

Effective Graphical Representation of Tumor Data

Sanjay Matange

ABSTRACT

The “Duration of Treatment” and “Tumor Response” information for subjects in a study has traditionally been visualized in separate graphs where the subjects may be sorted by different criteria. In such cases, the Clinician must work harder to associate the subject across different graphs. Displaying the data together, sorted by the tumor response makes it easier for the Clinician to understand this information.

3D waterfall graphs have been proposed for such visuals. This paper will show you how to build a 3D Waterfall Graph with tumor response and duration of treatment in one graph using SAS®. We will also discuss the features of this approach. Finally, we will present effective 2D alternative displays to visualize all the data in one combined graph thus making it easier to decode the information.

INTRODUCTION

In the paper “3D Waterfall Plots: A better graphical representation of tumor response in oncology”, authors Alvarez et. al. provide a compelling case for inclusion of more relevant data in graphs for the analysis of tumor data.

Classical waterfall plots traditionally display the tumor response data sorted by response. The authors indicate that readers often need to consult additional related information such as duration of treatment to get a better understanding of the study results. The authors propose to add the display of duration of treatment into a 3D waterfall plot that provides combined information useful to the reader.

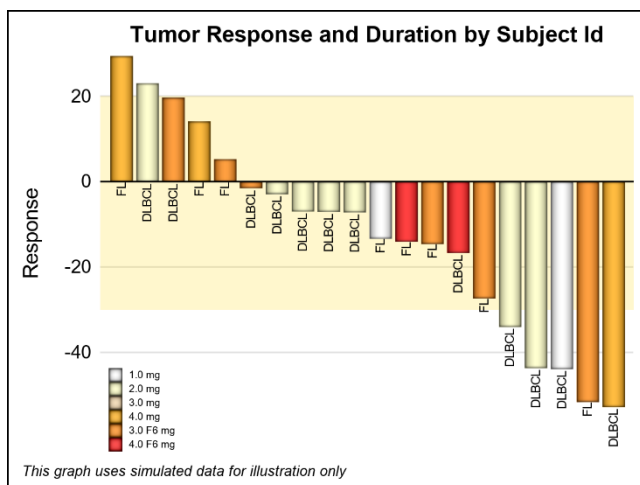


Figure 1 – Traditional Waterfall plot

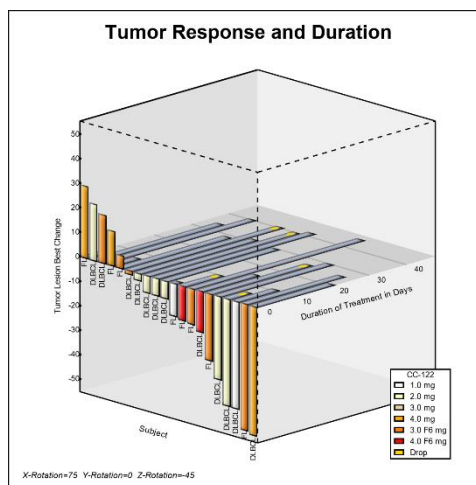


Figure 2 – 3D Waterfall plot

Figure 1 shows the tumor response in a study sorted by increasing response in a traditional waterfall plot. In this example, the bars are colored by the treatment level and labeled by type of lymphoma (DLBCL or FL) at the bottom of the bar. A band is displayed between +20% and -30% change in tumor size.

Figure 2 shows a 3D graph where the tumor response sorted by increasing response is displayed on the front vertical face of the graph. Again, the bars are colored by treatment level and the type of lymphoma is shown at the bottom of the bar. The duration of treatment for each subject is displayed on the middle horizontal plane of the graph. Note, the data for both graphs are simulated for illustration purposes only.

3D WATERFALL GRAPH

The simulated data for 3D Waterfall graph is shown in Figure 3. The data is sorted by increasing response for tumor size. Data columns include the treatment, code, response, duration of treatment, etc.

