Paper 262-2008

An Interpretation of Sparkline Graphs and Their Creation with SAS® Software

Paul OldenKamp, POK Programming Co., Seattle, WA

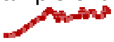
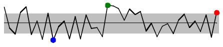
ABSTRACT

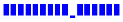
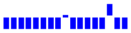
Edward Tufte has defined 'sparkline' as a "small, high resolution graphics embedded in a context of words, numbers, and images"[1]. They allow for very high density graphics information to be integrated with the text that discusses the information being presented and have been implemented in SAS infrequently.

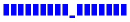
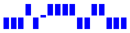
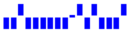
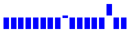
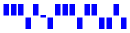
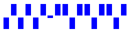
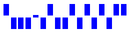
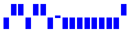
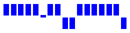
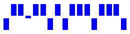
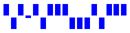
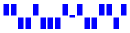
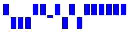
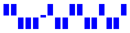
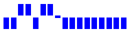
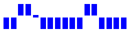
This paper discusses a variety of sparkline styles and how sparklines can provide more information in a given space when designed with good graphical presentation principles. The basic three styles to be discussed in detail are: 1) the line graph of continuous time series data; 2) a win/loss/tie sparkline for binary results; and 3) a table of column sparklines. The interpretation of this later column table will be used to demonstrate how high dimensional graphical presentations enhance information content.

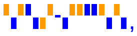
As with most SAS Software programming there are several ways to accomplish the same task. Sparklines have been produced using the SAS/Graph Annotate facility, PROC GCHART, and PROC GPLOT. These methods will be described and contrasted with the use of the SAS® Data Step Graphics Interface method presented in this paper and the use of True Type fonts and the Javaobj SAS interface to Java programs to generate sparklines with Base SAS programming.

EXAMPLES OF SPARKLINE STYLES

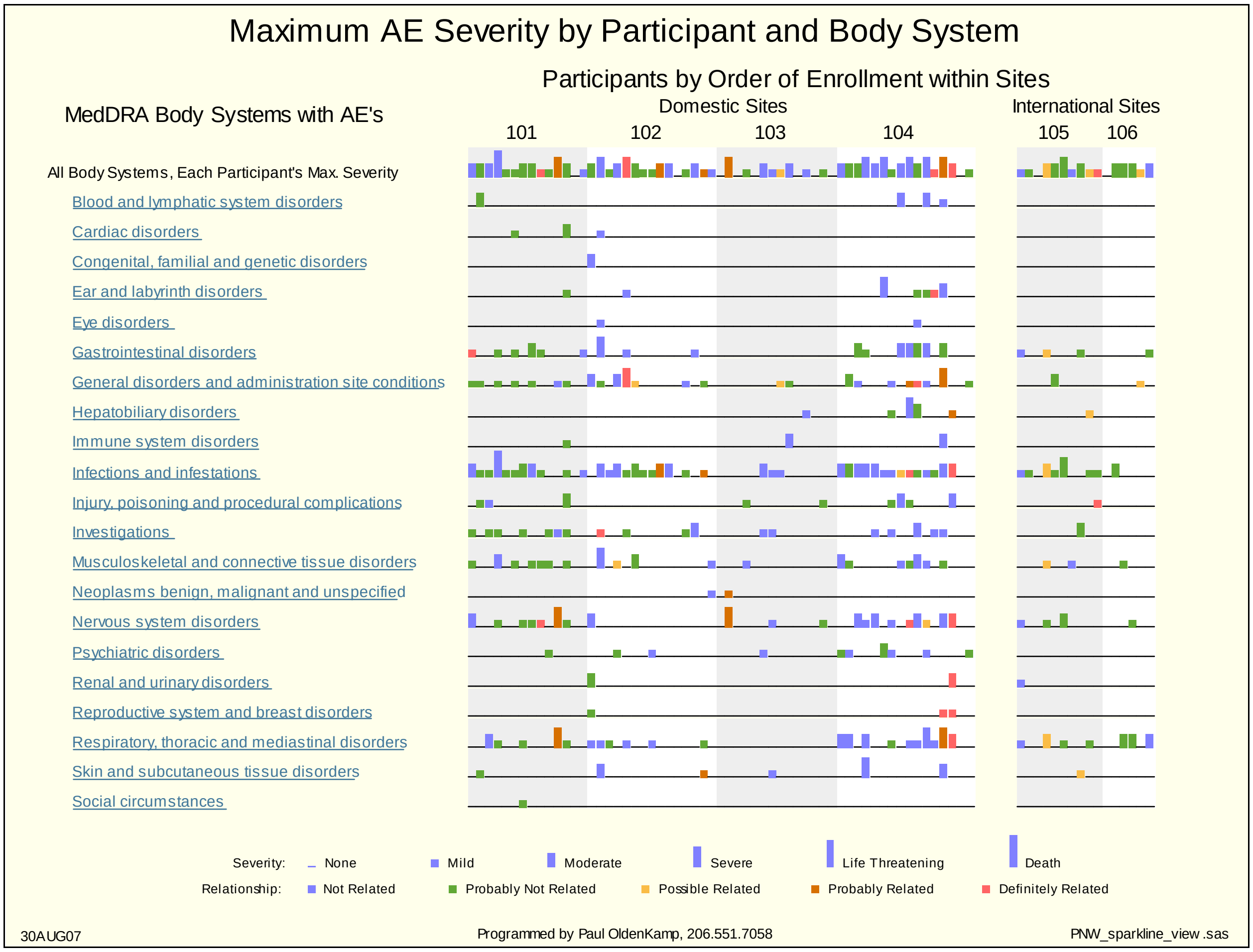
Tufte begins his discussion of sparklines with the line style. An example of this style generated for this presentation shows the attendance at SUGI and last year's SGF conferences since 1976, . One can see that SUGI attendance steadily increased to a peek in 1989, when the conference was held in San Francisco. The all time high reached in 2006 was also in San Francisco. The high resolution of sparklines is demonstrated with over thirty distinct attendance numbers shown in a space of less than three quarters of an inch. The time series pattern of attendance is clearly discernable. A number of enhancements adding reference information to line style sparklines,  temperature 7.9[2], are recommended by Tufte to enhance their usefulness. The combination of graphics and text demonstrated in this paragraph is an example of Tufte's objective for the use of sparklines, i.e., the close proximity of graphics and the text discussion.

