PharmaSUG 2021 – Paper DV-200

A Sassy substitute to represent the longitudinal data – The Lasagna Plot

Soujanya Konda, IQVIA, Germany

ABSTRACT

Data interpretation becomes complex when the data contains thousands of digits, pages, and variables. Generally, such data is represented in a graphical format. Graphs can be tedious to interpret because of voluminous data, which may include various parameters and/or subjects.

Trend analysis is most sought after, to maximize the benefits from a product and minimize the background research such as selection of subjects and so on. Additionally, dynamic representation and sorting of visual data will be used for exploratory data analysis.

The Lasagna plot makes the job easier and represents the data in a pleasant way. This paper explains the tips and tricks of the Lasagna plot.

INTRODUCTION

In longitudinal studies, representation of trends using parameters/variables in the fields like pharma, hospitals, companies, states and countries is a little messy. Usually, the spaghetti plot is used to present such trends as individual lines. Such data is hard to analyze because these lines can get tangled like noodles.

The other way to present data is HEATMAP instead of spaghetti plot. Swihart et al. (2010) proposed the name “Lasagna Plot” that helps to plot the longitudinal data in a more clear and meaningful way. This graph plots the data as horizontal layers, one on top of the other. Each layer represents a subject or parameter and each column represents a timepoint.

Lasagna Plot is useful when data is recorded for every individual subject or parameter at the same set of uniformly spaced time intervals, such as daily, monthly, or yearly.

MATERIALS & METHODS

Data of Injection Site Pain VAS Score and Injection Site Reactions such as Pain, Swelling, Tenderness etc. are the materials that are used for illustrative purpose. The data used in this presentation are dummy data created for demonstration.

INJECTION SITE PAIN VAS SCORE: Score is collected on a 0 to 100 visual analog scale, with 0 being the least pain score and 100 being the most severe pain score.

Pain score is collected at different timepoints after an injection. For example: “1-Min Post Injection, 5-Min Post Injection, 10-Min Post Injection, 15-Min Post Injection, 30-Min post Injection, 60-Min Post Injection, and 120-Min Post Injection”.

SAMPLE DATA

SUBJID	TRTA	AVISITN	ATPT	ATPTN	PARAM	AVAL
136-0099	Test Drug	5	1 Min	1	Injection Site Pain VAS Score	100
136-0099	Test Drug	5	5 Min	2	Injection Site Pain VAS Score	43
136-0099	Test Drug	5	10 Min	3	Injection Site Pain VAS Score	2
136-0099	Test Drug	5	15 Min	4	Injection Site Pain VAS Score	4
136-0099	Test Drug	5	30 Min	5	Injection Site Pain VAS Score	2
136-0099	Test Drug	5	60 Min	6	Injection Site Pain VAS Score	2
136-0099	Test Drug	5	120 Min	7	Injection Site Pain VAS Score	2

INJECTION SITE TOLERABILITY: Injection Site Tolerability collects various Injection Site Reactions like Pain, Swelling, Tenderness etc. Injection site reactions are collected at various timepoints like Immediately After Injection, 1-Hour Post Injection, 4-Hours Post Injection, 24-Hours Post Injection, 48-Hours Post Injection and 72-Hours Post Injection. Result is evaluated as None, Mild, Moderate, Severe and Very Severe.