Obs	j	Drug	Code	Response	Duration	Dropped	codeloc	baseline
1	1	4.0 mg	FL	29.3978	4.8311	.	0.0000	192.734
2	2	2.0 mg	DLBCL	23.0295	33.9761	.	0.0000	175.102
3	3	3.0 F6 mg	DLBCL	19.6369	40.1963	.	0.0000	93.218
4	4	4.0 mg	FL	14.0812	6.9348	.	0.0000	172.023
5	5	3.0 F6 mg	FL	5.2305	28.8763	.	0.0000	121.489

Figure 3 – Simulated data for 3D Waterfall graph

The graph shown in Figure 5 is created using the SAS macro %Waterfall_3D_Macro () using the data shown in Figure 3. The macro is published in the GraphicallySpeaking SAS web page. See the web page for the details and additional options. Some macro options are shown in Figure 4.

%WaterFall_3D_Macro (Data=tumor, Duration=duration, Response=response, Dropped=dropped, Group=Drug, Code=Code);

Parameter	Description
Data	Data set
Duration	Column for Treatment Duration
Response	Column for Tumor Response
Dropped	Column for dropped subject
Group	Column for Treatment Group
Code	Column for Lymphoma code displayed at end of bar

Figure 4 – 3D Waterfall graph data

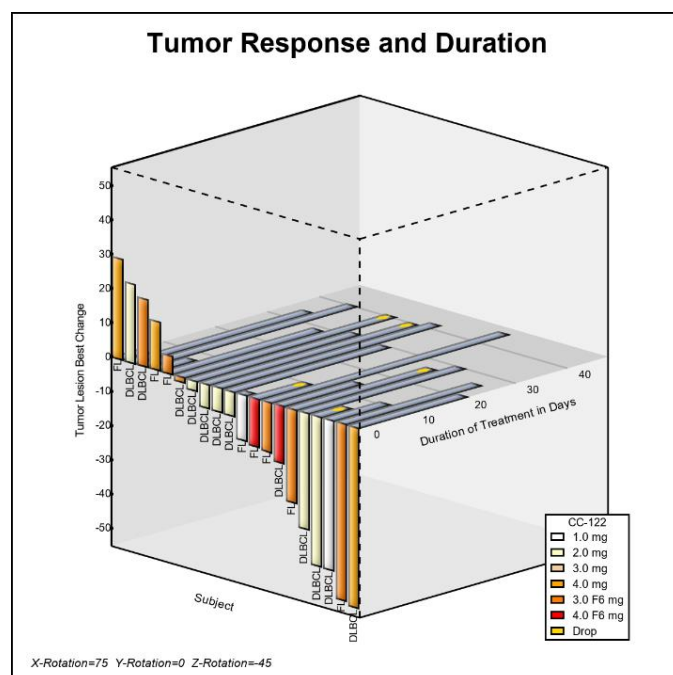


Figure 5 – 3D Waterfall graph

The data is sorted by increasing reduction in tumor size (Response). The %Waterfall_3D_Macro() lays out the provided data in 3D space as follow to render the graph:

- The 3D cube has horizontal x axis, y-axis going into the page and z-axis going up.
- All the bars are coded as polygons in 3D coordinate space in the appropriate plane.
- The tumor response bars are arranged on the front face ($y=0$) of the 3D cube.
- Lymphoma type is displayed at the bottom of the response bars.
- The duration of treatment bars are arranged on the horizontal plane at $z=0$.
- Markers are placed on the “duration” bars for dropped subjects.
- The 3D data is then transformed and projected into 2D space.
- The 2D polygonal data is now displayed using the SGPLOT procedure.

The 3D graph shown in Figure 5 has a modern and slick appearance. The following features of this visual representation are observed.

- Some of the “duration” bars are hidden behind the “response” bars.
- The “dropped” markers on the duration bars are not clearly visible.
- There is a lot of unused space in this visual. Density of information pixels may be less than 20%.
- This is a custom bar layout and adding more data to this display requires custom code.
- This data is really 2D in nature with one independent variable (subject id sorted by response). All response, duration and dropped information are dependent only the subject id.

One benefit of this arrangement is that the duration and response bars are touching in the middle, thus providing a connection between the data.

2D WATERFALL GRAPH WITH ADDITIONAL DATA

This paper proposes an alternate, simpler 2D representation of the same data as shown in Figure 6.