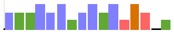
Our second example of a sparkline style is the win/loss/tie style for binary data. Many examples of this style use sports data and the table shown to the right summarizes the NFLAmerican Football Conference for the 2007 season. Here the table structure facilitates the comparison between teams and divisions and we can reproduce individual sparklines, like the New England Patriot's perfect regular season,  or Miami's disastrous year,  , in the text discussion so the reader need not constantly refer back and forth to the table.

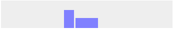
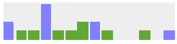
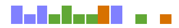
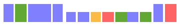
NFL 2007 Season			*W*	*L*	*T*
AFC East	New England		16	0	0
	Buffalo		7	9	0
	N.Y. Jets		4	12	0
	Miami		1	15	0
AFC North	Pittsburgh		10	6	0
	Cleveland		10	6	0
	Cincinnati		7	9	0
	Baltimore		5	11	0
AFC South	Indianapolis		13	3	0
	Jacksonville		11	5	0
	Tennessee		10	6	0
	Houston		8	8	0
AFC West	San Diego		11	5	0
	Denver		7	9	0
	Kansas City		4	12	0
	Oakland		4	12	0

The single color win/loss/tie sparklines emphasizes the strings of wins and losses the teams experienced during the season and whether the strings occur at the beginning or end of the season. If one adds another color when another attribute is present, say make the away games orange, the Seattle Seahawks season sparkline, , one can see the relationship between the win/loss dimension and the home/away dimension of the data. Strangely, Seattle won more away games than they lost and lost more home games than they won.

The column style of 'sparkline like' elements in the adverse event (AE) Summary graph on the following page summarizes the AE's experienced in an entire clinical trial and combines information on over nine dimensions of the data on a single page. High dimensionality is a very helpful feature for making comparisons with the data.
Combining the information on the severity of an AE in the study with information on how related the event is to the study product assists in understanding the importance of the event. In my experience, severity and relatedness counts have been reported in separate reports and analysts are continually looking up the detail record of an AE to learn whether a severe AE is related to the study product. By color coding the AE's the analyst is immediately aware to the relatedness and can investigate only the related events of high severity.



Additional information can be provided by controlling the order and location of graphic elements in sparklines. Since the left to right order of the top row, where there is one bar or blank space for each participant, is in sequence of enrollment in the study, one can identify where more recent participants have events more closely related to the study treatment than earlier participants,  at site 104. The separation of the international sites from domestic sites makes their comparison easier.