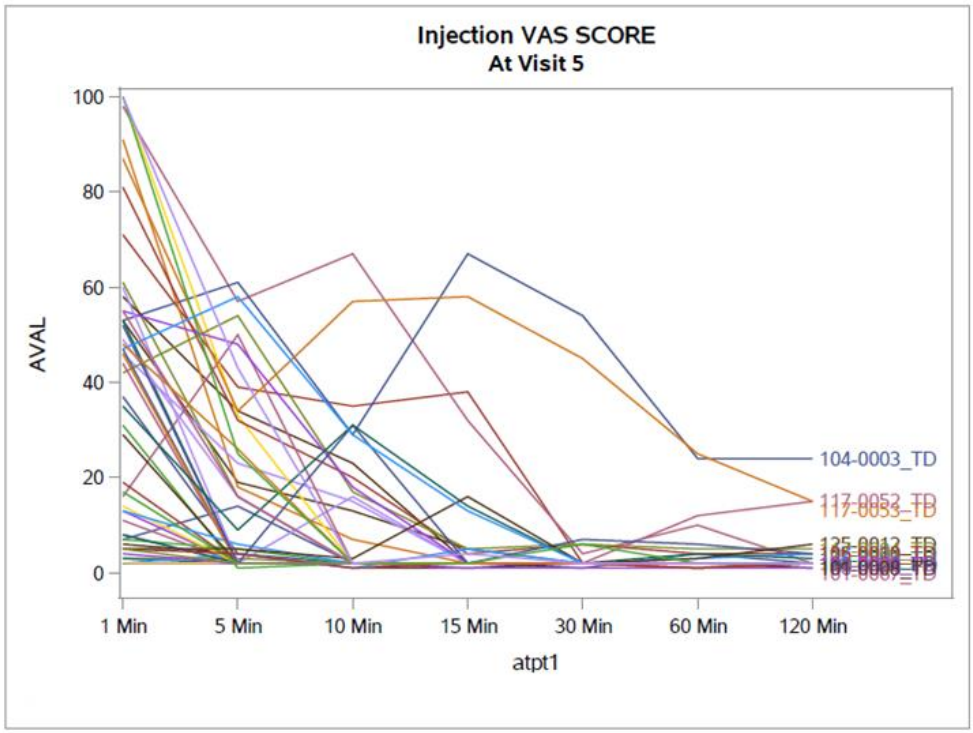
SAMPLE DATA

SUBJID	AVISITN	ATPT	ATPTN	PARAM	AVAL	AVALC
898-1002	1	Immediately After Injection	1	Injection Site Reaction Pain	4	Very Severe
898-1002	1	1 Hour Post Injection	2	Injection Site Reaction Pain	4	Very Severe
898-1002	1	4 Hours Post Injection	3	Injection Site Reaction Pain	3	Severe
898-1002	1	24 Hours Post Injection	4	Injection Site Reaction Pain	3	Moderate
898-1002	1	48 Hours Post Injection	5	Injection Site Reaction Pain	1	Mild
898-1002	1	72 Hours Post Injection	6	Injection Site Reaction Pain	0	None

SGPLOT procedure is used to develop various graphs presented in this paper. More details are presented in subsequent sections.

WHY LASAGNA OVER SPAGHETTI PLOTS?

This paper attempts to explore both the graphs for Injection Site Reactions like Pain, Swelling VAS Score collected at various visits.



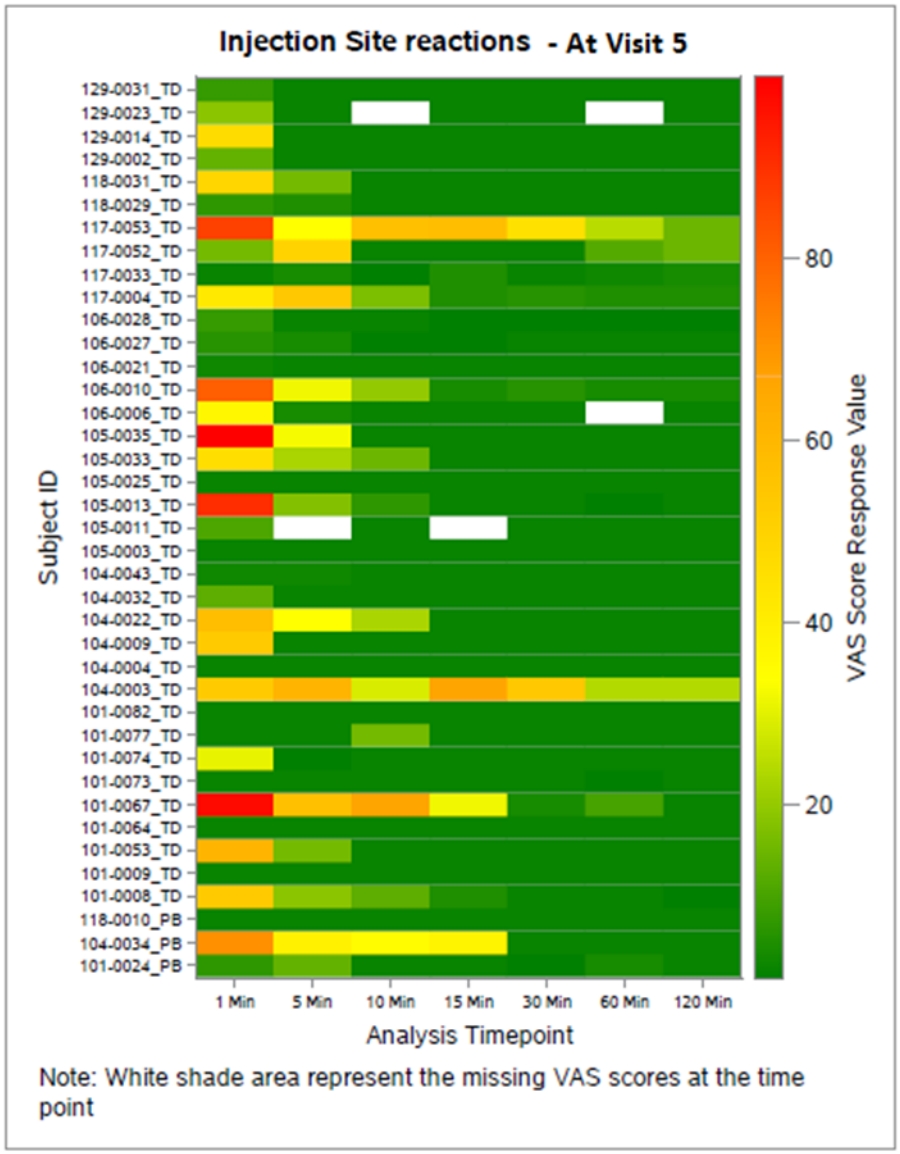
Graph 1. Spaghetti Plot

Each line represents a subject and its respective VAS score at each timepoint. By looking at this graph, it is very difficult to make out variability of each subject's VAS score. In case of large data, the lines will become a heap of noodles and it will be difficult to draw an inference about each subject or between subjects.

