

Unnoticed individual and mean line plot issue while using the SGPLOT procedure and the solution to avoid

Jagadish Katam, Princeps Technologies Inc.

ABSTRACT

This paper talks about the issue with overlaying time series for individuals and mean values in a single graph using PROC SGPLOT and the solution to it. Using the SGPLOT procedure to develop the overlaying time series plot, the individual as well as mean values lines in graph will not terminate to the end of X-axis. So, they will appear as unending lines circling back. This paper will have the images of the plots through which the issue will be detailed. In addition to this, I will talk about the changes that a dataset needs before passing into SGPLOT procedure and other required steps. This is an unnoticed issue in the SGPLOT procedure and the solution to it will help the peers to pay more attention while working on the individual/mean line plots.

INTRODUCTION

Generating simple graphs using SAS/GRAPH® can be a little challenging. When using the SGPLOT especially for line plot we miss to notice the error in the generated graph. If the graph is a simple single line plot or a mean line plot alone then there won't be any issue, however if you are using the SGPLOT for generating a multiple line plot with or without the mean then we need to pay close attention. Since the graph generated will not be in the expected format if we observe the lines plotted, we notice that the multiple lines appear as unending lines circling back which is incorrect. Sometimes we miss to notice it if there are many multiple lines in the same plot. In this paper we will investigate the issue and find a solution on how to avoid the multiple lines circling back.

CREATING INDIVIDUAL LINE PLOT

The example below creates a simple line plot using PROC SGPLOT. The following datalines create an example dataset of pharmacokinetic (PK) data, which includes an example subject 100. The other three variables used in this example are defined as VISIT (in weeks), VISITNUM (numeric version of VISIT) and AVAL (plasma concentration). This is a single ARM example, which means a single treatment.

In general, this graph is very effective to show the individual or mean concentration-time curves for a dataset with enough quantifiable concentrations over time.

```
data adpc;
input subjid visitnum visit&$ aval;
datalines;
100 20 Baseline 406000
100 30 Week 1 52300
100 40 Week 2 75200
100 50 Week 3 59200
100 60 Week 4 102000
100 70 Week 5 52200
100 80 Week 6 51000
100 90 Week 7 65600
100 110 Week 9 10500
100 130 Week 11 15700
100 140 Week 12 27200
100 150 Week 16 218
;
run;
```

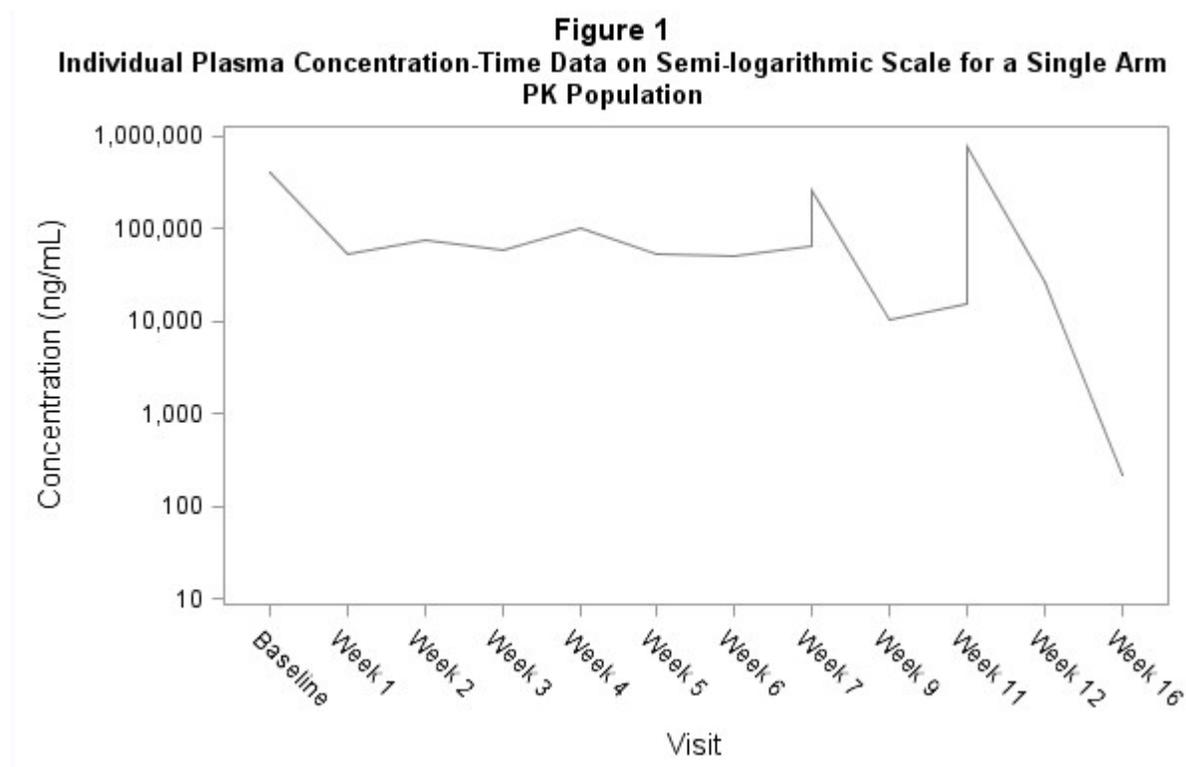
When we run the above datalines in SAS we get the below SAS dataset.

