

# VIEWS

International SAS Programmer Community



## News, Hints, Tips and Information for SAS® Users

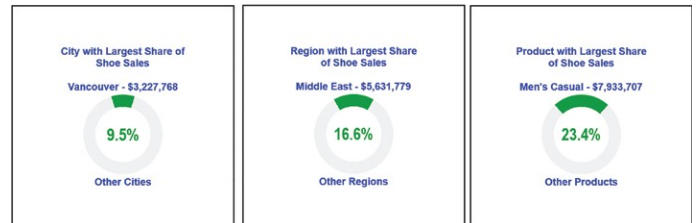
Issue 82, 2nd/3rd Quarter 2026

This newsletter is the home of Foundation SAS and brings you information about making the most of SAS 9 and Altair SLC. This newsletter issue includes an article from LeRoy Bessler about Donut Charts, Anders Sköllermo discusses quantiles for very large data sets, and there will be a launch of 3 Altair SLC books in Hamburg on 20Aug2026 and you are welcome to join the authors at lunchtime. 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](mailto:views-uk@hollandnumerics.org.uk).

**Philip R Holland (Newsletter Editor)**

donut charts in the Bank of America's Annual Report for its Year 2020. You can see them online on Page 31. Their layout is different, and the data is different. But they inspired me to create this triptych. A triptych is a three-part side-by-side composite usually seen in religious art.



Viewing this set of data graphics is facilitated by not having to "recalibrate your perception" from image to image. Consistency is always a communication asset.

Below is an enhanced version of the left hand side donut chart. The label for the "Other" bite now usefully includes its subtotal.

Membership of VIEWS is **free** and you can register for notification of all future VIEWS News issues by registering at [blog.hollandnumerics.org.uk](http://blog.hollandnumerics.org.uk), and then emailing [views-uk@hollandnumerics.org.uk](mailto: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.

## The Ineffable Power of Simplicity

In Issue 35, Third Quarter 2006, I showed you the Irrefutable Power of Simplicity with my "PacMan Pie Chart" in a pair of examples (from the early 1990's) that I call "The Extremes of Other". They are displayed at the bottom of this article for reference. In Case 1, the actual content of the Other slice was, at least by me, deemed insignificant and not worth exploring. In Case 2, the communication objective was to stimulate viewer curiosity about What IS in Other? What are the other entities contributing to my tax burden? In the 1990's, I served as a Trustee on the Fox Point Village Board.

In the donut chart examples below, the Other bite is simply a counterpoint to the donut bite that is the focus of the data graphic. The message in each chart is the title, the label of the small bite, and the small bite's percent of total in the hole label.

In March 2021, while working on Chapter 5 of my book (Pie Charts & Donut Charts), I found an interesting composite of four



Below is the code used to create the image.

```
ODS RESULTS OFF;
ODS _ALL_ CLOSE;

ODS LISTING GPATH="C:\temp" DPI=300;

PROC SUMMARY DATA=sashelp.shoes;
CLASS Region Subsidiary Product;
VAR Sales;
OUTPUT OUT=work.FromSummary SUM=;
RUN;
```