There are around 30 subjects presented in this graph. It is not pleasant to analyze as the lines are overlapped. The curve label option is used to identify the subject identifier. However, most of the subjects' data are overlapping and the subjects cannot be distinguished.

Inference:

For most of the subjects, the score can still be analyzed at 120 Min where the pain score is between 0 and 20. In between 5 to 10 Min, around 5 subjects have reached the maximum score



Graph 2. Lasagna Plot

X-axis represents the timepoint when the VAS pain score is collected. Y1 axis represents each subject and Y2 axis represents the indicator scale of plotted AVAL (result/response) values.

The Lasagna Plot is appealing as the subjects are displayed as layers and each subject score range can be seen per the timepoint. It is very easy to compare within the subjects and between the subjects. Also, it can be identified easily if the subject missed an assessment at any timepoint.

The graph displays the subject's VAS score with respect to the timepoints. Green indicates the least VAS Pain score. Red indicates the worst VAS Pain Score.

The color changes from Green-Yellow-Orange- Red as the intensity (Change of Pain Score over the period) of the Pain increases. The color indicator with specific AVAL value ranges is useful to observe the intensity Change of Pain Score over the period). The White shaded area represents the missing VAS scores at that point and the same can be indicated in the FOOTNOTE statement of the SGPLOT code.

Many rows have a dark shade on the left and a lighter shade to the right, which confirms the improvement of the response variable.

Inference:

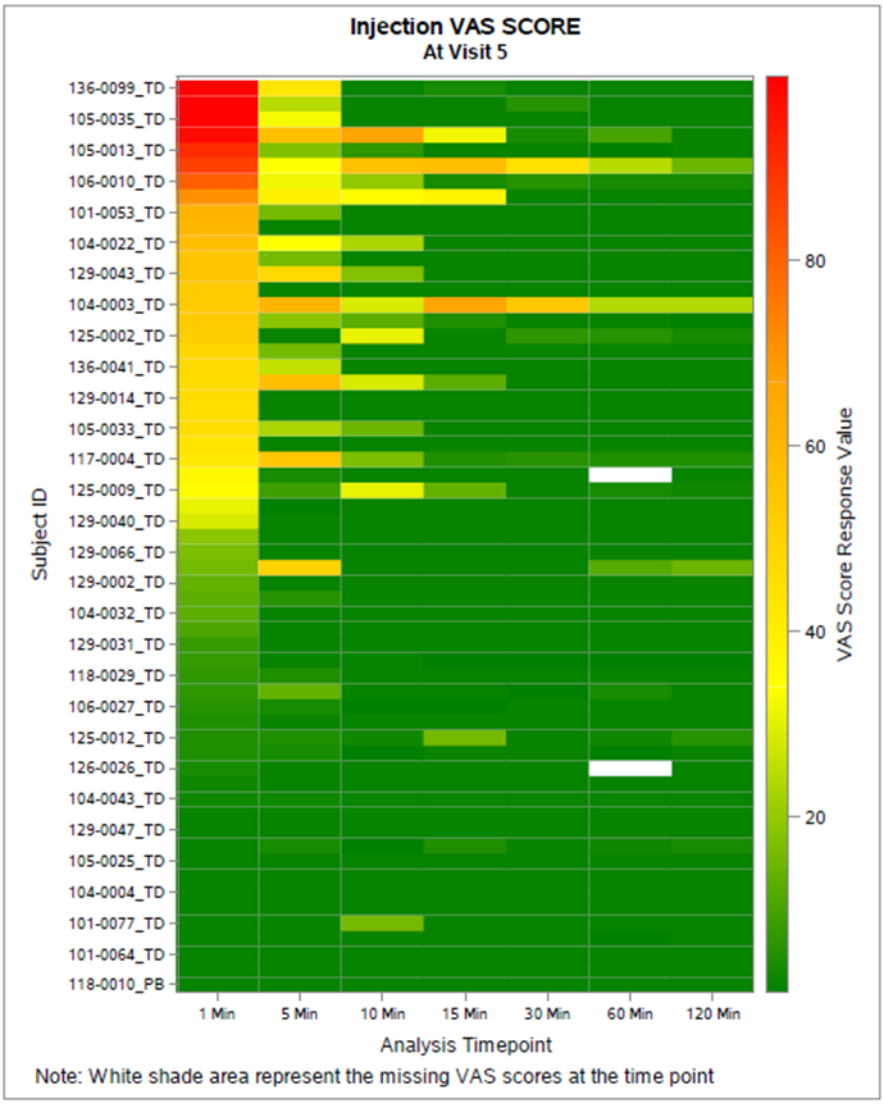
This graph helps to understand the variability of the response between subjects and within the subjects across time points. Trend analysis for every individual subject can be assessed/analyzed in a simple and precise manner. Also, a missing pattern or assessment can be easily identified with this graph. The color representation of the graph is visually appealing and helps identify the VAS score range just by looking at it.