SUBJID	VISITNUM	VISIT	AVAL
100	20	Baseline	406000
100	30	Week 1	52300
100	40	Week 2	75200
100	50	Week 3	59200
100	60	Week 4	102000
100	70	Week 5	52200
100	80	Week 6	51000
100	90	Week 7	65600
100	110	Week 9	10500
100	130	Week 11	15700
100	140	Week 12	27200
100	150	Week 16	218

The following PROC SGPLOT creates a line plot of individual plasma concentration overtime for a single subject. Several options are used in the axis statements to control the labels, labels size and axis range in graph. In the series statement we plot the VISIT on X axis and AVAL on Y axis.

```
ods listing close;
title1 j=c color= black "Figure 1";
title2 j=c color= black "Individual Plasma Concentration-Time Data on
Semi-logarithmic Scale for a Single Arm";
title3 j=c color= black "PK Population";
ods graphics on / width=6.25in imagemap=on border=off;
PROC SGPLOT DATA = adpc;
yaxis label='Concentration (ng/mL)' labelattrs=(size=11pt)
values=(10 100 1000 10000 100000 1000000) logbase=10 type=log
logstyle=logexpand ;
xaxis label='Visit' labelattrs=(size=11pt) fitpolicy = rotate offsetmin=0.05
offsetmax= 0.05 TYPE = DISCRETE;
SERIES X = visitnum Y = aval / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
format visitnum vis. aval comma10.;
RUN;
ods _all_ close;
ods listing;
```

The output for the PROC SGPLOT is shown below. Notice the simple line plot without any issue. This is a perfect example of a simple line plot.