<https://hollandnumerics.org.uk/forums/forum/views-uk>

```

DATA work.Cities work.Regions work.Products;
SET work.FromSummary;
IF _TYPE_ EQ 0
THEN CALL
SYMPUT('GrandTotalSales',TRIM(LEFT(Sales)));
ELSE IF _TYPE_ EQ 1 THEN OUTPUT work.Products;
ELSE IF _TYPE_ EQ 2 THEN OUTPUT work.Cities;
ELSE IF _TYPE_ EQ 4 THEN OUTPUT work.Regions;
RUN;

PROC SORT DATA=work.Cities;
BY DESCENDING Sales;
RUN;

DATA work.MaxSalesCityAndOthers;
LENGTH CityLabel $ 26;
SET work.Cities;
IF _N_ EQ 1;
CALL SYMPUT('PctMaxCity',
TRIM(LEFT(PUT(Sales/
&GrandTotalSales,PERCENT6.1))));
/* It is necessary to use the PERCENT6.1 FORMAT,
to allocate enough width for (N.N%), where the
parenthesis is for negative, even though the
value is positive */
/* In DATA steps for MaxSalesRegionAndOthers
and MaxSalesProductAndOthers, use PERCENT7.1 */
CityLabel = TRIM(LEFT(Subsidiary)) || ' - ' ||
TRIM(LEFT(PUT(Sales,dollar10.)));
OUTPUT;
Sales = &GrandTotalSales - Sales;
CityLabel = 'Other Cities' || ' - ' ||
TRIM(LEFT(PUT(Sales,dollar11.)));
OUTPUT;
STOP;
RUN;

ODS GRAPHICS / RESET=ALL SCALE=OFF
WIDTH=3.65in HEIGHT=3.65in
IMAGENAME='LeftDonutHoleInfo_MaxSalesPct_City';

TITLE1 HEIGHT=11pt FONT='Arial/Bold' COLOR=Blue
'City With Largest Share of Shoe Sales';

TITLE2 HEIGHT=3pt COLOR=WHITE 'white space';

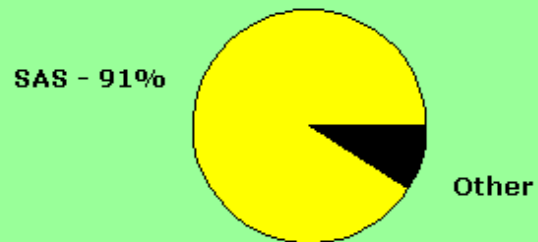
PROC SGPIE DATA=work.MaxSalesCityAndOthers;
STYLEATTRS DATACOLORS=(
CX009900 /* 60% from True Green to Black */
CXEEEEEE /* very close to White (CXFFFFFF) */
);
DONUT CityLabel / RESPONSE=Sales
SLICEORDER=RESPASC /* Start
with the smallest DONUT bite */
HOLELABEL="&PctMaxCity"
HOLELABELATTRS=(FAMILY='ARIAL' COLOR=CX009900)
/* The only HOLELABEL text controls are
FAMILY, COLOR,
WEIGHT (Bold or Default Normal) and
STYLE (Italic or Default Normal)
ODS Graphics determines the SIZE,
and determines the SAS FORMAT if the HOLELABEL
is numeric */
STARTANGLE=90
/* accepting the Default STARTPOS=CENTER,
i.e., center the starting bite at STARTANGLE */
DATALABELATTRS=(FAMILY='Arial' WEIGHT=Bold
SIZE=11pt COLOR=Blue)
DATALABELDISPLAY=(CATEGORY)
DATALABELLOC=OUTSIDE;
RUN;

ODS LISTING CLOSE;

```

Here are those simplest but powerful pie charts that I first developed in the early 1990's. These are the versions previously shown in VIEWS News Issue 35. The code for them is there. In 1993, R was probably part of "Other". Today, it can no longer be deemed insignificant.

### Case 1. Content of Other Is Insignificant Mainframe Data Analysis Market Shares - 1993



### Case 2. Content of Other Will Get Attention Where Your Property Tax Is Going This Year



LeRoy Bessler, Visual Data Insights™

## Ask The Consultant

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

## Fast and exact computation of all quantiles for very large amounts of data

**Q:** Can large amounts of data be converted into quantiles?

**A: [SAS 9.4]** We present three fast and robust methods for the computation of all specified quantiles for very large amounts of data. We have computed 15 quantiles for up to 2000 million datapoints on a Windows computer, where Word is a bit slow. Very little memory and disc space are used.

The basic idea is simple: Read the data values as few times as possible and sort as few values as possible!

**Definition of quantile:** Consider a file with N numeric values, sorted in increasing order. We denote them by Z1, Z2,

..., ZN. The data value for the p-quantile ( $0 < p < 1$ ) is defined as the value of the data point  $Z_j$ , where  $j$  is the smallest integer such that  $j \geq p \cdot N$ . The value  $p \cdot N$  is the X-index of the quantile,  $Z_j$  is the quantile point.

**Computation of quantiles:** The traditional way to compute quantiles is to sort the entire set of data in increasing order and then identify the data point with index  $j$  in the sorted set.

This requires huge resources of both time and memory when the amount of data is very large, and it works only up to a certain size of data. In our case we can handle amounts up to 500 million data points in this manner.

**Three robust methods that compute quantile points in different ways:**

- The *Slot-method* can be applied on data from a known, limited interval. The quantile points are determined with an absolute accuracy defined by the user.
- The *ES-method (Exact Slot)* for data from an unknown, unlimited interval. The quantile points are computed exactly. You may also compute any number of outliers.
- The *Hybrid-method* combines the Slot-method and the ES-method.