SAS CODE

```
ods graphics on / width=700px height=900px discretemax=10000;

proc sgplot data=vas; /*Input dataset*/

where visitnum=5; /*Sub-setting condition*/

   heatmap x=atpt y=subjid / colorresponse=aval   /*Pass the Timepoint, Subjid and
Aval values in Heatmap statement */

                colormodel= (green yellow orange red); /*We can define colors*/

   yaxis valueattrs=(size=7pt) labelattrs=(size=10pt); /* We can define attributes for
y1axis*/

   y2axis label="VAS Score Response Value";/* We can define attributes for y2axis*/

   xaxis valueattrs=(size=7pt) labelattrs=(size=10pt) fitpolicy=thin; /* We can define
attributes for xaxis*/

   label aval="VAS Score Response Value" /* Define labels for all axis*/

         subject="Subject ID"

         atpt="Analysis Timepoint";

   footnote justify=left "Note: White shade area represent the missing VAS scores at
the time point"; /* Specify the footnote */

run;

ods graphics off;
```

DYNAMIC SORTING

Dynamic sorting is a useful strategy to better represent the data in the graph. Such a graph is shown below in Graph 3. The graphs that are represented by dynamic sort visualize data in a better way and per the user's point of view. These graphs are easy to analyze as required.



Graph 3. Dynamic sorting of VAS Score

Observe when the scores get improved, worsen, or stagnant over time. Also, it is easy to compare each subject because they are placed in horizontal rows. Retrieve the count of the subjects exclusively and easily within the required group per the sorting condition.

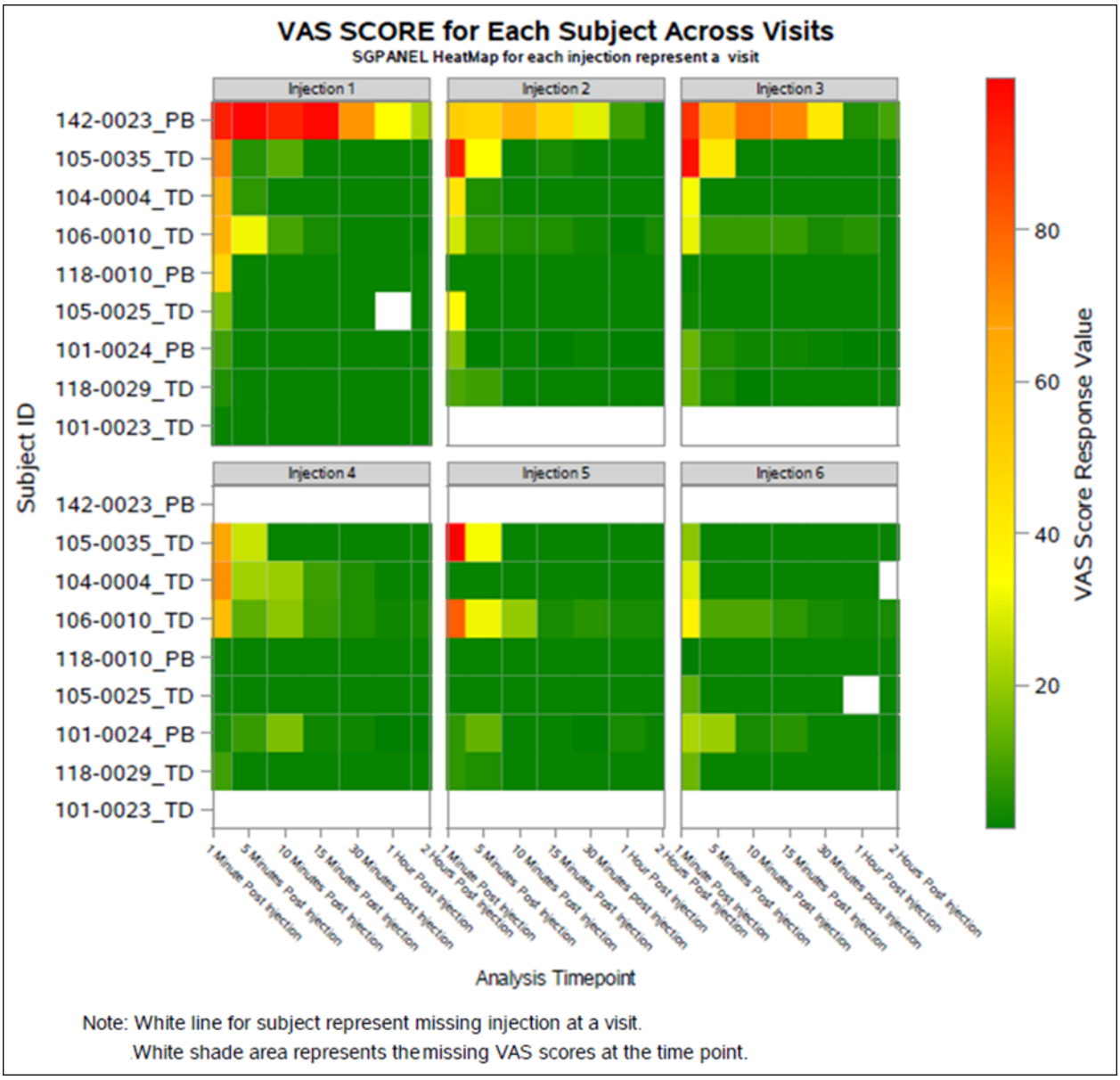
For example, when the graph is plotted based on various conditions, it is very easy to visualize the graph. It can be displayed with the specified groups based on the colored patterns. Handpick and count them easily, which is very difficult in the spaghetti plot.