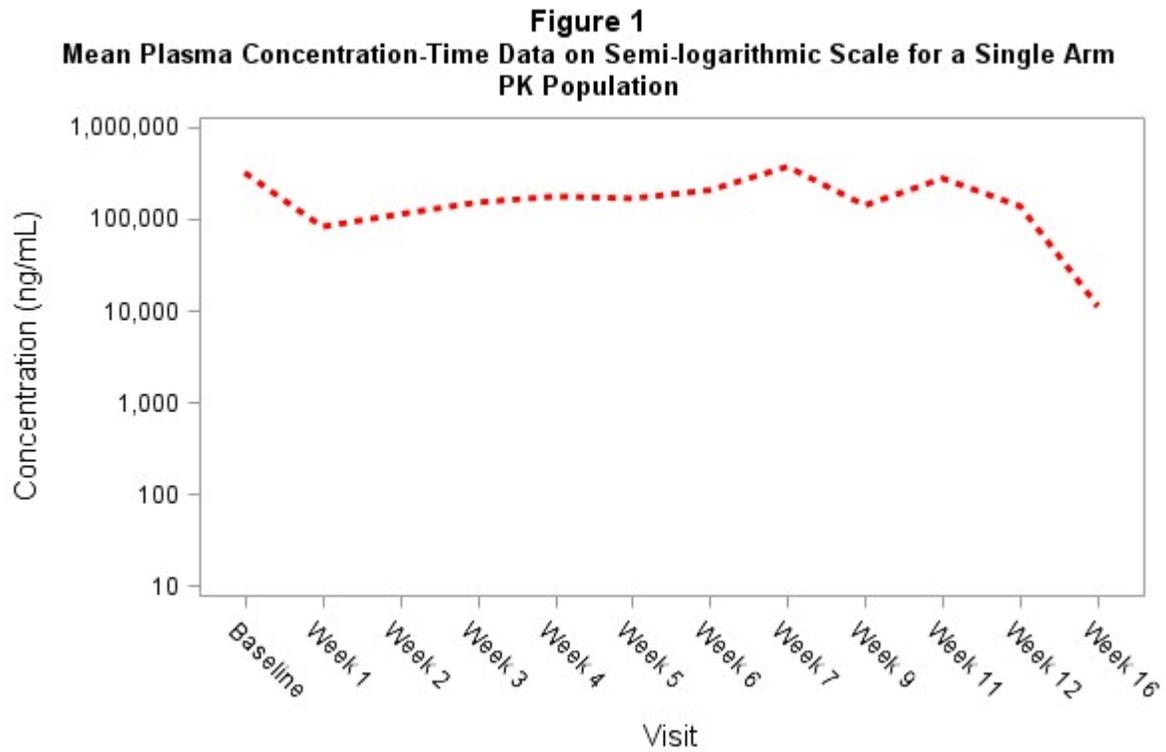


CREATING MEAN LINE PLOT

Similarly, the following PROC SGPLOT creates a line plot of mean plasma concentration overtime. The only difference between the above graph and the mean graph is that we replaced the Y axis with MEAN variable instead of AVAL.

```
ods listing close;
title1 j=c color= black "Figure 1";
title2 j=c color= black "Mean Plasma Concentration-Time Data on Semi-logarithmic
Scale for a Single Arm";
title3 j=c color= black "PK Population";
ods graphics on / width=6.25in imagemap=on border=off;
PROC SGPLOT DATA = adpc;
yaxis label='Concentration (ng/mL)' labelattrs=(size=11pt) values=(10 100 1000
10000 100000 1000000) logbase=10 type=log logstyle=logexpand ;
xaxis label='Visit' labelattrs=(size=11pt) fitpolicy = rotate offsetmin=0.05
offsetmax= 0.05 TYPE = DISCRETE;
SERIES X = visitnum Y = mean /name='mean' legendlabel='Average' LINEATTRS =
(pattern=2 THICKNESS = 3 color=red);
keylegend 'Mean' / location=outside position=bottom;
;
format visitnum vis. mean comma10.;
RUN;
ods _all_ close;
ods listing;
```

The output for the PROC SGPLOT is shown below. Notice the mean line plot also without any issue.