**Slots:** When we know the lower and upper limit of the interval, where all the data points lie, we split this interval into a set of small intervals, equally wide. We call these short intervals slots. An example might be yearly income in millions of Swedish crowns in the interval 0 to 8.

We define an array with one element per slot. We will store the number of data points that belong to each slot in the corresponding array element.

We read a data point from the large set of data. We compute the index of the corresponding slot using a simple linear transformation. We increase the value of the array element by one.

We do not store the data value, but continue with reading the next data point.

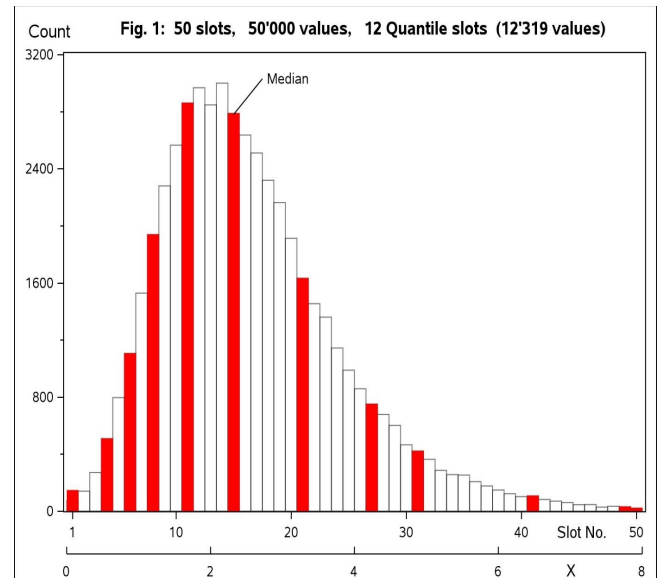
When all data points have been read, we have an array with the number of data points per slot. This array provides a good picture of how the data are distributed along the X-axis.

**Quantile Slot:** The next step is to determine which slot each quantile point belongs to. For each array element we compute the sum of all the preceding elements and add the value of the current element. We compare each quantile X-index ( $p \cdot N$ ) to the successive elements in this array of accumulated values. The index of the first element where the accumulated sum exceeds or equals the quantile X-index is the index of the quantile slot.

More than one quantile may belong to the same slot. We can also determine the order number for the quantile point within the slot, i.e. how many sorted data points we need within the quantile slot to reach the quantile point exactly.

Figure 1 shows a distribution of 50,000 data points in 50 slots with the quantile slots shown for 15 p-values: 0.001, 0.01, 0.1, 1, 5, 10, 25, 50, 75, 90, 95, 99, 99.9, 99.99, 99.999%.

Some of the extreme quantiles belong to the same quantile slot. We have used random data from a skew distribution. The quantile slots contain 12,319 data points all together.



**The Slot-method - Midpoint rule:** Each quantile data point is approximated by the midpoint of the corresponding quantile slot. We call this the Midpoint rule. The absolute accuracy is half the slot width. By choosing the number of slots large enough we can determine all the quantile points with a predefined absolute accuracy.

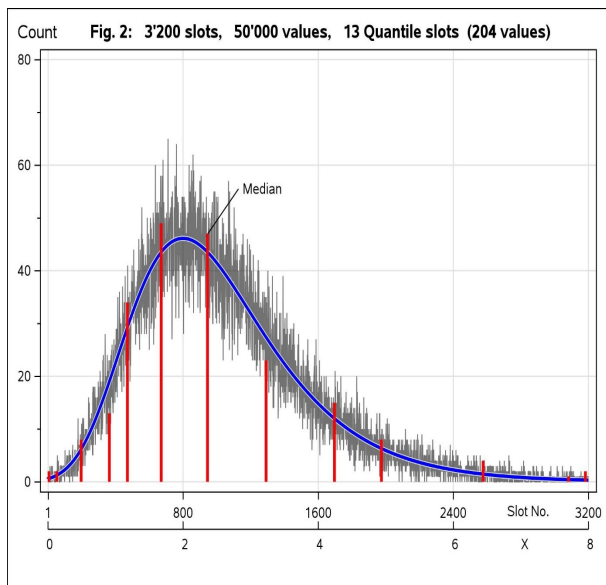
We read all the data once and we perform no sorting. Very little memory and disc space is required for this method.

