/en_us/events/sas-global-forum.html.

Q2 2022

?? **SAS Innovate EMEA** (dates/venue to be confirmed)

A face-to-face conference for Business. For more details please go to the SAS Events web site at www.sas.com/en_us/events/sas-global-forum.html.

?? **SAS Innovate Asia Pacific** (dates/venue to be confirmed)

A face-to-face conference for Business. For more details please go to the SAS Events web site at www.sas.com/en_us/events/sas-global-forum.html.

Q3/Q4 2022

?? **SAS Explore**, Virtual

A virtual conference for SAS Technologists. For more details please go to the SAS Events web site at www.sas.com/en_us/events/sas-global-forum.html.

Contacts

Robert Allison, The Graph Guy

Email: robincary@gmail.com
Web: robslink.com

SAS UK

Tel: +44-(0)1628-486933
Web: www.sas.com/uk

VIEWS

Web Site

Web: blog.hollandnumerics.org.uk

Newsletter Editor

Email: views-uk@hollandnumerics.org.uk

***VIEWS News** is published quarterly for the VIEWS International SAS Programmer Community by Holland Numerics Limited.*