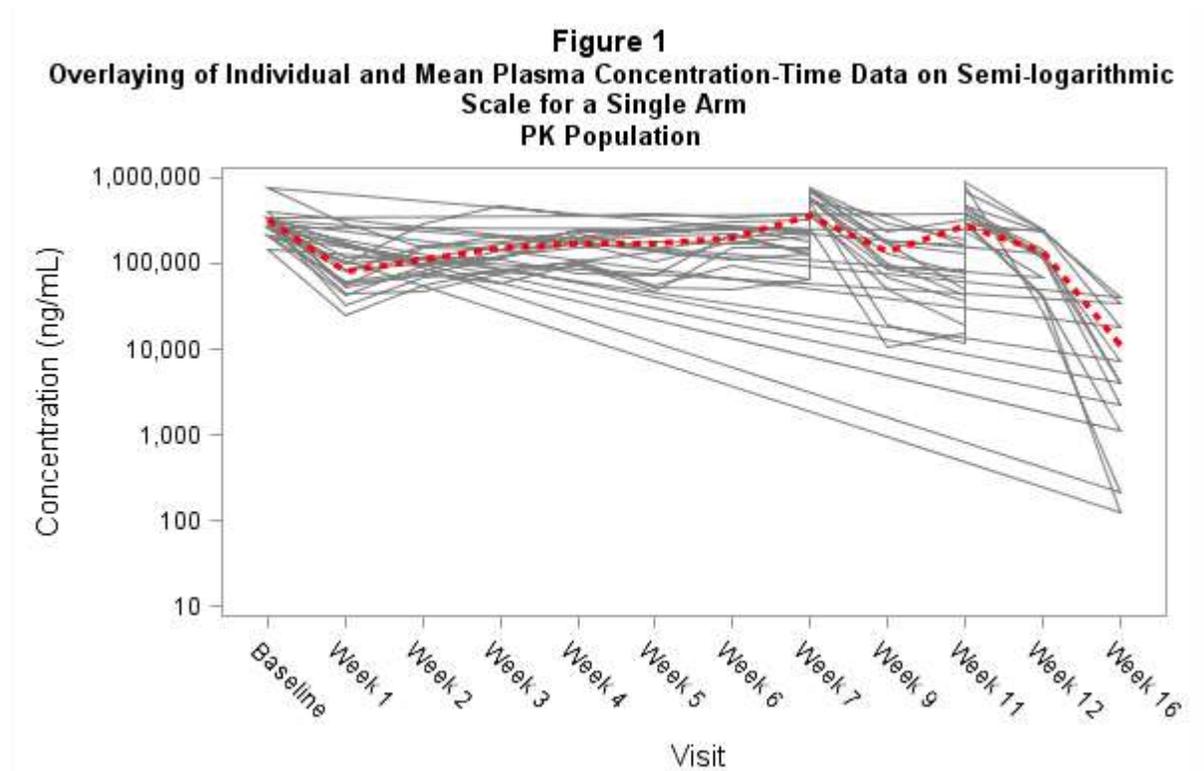
CREATING INDIVIDUAL AND MEAN LINE PLOT WITH UNENDING LINES

When we move from single subject or mean data to multiple subject's data and plot the individual and mean line plot, we can see the issue. Let's consider the below sample data as shown in the dataset below, due to space constraints here we showed only two subjects' data as example, but we considered several subjects data.

SUBJID	VISITNUM	VISIT	AVAL
100	20	Baseline	406000
100	30	Week 1	52300
100	40	Week 2	75200
100	50	Week 3	59200
100	60	Week 4	102000
100	70	Week 5	52200
100	80	Week 6	51000
100	90	Week 7	65600
100	110	Week 9	10500
100	130	Week 11	15700
100	140	Week 12	27200
100	150	Week 16	218
101	20	Baseline	342000
101	30	Week 1	84400
101	40	Week 2	97400
101	50	Week 3	118000
101	60	Week 4	175000
101	70	Week 5	206000
101	80	Week 6	224000
101	90	Week 7	132000
101	110	Week 9	101000
101	130	Week 11	43100
101	140	Week 12	114000
101	150	Week 16	4040

The following PROC SGPLOT creates an overlaying of individual and mean plasma concentration overtime. Here we used two series statements one for AVAL and another for Mean. However, for Mean values within the data there will be only distinct records per VISIT whereas for AVAL since there are many subjects there will be separate VISIT and AVAL values per subject. When we plot such a data in SGPLOT we can notice the issue in the generated output.

```
ods listing close;
title1 j=c color= black "Figure 1";
title2 j=c color= black "Overlaying of Individual and Mean Plasma Concentration-
Time Data on Semi-logarithmic Scale for a Single Arm";
title3 j=c color= black "PK Population";
ods graphics on / width=6.25in imagemap=on border=off;
PROC SGPLOT DATA = adpc;
yaxis label='Concentration (ng/mL)' labelattrs=(size=11pt) values=(10 100 1000
10000 100000 1000000) logbase=10 type=log logstyle=logexpand ;
xaxis label='Visit' labelattrs=(size=11pt) fitpolicy = rotate offsetmin=0.05
offsetmax= 0.05 TYPE = DISCRETE;
SERIES X = visitnum Y = aval / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = mean /name='mean' legendlabel='Average' LINEATTRS =
(pattern=2 THICKNESS = 3 color=red);
keylegend 'Mean' / location=outside position=bottom;
format visitnum vis. aval var: comma10. mean comma10.;
RUN;
ods _all_ close;
ods listing;
```



The output for the PROC SGPLOT is shown above. Notice the overlaying of individual and mean line plot with particularly the individual lines which represent plasma concentrations per subject per visit shows unending lines circling back. This output makes the interpretation difficult and confusing. We can also notice that the mean line is not circling back. Reason being that the mean values are derived in a separate variable MEAN. If the mean values are also derived in AVAL and a single series statement is used in the SGPLOT then it will make the mean values also circle back.