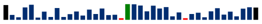
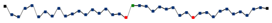
One can compare the small number of 'Infections and Iinfestations' at site 103, , with the larger number of participants with this category of events at site 101, , site 102, , and site 104,  and ask why the experience at one site is so different. Is it likely that participants could be that different at site 103 or is there perhaps is a different practice in coding adverse events at site 103?

There are nine dimensions of data displayed in this graph. The height of the bars shows the severity (1) of the AE. The color of the bar displays the relatedness (2) of the AE to the study treatment. Severity and life threatening events that are probably and definitely related are of greatest concern. The number of bars or blank spaces across the top of the graph shows the total number of participants currently enrolled (3). One can easily count in that row the number of participants with a maximum grade of each severity (4) and the total number of participants for whom this maximum event is of each relatedness (5). The number of enrollees at each site (6) is shown and the sites with the most events of high severity (7). The count of domestic vs. international participants (8) is shown and the list of MedDRA body system categories (9) is displayed on the left side of the table.

'Drilling down' form a summary display like this to a display of the underlying detail categories or a complete patient profile of a participant with an unusual symptom is a common technique to get even higher resolution information than can be displayed on one page. The MedDRA body system text is underlined and displayed in blue to clue the reader to the presence of a link to the detail reports.

SOME COMMERCIAL SPARKLINE PRODUCTS

Bissantz markets an add-in for Microsoft Office that generates sparklines in Word and Excel documents called Sparkmaker. It uses the True Type fonts that are used to produce the sparklines in the NFL table on page one of this paper. SparkMaker is available as a 30 day free trial.

Here are some examples of sparklines produced with the Bissantz SparkMaker: SparkMaker can produce a bar sparkline, 9.3  7.8 [1|9.9], and a line style sparkline, 9.3  7.8 [1|9.9] There are well designed and can be used as benchmarks for sparklines produced with SAS programs. Notice that the values of the first and last values in the graph are also reported as numeric reference points and the maximum and minimum values are color coded to aid in the interpretation.

An innovative implementation of sparklines is as an enhancement to a stock price ticker scrolls across the Bissantz web page. It's produced with their beta product SparkTicker.

Bonavista Systems has the Microcharts software that produces more very well designed styles of sparklines than other commercial products. Microcharts is also available for a 30 day free trial.

These products share a limitation that has frustrated Tufte: most methods for producing sparklines are manually intensive. After preparing the data one wants to display in sparklines one has to manually transfer it to the SparkMaker application. Hopefully, this will be a temporary difficulty as better applications are developed

The Business Refinery has a Sparkllines for Java product for serious developers of Java applications.

WHAT ARE SOME CHARACTERISTICS OF SPARKLINE GRAPHS?

Several useful characteristics of sparklines have been identified by Tufte, in the online discussions of sparkline design and by observation of the sparklines produced in various contexts.

ASPECT RATIO IN LINE SPARKLINES

William Cleveland advised that time series line graphs should be designed so that the lines have a slope of about 45° for the reader to get an objective understanding of the relative fluctuation. A sparkline should have a vertical dimension that uses the available space and then the length should be adjusted to get the average 45° slope. Kate Davis[4] provided a formula and programs for SAS and JMP to compute sparkline dimensions that satisfy the 45 degree rule in her NESUG 2007 paper.

HIGH DENSITY/RESOLUTION INFORMATION

One goal to pursue on a high quality graphics publication is the high density of information. This doesn't mean that there should be a lot of pixels on the page, but that there should be a very large volume of information displayed. The goal is a high information/ink ratio. One should avoid 'chart junk' used to simply decorate the page without conveying information. The guidance here should be for a minimalist approach for graphic elements

The human eye is capable of discerning differences in graphical displays with very small details. Taking advantage of this ability by displaying the highest resolution possible with a given media is important in achieving high density. The designer of statistical graphs should take advantage of this to combine many small related graphs on the same page so it is easy for the reader to compare different series of variables in the data and perform an analysis.

MULTI-DIMENSIONAL DISPLAY

Combining several dimensions of data on the same page aids the analysis. The famous graph by Minard of Napoleon's Moscow campaign described by Tufte in his book *Beautiful Evidence* displays six different variables of information on the page in ways that facilitate comparison.

As described above, sparklines can use color and position to create dimensional comparisons for analysis.