Inference

Easy to analyze per the requirement when the data is dynamically sorted. Sorting helps to handpick the categories or subjects per graph, which has worse or improved cases at initial and final timepoints. Hence, it will be easy to understand the data like the number of subjects.

SGPANEL PLOT

In this graph, data is showed across the visits. Each visit represents an injection like Injection 1 is Visit 1, Injection 2 is Visit 2 etc. SG Panel procedure is used to plot the graph. the dataset is sorted and then it is fed to the SGPANEL PLOT procedure as an input. Data is sorted by Analysis Timepoint (N) (ATPTN) and Analysis Value (AVAL). Analysis Visit (AVISIT) is used as by parameter for SGPANEL procedure.



Graph 4. Dynamic sorting of VAS Score by SGPANEL

X-axis represents the timepoint when the VAS pain score is collected. Y1 axis represents each subject and Y2 axis represents the indicator scale of plotted AVAL (result/response) values.

Observe when the scores get improved, worsen, or stagnant over time across the analysis visits or injections. It is easy to study each subject across the analysis visits or injections as they are placed in adjacent each other. Also, it is easy to compare each subject because they are placed in horizontal rows. Retrieve the count of the subjects exclusively and easily within the required group per the sorting condition.

For example, when the graph is plotted based on various conditions, it is very easy to visualize the graph. It can be displayed with the specified groups based on the colored patterns across the injections. Handpick the subjects with the specific trends of VAS scores across the analysis visits and count them easily.

Inference

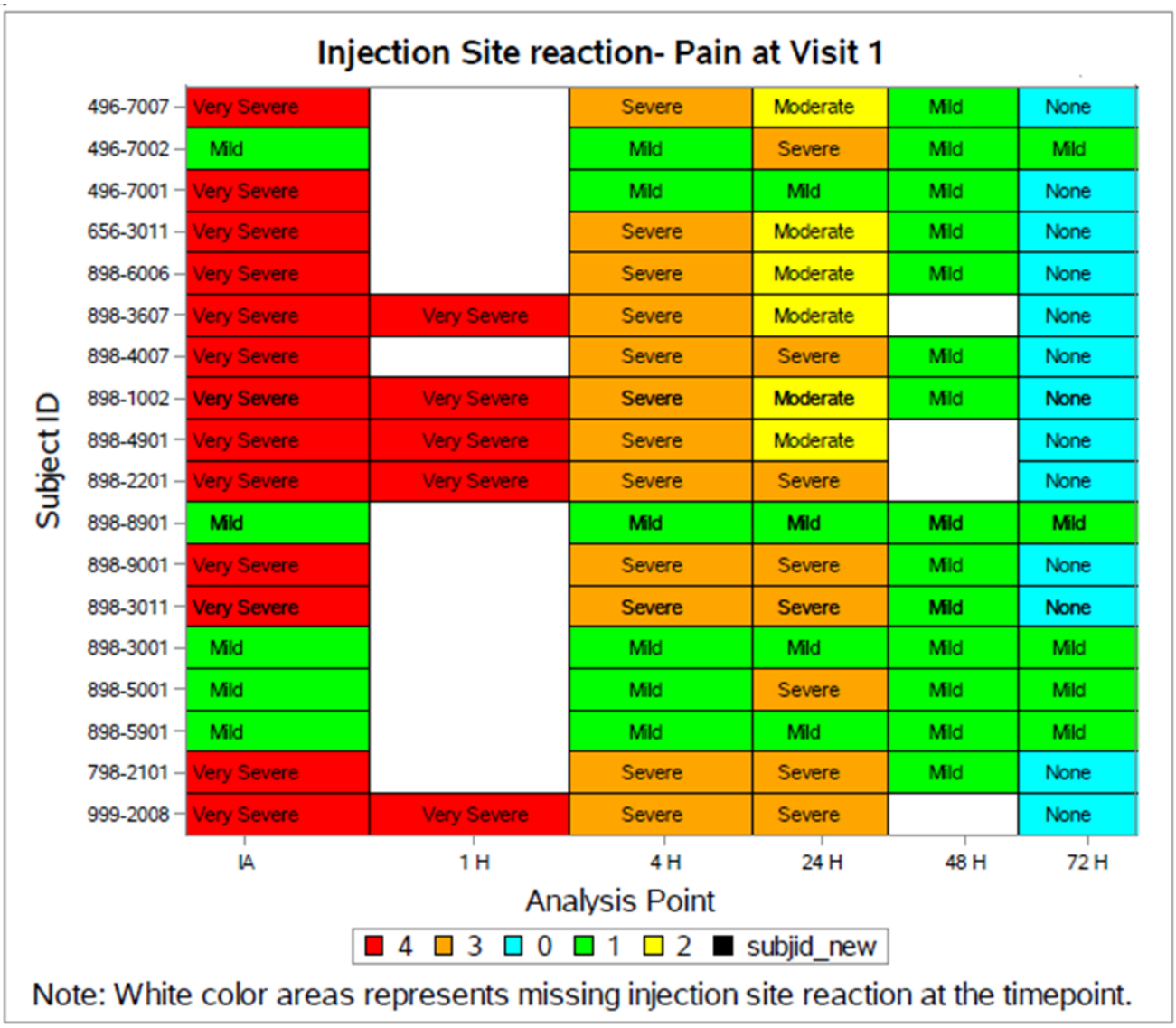
Easy to analyze per the requirement when the data is dynamically sorted and plotted across injections. SGPANEL Plot and Sorting helps to handpick the categories or subjects per visit, which has worse or improved cases at initial and final timepoints at various visits. Hence, it will be easy to understand the data like the number of subjects, their VAS Scores trends. Also, it gives an overview of total picture VAS scores of subjects and their trends across injections or analysis visits. It helps to see the bigger picture of data and to analyze the pain score.