CREATING INDIVIDUAL AND MEAN LINE PLOT WITHOUT UNENDING LINES

As noticed in the above graph, we need to find a solution to avoid the circling back of the individual lines. The below code explains on how we can modify the dataset and pass in SGPLOT to avoid the unending individual lines.

```
/*subset the analysis dataset adsl on PK population*/
proc sort data=adam.adsl out=adsl(keep=usubjid) ;
by usubjid ;
where PKFL='Y' ;
run;

/*subset the analysis dataset adpc on analysis flag if any*/
proc sort data=adam.adpc out=adpc ;
by usubjid visitnum aval;
where anl01fl='Y' and aval>0;
run;

/*merge the adsl with adpc to get the same subjects as in PK population*/
data adpc;
merge adpc(in=a) adsl(in=b);
by usubjid;
if a and b;
run;

proc sort data=adpc;
by trt01pn trt01p visitnum visit;
run;

/*to get the mean line plot, derive the mean using PROC MEANS*/
proc means data=adpc noprint;
by trt01pn trt01p visitnum visit;
var aval;
output out=means mean=mean;
run;

proc sort data=adpc;
by usubjid visitnum visit;
run;

/*get the distinct subjects into a separate adpc2 dataset*/
data adpc2;
set adpc;
by usubjid;
if first.usubjid;
keep usubjid ;
run;
```

```

/*assign the incremental numbers by _n_ to subjects in ord variable*/
data adpc3;
set adpc2;
by usubjid;
ord=_n_;
run;

/*Merge the adpc with adpc3 dataset that has the ord variable*/
data adpc4;
merge adpc(in=a) adpc3(in=b);
by usubjid;
if a;
run;

/*create a macro variable subj which stores the total count of subjects*/
proc sql noprint;
select count(distinct usubjid) into : subj trimmed from adpc4;
quit;
%put &subj;

/*use the macro variable subj in the array and derive var variables with aval
values so that separate var variable is created for each subject*/
data adpc5;
set adpc4;
by usubjid;
array vars(*) var1-var&subj.;
do i = 1 to dim(vars);
if ord=i then vars(i)=aval;
end;
run;

```

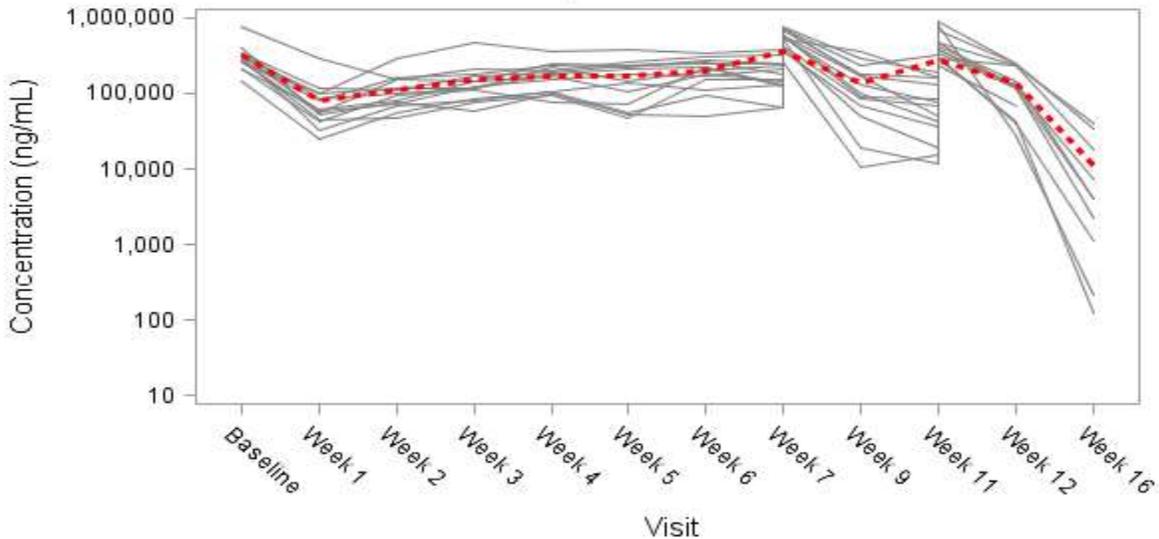
SUBJID	ORD	AVAL	VAR1	VAR2
100	1	406000	406000	.
100	1	52300	52300	.
100	1	75200	75200	.
100	1	59200	59200	.
100	1	102000	102000	.
100	1	52200	52200	.
100	1	51000	51000	.
100	1	65600	65600	.
100	1	255000	255000	.
100	1	10500	10500	.
100	1	15700	15700	.
100	1	784000	784000	.
100	1	27200	27200	.
100	1	218	218	.
101	2	342000	.	342000
101	2	84400	.	84400
101	2	97400	.	97400
101	2	118000	.	118000
101	2	175000	.	175000
101	2	206000	.	206000
101	2	224000	.	224000
101	2	132000	.	132000
101	2	566000	.	566000
101	2	101000	.	101000
101	2	43100	.	43100
101	2	464000	.	464000
101	2	114000	.	114000
101	2	4040	.	4040