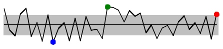
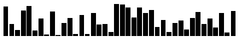
SUMMARY-DETAIL LINKS

Another element is to display different levels of granularity in the information. In the AE Summary graph, hyperlinks are implied with the blue underlined text to connect this one page 'detailed' summary with separate reports with subcategory charts

HOW HAVE SPARKLINES BEEN PRODUCED WITH SAS SOFTWARE?

There are three SAS programs published on the web to produce sparklines. Marc Hendel posted his program using the SAS/Graph PROCs and the SAS Annotate facility to produce bar and line sparklines in November, 2005. Last year the SAS Institute added sample number 1700 to its sample library. That SAS program produces a "Web Marketing Dashboard" that was first published in Stephen Few's book *Information Dashboard Design,* [3] and produces a group of simple line style sparklines. The third program is a SAS macro written by Andrew Ratcliffe in the October 2006 Edition of his online newsletter *Note:* and produces another line style sparkline. URL's to these examples are listed in the References at the end of this paper. These examples of sparklines raise a number of issues regarding their practical construction and design.

HENDEL: SAS/GRAPH PROCS WITH ANNOTATE FEATURES

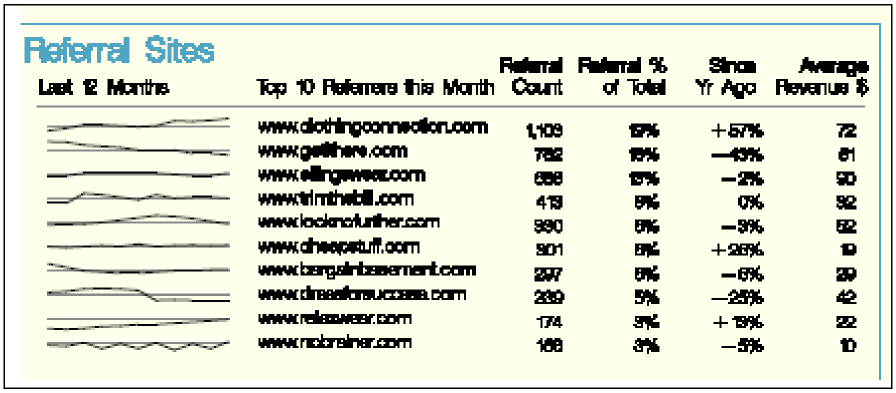
Marc Hendel's macro program produces a line sparkline,  temperature 7.9, with a normal range in gray and a bar sparkline,  temperature 7.9. For comparison, these sparklines use the same series of data as the sparklines shown in the middle of the previous page. Hendel's line sparkline program is the only one of the three examples that provides the gray background of the inner quartile range, the color coding of the minimum, maximum, and last value to provide a reference for the reader. The text label and red numeric value of the last value are additional features described by Tufte and are frequently included with sparklines when the context makes them useful. Hendel's program produces sparklines that are quite true to the design proposed by Tufte.

Comments of Hendel's web page explain the practical difficulty of outputting sparklines in a form that can easily be incorporated in a document. Tufte also discusses the need for an elaborate combination of software programs in order to produce high quality sparklines.

Hendel claims to be a novice SAS/Graph programmer and has invited comments for improving his programs. The programs are fine for a proof of concept experiment but could be more parameter driven and separated into several modules for regular production work. Like many of these programs, a very specific form of input is necessary and more robust data input would be helpful for general use.

SAS INSTITUTE: PROC GPLOT IN A DASHBOARD DISPLAY

In 2007, the SAS Institute announced some new sample programs that reproduce the executive dashboard examples published in Stephen Few's book *Information Dashboard Design*. I had some difficulty coping the section of the dashboard with sparklines and the poorer quality is a result. Here's a section of the page that contains the sparklines:



Referral Sites

Last 12 Months	Top 10 Referrers this Month	Referral Count	Referral % of Total	Since Yr Ago	Average Revenue $
	www.clothingconnection.com	1,103	17%	+57%	72
	www.getthere.com	792	15%	—43%	61
	www.ellingwear.com	658	12%	−2%	90
	www.trimthebill.com	419	8%	0%	92
	www.looknofurther.com	390	8%	−3%	62
	www.cheapstuff.com	301	6%	+26%	19
	www.bargainbasement.com	297	6%	−6%	29
	www.dressforsuccess.com	280	5%	—25%	42
	www.relaxwear.com	174	3%	+19%	22
	www.rickrelner.com	168	3%	−5%	10

The sample program has this interesting comment:

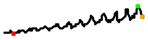
```
/* These are the sparklines to the left of the bar chart. "Sparkline"
   graphs are challenging to create with SAS/GRAPH software.  This example
   uses the GPLOT procedure, and for each line an offset is
   added so that the lines are plotted above or below each other, rather
   than on top of each other. */
```

We need a way that isn't so 'challenging'.