SAS CODE

```
ods pdf file="/Desktop/vas_spa_panel_pot.pdf" ;
options orientation=portrait;
ods graphics / reset width=7.7in height=7.4in imagename='HeatMapPanel';
title 'VAS SCORE for Each Subject Across Visits';
title2 h=7pt 'SGPANEL HeatMap for each injection represent a visit';
proc sgpanel data=VAS ; /*Input dataset*/
  panelby avisit / novarname spacing=7 headerattrs=(size=7)
headerbackcolor=lightgray;/* Specifies one or more classification variables
for the panel*/
  heatmap x=atpt y=subjid /colorresponse=aval /*Pass the Timepoint, Subjid
and Aval values in Heatmap statement */
  colormodel= (green yellow orange red); /*We can define colors*/
  colaxis integer offsetmin=0 offsetmax=0 display=(noline)
        valueattrs=(size=6) labelattrs=(size=8);/*Specifies the axis options
for each X axis in the panel */
  keylegend;/* Adds a legend to the plot*/
    label aval="VAS Score Response Value"
         subjid1="Subject ID"
         atpt="Analysis Timepoint";/* Define labels for all axis*/
footnote justify=left "Note: White line for subject represent missing
injection at a visit."
 "White shade area represents the missing VAS scores at the time point."; /*
Specify the footnote */
run;
ods graphics off;
ods pdf close;
ods listing;
```

CATEGORICAL DATA:

Heatmap can be used to plot categorical data, which is not possible in the spaghetti plot. Other plots like the Mosaic plot can be used but it is not so feasible when there are more variables. To plot this graph, Injection Site Reaction Pain data is used. It helps to plot the longitudinal categorical data that is collected at various time intervals.



Graph 5. Injection Site Reaction Pain

X axis has timepoints as IA (Immediately After Injection), Post Injection 1 Hour (1H), 4 Hours (4H), 24 Hours (24H), 48 Hours (48H) and 72 Hours (72H). Y axis has Subject ID. Cyan color represents the response value as “None” which denotes there is no injection site reaction, Green color represents value as “Mild”, Yellow color represents value as “Moderate”, Gold color represents as value Severe and Red color represents Very Severe injection site reaction.

Inference:

Most of the “very severe” injection site reaction is seen immediately after injection and above graph depicts the changes in intensity at different timepoints and for most of the subjects at 72 Hour post