We can see the above dataset with separate VAR variables with AVAL values for individual subjects. The number of the VAR variables will depend on the number of the subjects in the data. The following PROC SGPLOT uses the separate VAR variables instead of AVAL in PROC SGPLOT. The 12 VAR variable from VAR1-VAR12 here indicate that there are 12 subjects.

```
ods listing close;
title1 j=c color= black "Figure 1";
title2 j=c color= black "Overlaying of Individual and Mean Plasma Concentration-
Time Data on Semi-logarithmic Scale for a Single Arm";
title3 j=c color= black "PK Population";
ods graphics on / width=6.25in imagemap=on border=off;
PROC SGPLOT DATA = adpc5;
yaxis label='Concentration (ng/mL)' labelattrs=(size=11pt) values=(10 100 1000
10000 100000 1000000) logbase=10 type=log logstyle=logexpand ;
xaxis label='Visit' labelattrs=(size=11pt) fitpolicy = rotate offsetmin=0.05
offsetmax= 0.05 TYPE = DISCRETE;
SERIES X = visitnum Y = var1 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var2 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var3 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var4 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var5 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var6 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var7 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var8 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var9 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var10 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var11 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = var12 / name='aval' markers markerattrs=(size=2pt
symbol=squarefilled color=grey) LINEATTRS = (pattern=1 THICKNESS =0.5
color=grey);
SERIES X = visitnum Y = mean /name='mean' legendlabel='Average' LINEATTRS =
(pattern=2 THICKNESS = 3 color=red);
keylegend 'Mean' / location=outside position=bottom;
format visitnum vis. var: comma10. mean comma10.;
```

```
RUN;  
ods _all_ close;  
ods listing;
```

Figure 1
Overlaying of Individual and Mean Plasma Concentration-Time Data on Semi-logarithmic Scale for a Single Arm PK Population



The output for the PROC SGPLOT is shown above. Notice the overlaying of individual and mean plasma concentration is without the unending lines. Now the graph is interpretable.

CONCLUSION

This paper presents just one of the approaches that can be used to create overlaying of individual and mean plasma concentration plot. There could be other approaches as well to achieve the same output. Here I tried to take a simple and easy approach to avoid the unending lines in the plot when we plot several subjects in the same plot. Also, with this paper I brought to the attention an issue sometimes is unnoticed when we have huge data with several subjects in the same plot.

REFERENCES

Susan J. Slaughter and Lora D. Delwiche. 2010. "Using PROC SGPLOT for Quick High-Quality Graphs." SAS Global Forum 2010, Paper 154-2010.

Available at <https://support.sas.com/resources/papers/proceedings10/154-2010.pdf>.

ACKNOWLEDGMENTS

The author would like to thank Bharat Buchupalli and Vijay Krishna, for their insightful comments and review.

CONTACT INFORMATION

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

Jagadish Katam
Princeps Technologies Inc.
Jagadish.katam@princepstechnologies.com

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.