It is tempting to interpolate within the slot to achieve even better accuracy, but it is in general not possible to improve the accuracy in that way. Random data are usually very unevenly distributed within the slots.

### Concepts and Terminology:

| Slot               | Interval on the X-axis                                                                     |
|--------------------|--------------------------------------------------------------------------------------------|
| N                  | Total number of data points                                                                |
| Quantile p-value p | Value between 0 and 1 (0 and 100%). Median = 0.50                                          |
| Quantile X-index j | Smallest integer $j \geq p \cdot N$                                                        |
| Quantile point     | The data point with index j, in the sorted data set. This is the X-value of the p quantile |
| Quantile slot      | X-interval that contains one (or several) data points.                                     |
| Quantile file      | Data file with all data points that belong to a quantile slot.                             |

Figure 2 shows how the data points are distributed when we use very small slots. The random data are "jumping up and down", even when the function (e.g. Normal or Gumbel) generating the random numbers, is very smooth.



The Slot-method needs to be extended to handle data from an unknown, infinite interval. We split the X-axis into 3 regions: a lower infinite region, a middle finite region and an upper infinite region.

**Region boundaries:** We need to determine the lower and upper boundaries of the middle region. We want this region to cover the main part of the data. Useful limits may be computed in the following way:

Analyze a small sample of the original data, say 5,000 values. Sort this small set and compute the 2% and 98% quantile points. Use these values as the lower and upper limits of the middle region.

**Extension 1 - The ES-method, Exact Slot, exact quantile points.** The idea is to use slots across the entire X-axis in the following way: We use slots of equal width in the entire middle region and slots in the lower and upper regions of expanding widths, as we move away from the middle region. The first slot in the lower region and the last slot in the upper region both have infinite width. We have chosen to let the slots closest to the middle region have the same width as the slots in the middle.

We use a simple linear transformation to map an X-value from the middle region to the appropriate slot index. For X-values from the lower and upper regions we use a mapping of the type  $1/x$ .

We proceed as before: We read a data point, determine which slot it belongs to and increase the corresponding slot array element by 1. Then we read the next datapoint. When all datapoints have been read we compute the slot index for each quantile slot - in the same way as for the simple Slot-method. We also compute the order number of each quantile point within its slot. No datapoints are saved at this stage.

To compute exact quantile points we need to sort the data points that belong to each quantile slot. We need to read all the data a second time. We read a data point and compute its slot index.

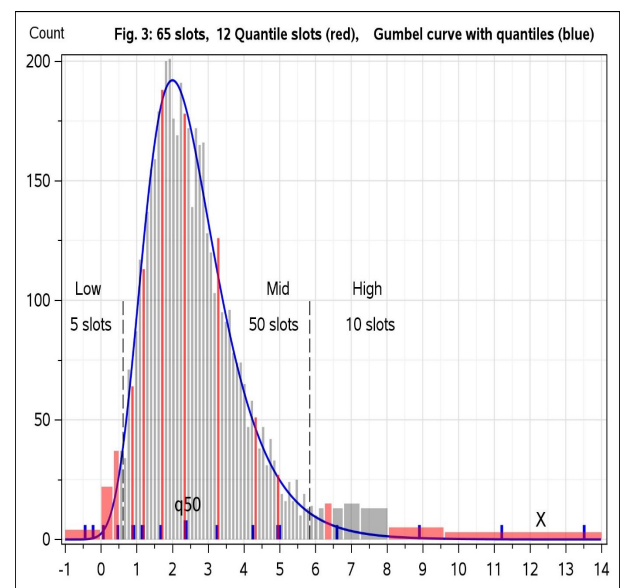
If this is a quantile slot we save the data value to a corresponding quantile file. Then we read the next data point. When all the data has been read a second time, we sort each small quantile files. Finally, we use the saved slot internal order number to identify the exact quantile point.

In the simple Slot-method, the width of the slots is determined by the accuracy requirement. In the ES-method we need instead to consider how large the quantile files may become - we want to avoid sorting very large files. In our program we estimate how dense the data points are and compute a reasonable choice of slot width in the middle region from this estimate.

The ES-method gives us the exact quantile points. It requires two reads of all the data and sorting of one file per quantile. Only a few data points are saved in the quantile files and very little memory, and disc space is used.

If we require outliers, we save the data values from the extreme slots in two outlier files and compute all wanted outliers (this is the default).