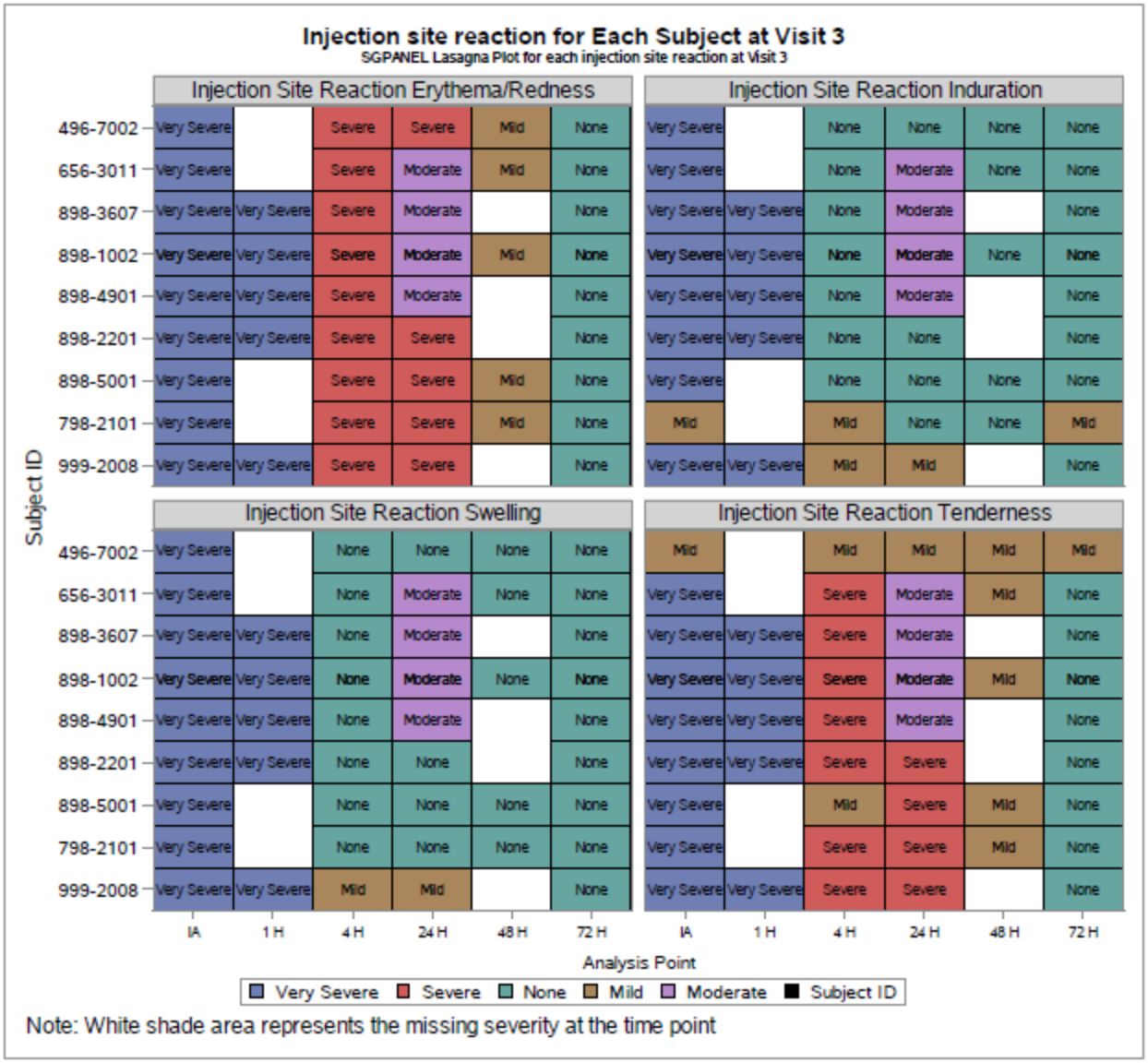
injection response value is None. It shows the improvement in pain post injection over the time. It is easy to understand the variability within or between subjects

SAS CODE

```
ods graphics on / width=1000px height=500px discretemax=10000;
proc sgplot data=isr dattrmap=isr1; /* Here we have specified attribute map
dataset*/
     where avisitn=1; /*Sub-setting condition*/
     heatmapparm x=atptn y=subjid colorgroup=aval  / outline attrID=avalc /*
Specify X, Y and colour group values*/
      tip=(atptn subjid avalc); /* gives the details when we place cursor on
the graph*/
      text x=atptn y=subjid text=avalc; /*Helps to annotate the graph with
values present in AVALC*/
        xaxis values=(1,2,3,4,5,6) valuesdisplay =('IA' '1 H' '4 H' '24 H'
'48 H' '72 H')  valueattrs=(size=7) label="Analysis Point"; /* Values
specifies values range on X axis. Valuesdisplay displays values specified as
values of X axis. Label- specifies label for x axis*/
keylegend; /* Adds a legend to the plot*/
yaxis valueattrs=(size=7) label="Subject ID" fitpolicy=none;
footnote justify=left "Note: White color areas represents missing injection
site reaction at the timepoint."; /* Specify the footnote */
run;
ods graphics off;
```

SGPANEL PLOT

In this graph, the various Injection Site Reactions like “Injection Site Reaction Erythema/Redness”, “Injection Site Reaction Swelling”,” Injection Site Reaction Induration” and “Injection Site Reaction Tenderness” of Analysis Visit 3 are shown. It is easy to observe the severity scores for injection site reaction.