The objective of the SAS Institute programming was to reproduce the Few dashboard as closely as possible and the Institute programmers can not be responsible for issues in the design of the sparklines included. Because of the need to clearly display information is such a small area close care should be given to graphic design principles like the aspect ratio of the graph. The sparklines shown above are too flat to present a good display of the variability of the data and should follow the Cleveland rule of 45° line slopes for better understanding. The sparklines also do not have a quantifying reference to aid in interpretation since the different sparklines seem to have a different vertical scale.

The quality of the font could not be improved easily since a font other that the SAS software font SWISSB changed the relative position of graph elements.

ANDREW RATCLIFFE: MACRO USING PROC GPLOT

In the October 2006 issue of his online newsletter, *Note*, Andrew Ratcliffe commented on the increasing interest people were having in sparklines. He included a SAS macro program to produce a simple sparkline with a color coded minimum, maximum and last point,  This sparkline style is particularly weak in it's reference context.

The bulk of the SAS macro program used to produce this sparkline is for data manipulation and to calculate the color coded points and then a simple SAS/Graph program uses the gif driver and PROC GPLOT.

THIS PROGRAM'S APPROACH

The objective of the programming to produce the graph on page one is to demonstrate the high resolution of information possible using a variation of a sparkline design and other graphical color and location attributes to display many dimensions of data on the same page for easy comparison. I didn't see how to use SAS/Graph PROCs to display this large amount of information on one page and turned to the Data Step Graphics Interface. The DSGI uses the graphic kernel system and allows one great control to put graphic elements wherever one wants on the page, but at the cost of a much more detailed low level programming.

A secondary objective was to learn about how to integrate SAS/Graph programming with the Output Delivery System and get my SAS/Graph programming skills up-to-date.

The input data for this program comes from AE, Adverse Event domain, and DM, Demographics domain, datasets that follow the CDISC, Clinical Data Information Standards Consortium, SDTM, Study Data Tabulation Model, standard. I hoped that this would be a useful starting point.

The Model-View paradigm was followed to create the fundamental logic of the application after an initialization phase to establish the general SAS programming and SAS/Graph environments, and input/output locations. The model SAS macro program module the sets up the data for the view SAS macro program module that generates the page of graphics.

The programs are available on the PNWSUG conference web site with my paper presented at the PNWSUG 2007.

THE CONCEPTUAL AND SAS CODE DETAILS

THE MODEL PROGRAM MODULE

The model program combines data from the two CDISC domains, calculates a few additional categorical variables, and then uses executions of PROC MEANS to select the maximum severity and relatedness from cases of multiple observations for an individual participant. The COMPLETETYPES and PRELOADFMT options were used to fill out the datasets so they have an observation for each participant for each MedDRA symptom that is in the data. This sets the data up so one can then loop through the data producing each column of participant bars for each row of the graph.

THE VIEW PROGRAM MODULE

The overall design of the DSGI program used here is somewhat crude. A DSGI graph is produced in a single SAS data step that can get long and complex. This program is definitely a 0.5 prototype version that should be redesigned for more flexible handling of more columns horizontally or more rows vertically.

There are a variety of initialization steps that are performed when the retained variable **initialization** has its initial value of TRUE. The RGB color specifications are assigned to six color index values for use in other DSGI statements and the initial value of the text font size is set for example. While the first row is being produced the column headings positions are calculated and the text is placed on the page. The **initialization** variable is set to FALSE after the first line in the main table is written.

The data step loops through the individual participant columns and site and location groups of columns for each block of observations for each row.

When the program gets to the end of the data after writing the last line of the table the legend is written at the bottom of the page with do loops to calculate the position of text element and to write the symbols and text.

GRAPHICS PROGRAMMING ISSUES

I learned of a few programming issues in the interaction of the DSGI and ODS output that are unresolved.

The first issue is the DSGI calculation of the length of text strings is not accurate for several ODS output locations and is different for output in the SAS graphics output window in the Display Manager. One can see the results of this on the graph where the lines drawn under the text column at the left side of the table are too short for long text strings and too long for short text strings. This effect depends on the font used and may not occur for SAS software fonts. This creates a real trade off for the programmer since the SAS software fonts produce poor quality in several ODS output locations.