Figure 3 shows the X-axis split into 3 regions with 50 slots in the middle, 5 in the lower and 10 in the upper region.



**Extension 2 - The Hybrid-method, approximative and exact quantile points.** The Hybrid-method is a combination of the Slot-method and the ES-method. We use the simple Slot-method and compute approximative quantile points in the middle region using sufficiently narrow slots to obtain the desired accuracy in the quantile points. In the lower and upper regions, we use the ES-method and compute exact quantile points.

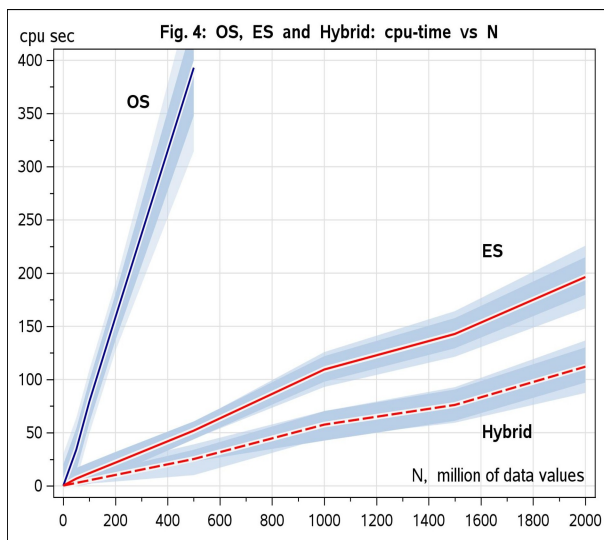
We want to avoid reading all the values that belong to the middle region a second time. Therefore, we save all the data values that belong to the lower and upper regions in two separate files already during the first read. We read only those two files a second time and create the quantile files that belong to the lower and upper regions. Finally, we compute the exact quantile points in the lower and upper regions.

We sort very few data values in the quantile files that belong to the lower and upper regions. Very little memory is used. Every quantile point that belongs to the lower or upper region is computed exactly. All quantile points that belong to the middle

region are computed to the required accuracy. All wanted outliers are computed.

The number of slots in the middle regions is determined by the accuracy requirements. The number of slots in the lower region is set as a percentage of the number of slots in the middle e.g. 10% - and similarly for the upper region. If the lower and upper regions together contain e.g. 4% of the data values the Hybrid-method means that we read data 1.04 times. We save approximately 4% of all the data values in the two additional files.

**Some results from test executions:** In the classical "Order Statistics" (OS)-method the quantile points are computed by sorting all the data points. In Figure 4 we compare the OS-, ES- and Hybrid-methods with respect to cpu-time, for increasing amounts of data. We use an infinite interval and compute 15 quantiles.



The blue shadowed areas around each curve indicate that we get cpu-times that vary by +/-10% or more when we execute the code many times. The OS-method fails and gives no results at all when we use more than 500 million data points. The ES- and Hybrid-methods have been tested for up to 2,000 million data points.

We required an absolute accuracy of 0.00005 for the Hybrid-method. The cpu-time increases linearly with increasing number of data values, for all three methods with very different slope of the line. The ES-method requires twice as long cpu-time as the Hybrid-method.

Both the ES- and the Hybrid-methods require very little memory, whereas the OS-method requires very, very much more memory. The cpu-time and memory requirements are only two aspects to consider when computational methods are compared. Robustness is another important feature.

**Implementation and comments:** A detailed description and discussion of the methods can be found in (1). We have used the programming language in the SAS system to implement the Slot-, ES- and Hybrid-methods. The programs are included in (1). We have tried to provide good standard values for all internal parameters.

The programs may be used as they are. The user only needs to provide the name of a SAS table with all the data values and select ES (default) or Hybrid. There are several internal

parameters that the user may modify to suit specific needs. Normally nothing needs to be changed.

The ES- and Hybrid-methods are rather insensitive to the choice of parameter values, e.g. the selected lower and upper boundaries and the slot width of the middle region. Both methods have worked and given correct results even when we by mistake have given completely erroneous parameter values, the cpu-time increased, that was all. The methods have proven to be very robust and reliable.

**Other quantile definitions:** The quantile point may also be defined as the mean value of the data point immediately before and immediately after the quantile index. The computation of quantile points using the ES- or Hybrid-method is easily extended to this and other definitions.