Graph 6. SGPANEL- Injection Site Reactions

X axis has timepoints as IA (Immediately After Injection), Post Injection 1 Hour (1H), 4 Hours (4H), 24 Hours (24H), 48 Hours (48H) and 72 Hours (72H). Y axis has Subject ID. Teal color represents the response value as “None” which denotes there is no injection site reaction, Brown color represents value as “Mild”, Purple color represents value as “Moderate”, Brick Red color represents as value Severe and Blue color represents Very Severe injection site reaction.

Inference:

Most of the “very severe” injection site reactions are seen immediately after injection and above graph depicts the changes in intensity at different timepoints and for most of the subjects at 72 Hour post injection response value is None. It shows the improvement in injection site reactions post injection over the time. It is easy to understand the variability within or between subjects and within or between injection site reactions

SAS CODE

```
ods pdf file="/ Desktop/param_spa_panel_pot.pdf" ;
options orientation=landscape;
ods graphics / reset width=8in height=7.4in imagename='HeatMapPanel'
discretemax=10000;
 title'Injection site reaction for Each Subject at Visit 3';
 title2 h=7pt'SGPANEL Lasagna Plot for each injection site reaction at Visit
3';
proc sgpanel data= isr dattrmap=anno_isr; ; /* Here we have specified
attribute map dataset*/
panelby param / novarname spacing=7 headerattrs=(size=10)
headerbackcolor=lightgray; ;/* Specifies one or more classification variables
for the panel*/
    heatmapparm  x=atpt y=subjid colorgroup=avalc  /  outline attrID=aval
            tip=(atptnew subjid_new avalc) showxbins  ;/* Creates a two-
dimensional plot*/
          text x=atpt y=subjid text=avalc /textattrs=(size=7pt)
splitpolicy=split splitwidth=25 position=center ;
  colaxis values=(1,2,3,4,5,6) valuesdisplay =('IA' '1 H' '4 H' '24 H' '48 H'
'72 H')
  offsetmin=0 offsetmax=0    valueattrs=(size=7) labelattrs=(size=8)
label="Analysis Point" fitpolicy=none ; /*Specifies the axis options for each
X axis in the panel */

keylegend  ; /* Adds a legend to the plot*/
 label      subjid ="Subject ID"
               atpt="Analysis Timepoint";
  footnote justify=left "Note: White shade area represents the missing
severity at the time point"; /* Specify the footnote */
run;
  ods graphics off;
ods pdf close;
ods listing;
```

ADVANTAGES

- Comparison within subject or between subjects is easier
- Missing data pattern can be easily identified
- Trend analysis is made easy
- More effective when sort by required group or response value
- Visualize the shift in response values for subjects in longitudinal data
- Visualization is clear
- Ability to plot categorical data
- No over plotting

DISADVANTAGES

- Needs lots of place to display rows as it avoids over plotting
- Cannot fit larger number of subjects in a single graph

RESTRICTIONS

Data should be recorded for every individual subject or parameter at the same set of uniformly spaced time intervals, such as daily, monthly, or yearly.

CONCLUSION

The Lasagna plot is used when the data is uniformly spaced by different timepoints. Analysis can be done within and between subjects' comparison. This graph is a perfect fit when it is plotted per categories and/or with less number of subjects.

Using this graph, missing data pattern can be identified, data can be visualized clearly, shift in result response of longitudinal data can be observed.

Categorical data can have defined color patterns, which helps in hand picking the subjects per the requirements. Dynamic sorting helps analyze the trend of the outcome at the first and last timepoints. This sorting arranges the responses from a range of worst to best or vice versa. Analysis of data becomes simpler with such data categorization.

Panel Plots will make visualization much simpler to understand the variability within or between subjects and within or between selected subgroups or parameters

REFERENCES

- SAS Blogs – "Lasagna plots in SAS: When spaghetti plots don't suffice" by Rick Wicklin on The DO Loop, 2016
- BMC Medical Research Methodology Article – "Visualising and modelling changes in categorical variables in longitudinal studies" by Mark Jones, Richard Hockey, Gita D Mishra and Annette Dobson, 2014
- SAS Global Forum 2007 – Paper 193-2007 – "Procedures for Creating Statistical Graphics in Data Analysis" by Dan Heath, SAS Institute Inc., Cary, NC, 2007
- Swihart et al. (2010) – "Lasagna Plots: A Saucy Alternative to Spaghetti Plots", By Swihart, Bruce J.a; Caffo, Briana; James, Bryan D.b; Strand, Matthewc; Schwartz, Brian S.d; Punjabi, Naresh M.e, 2010
- SAS® Help Centre https://documentation.sas.com/

ACKNOWLEDGMENTS

A special thanks to Srivalli Konda who gave us a helping hand in meticulously organizing these words.

RECOMMENDED READING

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

CONTACT INFORMATION

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

Soujanya Konda
IQVIA Pvt Ltd..
soujanya.konda@iqvia.com; soujitheuniq@gmail.com
gsk.com | Twitter | YouTube | Facebook | Flickr