A second issue is that my ODS pdf output required specifying a value of .25 for the SAS/Graph options horigin and vorigin to get a .5 inch margin.

The biggest issue for me personally though is the difficulty of producing automated unit test checks for this type of detailed graphics programming. How can one detect in a program whether a change made to the program produced an error or correct results?

PROGRAMMING OF SPARKLINES IN BASE SAS SOFTWARE

Several new possibilities are available in Version 9 of SAS Software for the production of sparklines without SAS/Graph software.

PROC REPORT WITH SPARKFONTS

The Bissantz SparkFonts were used in conjunction with a ODS PROC REPORT program to produce the NFL report on page one. The real work was in preparing the data for a fairly simple PROC REPORT program. The only novel aspect was to produce a string of +, -, and / characters, i. e. +---/-+--+-+-+-++, and then formatting it with the SparkFont in an inline style statement.

JAVA GRAPHICS LIBRARIES AND APPLICATIONS

Sparklines have been implemented in Java programs that are available at Larry Ogrodnek's web site at http://www.representqueens.com/spark/ . The javaobj functionality of the Version 9 data step allows SAS programmers to integrate the use of these programs with the data access and preparation capabilities of SAS software. Larry's Java programs are a useful starting point for developing production sparkline applications but lack most of the enhancements and design features discussed above.

SPECULATING ON FUTURE SAS PROGRAMMING OF SPARKLINES

There seems to be several new possibilities that new features in version 9 of SAS Software have opened up for the production of sparklines that I think should be explored in the future.

SAS ADD-IN FOR MICROSOFT OFFICE

Sparklines would be a natural addition to the SAS Add-in for Microsoft Office. The Bissantz SparkMaker works well for this but is limited because it requires one to manually type in the data values. One should be able to interface directly to a variety of data sources through the SAS system. I don't currently have access to this SAS product and don't know if a user extends this product or whether the SAS Institute would have to do it.

ENTERPRISE GUIDE STORED PROCESS

One can program a stored process in the SAS BI environment to provide a smooth interface the SAS system for producing sparklines. Who will be first to produce this stored process?

VERSION 9.2 OF SAS SOFTWARE NEW FUNCTIONALITY

What surprises will we find in the next major release of SAS software in the coming year? I haven't heard anything about PROC SPARKLINE so let's look for elements one can use for new and innovative graphics.

CONCLUSION

Sparklines are an exciting new format of graphical information display that improve interpretation by integrating large amounts of quantitative information with the text that discusses the numbers.

Traditional SAS/Graph programming can produce high quality sparklines but we need additional examples and programming techniques to get these good results into modern ODS output formats.

REFERENCES

[1] Tufte, Edward. 2006. *Beautiful Evidence.* Cheshire, CO: Graphics Press LLC

[2]Hendel, Marc, 2005. "Sparklines in SAS". http://home.mchsi.com/~marchendel/sparklines.html

[3] Few, Stephen. 2006. *Information Dashboard Design.* Sebastopol, CA: O'Reilly Media, Inc.

[4] Davis, Kate. 2007. "Sparklines for SAS® and JMP®", *Proceedings of the 2007 NESUG Annual Conference.*

[5] Ratcliffe, Andrew, 2006. "Note: Sparklines". http://www.ratcliffe.co.uk/note_colon/note18.htm#Sparklines

[6] SAS Institute Inc. 2007. Sample Program. " *Sample 1700:* Create a sample marketing analysis dashboard". http://support.sas.com/ctx/samples/index.jsp?sid=1700&tab=details

RECOMMENDED READING

Few, Stephen. 2006. *Information Dashboard Design.* Sebastopol, CA: O'Reilly Media, Inc.

OldenKamp, Paul. 2007 "A Variety of SAS® Programming Techniques to Produce Sparkline Graphics", *Proceedings of the 2007 Pacific Northwest SAS Users Group.*

Tufte, Edward, 2004. Ask E.T. Discussion site "Sparklines: theory and practice". http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR&topic_id=1&topic=

CONTACT INFORMATION

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

Paul OldenKamp

POK Programming Co.
8321 29th Ave. NW
Seattle, WA 98117

Work Phone: 206.551.7058
Fax/Alt Ph.: 206.789.7769

E-mail: POK_Programming@OldenKamp.org