**Computation of other statistics:** We can compute the mean and standard deviation and other statistical values with very good accuracy using the information in the slot array, the slot midpoint and the number of values in each slot.

**A note on cpu-time:** In our test executions we have assumed that all the data points are randomly created and stored in internal format in a SAS table. If we instead need to read them from an ordinary external flat file, the cpu-time will be 3 to 4 times longer.

**Conclusion:** The Slot-, ES- and Hybrid-methods are all very robust and fast methods to compute all required quantile points to a requested accuracy. They require very little internal memory and very little disc space. They can easily be modified according to user demands.

**Anders and Gunilla Skölleremo**

(1) "Fast and Exact Computation of All Quantiles for Very Large Amounts of Data", See [www.ResearchGate.net](http://www.ResearchGate.net).  
By Anders Skölleremo

## Stop Press

### Please come to the Launch of 3 x Altair SLC books in Hamburg on 20Aug2026

Philip R Holland and Pablo Orosco Joerger will be at The Locks, Marienhof 6, 23399 Hamburg, a bar and restaurant in northern Hamburg, between 1200 and 1600 on 20Aug2026 to talk to interested readers about their 3 new books:

- Advanced Altair SLC
- Altair SLC for Business Users
- Altair SLC for Companies

Discounted ebooks will be available to purchase at the venue. Discounted paperbacks can also be requested from the authors.

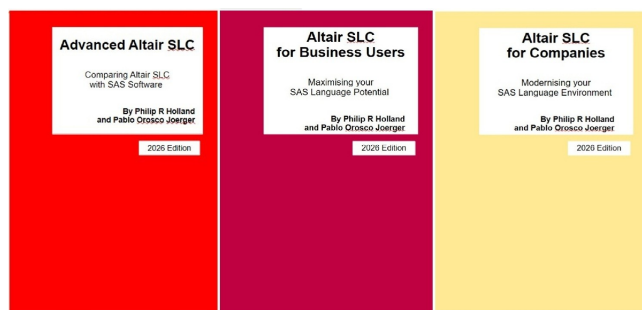
**Philip R Holland and Pablo Orosco Joerger**

## Book Releases

### 3 new Altair SLC books available in ebook and paperback format from July 2026

Each volume is self-contained but shares the common foundation chapters. The choice of title depends on the reader's role and what they need from Altair SLC.

<https://hollandnumerics.org.uk/forums/forum/views-uk>



| Volume                               | Primary audience                                                | Core value proposition                                                                                                                                                          |
|--------------------------------------|-----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Advanced Altair SLC</b>           | Experienced SAS/SLC programmers                                 | Technical depth: PROC R/PROC IML/PROC PYTHON with full programs, ODS gaps, clinical CDISC programming, non-standard file reading/writing, systematic SLC vs SAS 9.4 differences |
| <b>Altair SLC for Business Users</b> | SAS programmers and analysts adopting SLC as their primary tool | Broadest practical coverage: batch processing, server access, database access, R/Python, SASjs integration, clinical programming - all with working code                        |
| <b>Altair SLC for Companies</b>      | IT managers, architects, decision-makers evaluating migration   | Business case: cost and capability comparison with SAS 9.4 and Viya, migration planning checklist, database engine coverage, server architecture                                |

All 3 books are available on [hollandnumerics.org.uk](https://hollandnumerics.org.uk) and Lulu.com as ebooks and paperbacks:

- <https://hollandnumerics.org.uk/book-table>
- <https://www.lulu.com/spotlight/hollandnumerics>
- <https://www.lulu.com/search?sortBy=RELEVANCE&page=1&q=Altair+SLC>

**Philip R Holland and Pablo Orosco Joerger**

## Formats, Options, and Functions

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

[SAS and Altair SLC] The numeric format **YENw.d** displays values in a similar way to **DOLLARw.d**, but with a prefix of “¥” instead of “\$”.

[SAS and Altair SLC] The **SGANNO=** option available on the ODS Graphics procedures **PROC SGPLOT**, **SGPANEL** and **SGSCATTER** can be used to annotate the standard ODS Graphics. The SG Annotate statements are similar, but not identical to the Annotate statements in SAS/GRAPH.

[SAS and Altair SLC] The **DIMn()** and **LAGn()** functions are equivalent to **DIM(n,)** and **LAG(n,)** functions, where **n** is an integer. However, the Altair SLC Program Compatibility tool may flag them incorrectly as unrecognized, but Altair SLC will still recognize and run them correctly.

## In Brief

- Back issues of *VIEWS News* are available from the VIEWS UK forum at [blog.hollandnumerics.org.uk](https://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 **41,300** members. Everyone is welcome to request to join, but discussions here are restricted to SAS-related topics: [www.linkedin.com/groups?gid=70702](https://www.linkedin.com/groups?gid=70702).
- The SAS Opportunities group on LinkedIn, which currently has over **23,400** members. Everyone is welcome to request to join, but discussions here are also restricted to SAS-related topics: [www.linkedin.com/groups?gid=805217](https://www.linkedin.com/groups?gid=805217).
- Altair SLC Personal Edition provides named users with free access to Altair SLC and Analytics Workbench software (formerly WPS) including a SAS language compiler with coding and workflow tools on unlimited data volumes, with interfaces to a wide range of databases, as well as integrated R and Python programming environments. When the Personal Edition software is downloaded, it includes a licence for up to 12 months. At the end of this term, users should download a new copy of the software to renew their licence and/or to upgrade to the latest available full version. Download the software from [altairone.com/Marketplace](https://altairone.com/Marketplace). No technical support service is offered with the Personal Edition, but assistance can be requested using the Community forums at [community.altair.com](https://community.altair.com). The current General Access version of Altair SLC is 5.26.0 (2026).
- Do you use Microsoft Visual Studio Code Editor?
  - You may be interested in a SAS-related extension called **SASjs - DevOps for SAS**. You can develop SAS applications and run code on local SAS installations, SAS Viya, SAS 9 EBI, SASjs Server or local Altair SLC! Just set up the target, and click the running man to execute your SAS code. The SASjs Server could be your local SAS or Altair SLC installation. SASjs is an open source project, and can be obtained from the Visual Studio Code Marketplace, where the current version is v2.6.0.
  - The SAS Institute has also been developing its own extension called **SAS**. You can develop SAS applications and run code on local SAS installations, SAS Viya, or SAS OnDemand for Academics! Just set up the target, and click the running man to execute your SAS code. The extension can be obtained from the Visual Studio Code Marketplace, where the current version is v1.19.0.
  - Another Microsoft Visual Studio Code Editor extension has just been released called **Altair SLC**. This extension currently only allows you to connect to local Altair SLC installations, even if, until now, you have only been able to use Altair Analytics Workbench to access Altair SLC. The current version in the Visual Studio Code Marketplace is v0.7.1 (pre-release).

<https://hollandnumerics.org.uk/forums/forum/views-uk>

## 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.*

### August 2026

20 August **Altair SLC Book Launch**, Hamburg, Germany

Meet the authors of 3 Altair SLC books in Hamburg, where they will celebrate a successful release of paperbacks and ebooks. Full details in **Stop Press** above.

### November 2026

15-17 **MWSUG**, Madison, WI, USA

The conference committee is busy planning another excellent multi-day event filled with user presentations, networking opportunities, training, demonstrations, and more! Please stay tuned to **[www.mwsug.org](http://www.mwsug.org)** and MWSUG social media channels for more details.

15-18 **PhUSE EU Connect**, Glasgow, UK

The PhUSE Community allows you to enhance your learning and gain insights from our experts on today's most critical industry topics through a variety of engaging presentations, workshops, and networking opportunities. Individuals from all areas of the industry will be gathered in one place, providing an opportunity not only for you to advance your career but also to heighten the pharmaceutical industry's influence and findings. Full details at **[www.phuse.eu](http://www.phuse.eu)**.

## Contacts

**LeRoy Bessler**

Email: [besslerleroy2024@gmail.com](mailto:besslerleroy2024@gmail.com)

**Pablo Orosco Joerger**

Email: [pablo.joerger@icloud.com](mailto:pablo.joerger@icloud.com)

**Anders Skollermo**

Email: [anders.skollermo@one.se](mailto:anders.skollermo@one.se)

### VIEWS

Web Site

Web: [hollandnumerics.org.uk](http://hollandnumerics.org.uk)

Newsletter Editor

Email:  
[views-uk@hollandnumerics.org.uk](mailto:views-uk@hollandnumerics.org.uk)

**VIEWS News** is published quarterly for the **VIEWS International SAS Programmer Community** by Philip R Holland.