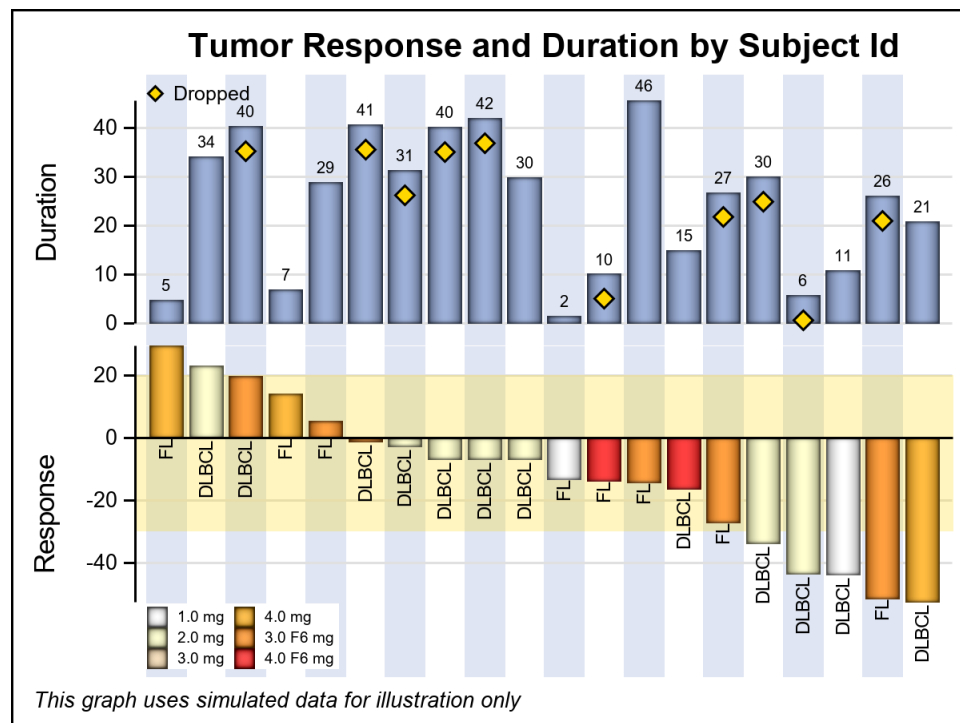


Figure 6 – 2D Waterfall Graph with duration data.

The program using the Graph Template Language (GTL) is shown below, using a two row Lattice.

```
proc template;
  define statgraph Waterfall_Plus;
    begingraph / axislineextent=data;
      entrytitle 'Tumor Response and Duration by Subject Id';
      entryfootnote halign=left 'This graph uses simulated data for ... ' /
        textattrs=(size=7pt style=italic);
      layout lattice / columndatarange=union rowweights=(0.45 0.55) rowgutter=0;
        columnaxes;
          columnaxis / display=none discreteopts=(colorbands=odd
            colorbandsattrs=(transparency=0.2));
        endcolumnaxes;

      /*--Define the upper cell with duration and dropped data--*/
      layout overlay / yaxisopts=(griddisplay=on offsetmax=0.1
        tickvalueattrs=(size=7) labelattrs=(size=9)) walldisplay=none;
        barchartparm category=j response=duration / datalabel=duration
          fillattrs=graphdata1 datalabelattrs=(size=5)
          dataskin=presseddisplaybaseline=auto;
        scatterplot x=j y=dropped / markerattrs=(symbol=diamondfilled size=9)
          filledoutlinedmarkers=true markerfillattrs=(color=gold)
          markeroutlineattrs=(color=black) name='d'
          legendlabel='Dropped';
        discretelegend 'd' / location=inside valign=top halign=left
          valueattrs=(size=7) border=false autoitemsizetrue;
      endlayout;

      /*--Define the lower cell for display of response data--*/
      layout overlay / yaxisopts=(griddisplay=on tickvalueattrs=(size=7)
        labelattrs=(size=9) offsetmax=0
        linearopts=(tickvaluepriority=true)) walldisplay=none;
        bandplot x=j limitupper=20 limitlower=-30 / extend=true
          fillattrs=(color=gold transparency=0.75);
        barchartparm category=j response=response / group=drug
          groupdisplay=cluster datalabelattrs=(size=5) dataskin=pressedd
          name='a' datalabelfitpolicy=rotate;
        textplot x=j y=codeloc text=code / rotate=90 position=left
          textattrs=(size=6) contributeoffsets=(ymin);
        discretelegend 'a' / location=inside valign=bottom halign=left
          order=columnmajor down=3 opaque=true valueattrs=(size=5)
          border=false;
      endlayout;
    endlayout;
  endgraph;
end;
run;

/*--Render the graph using the defined template--*/
ods graphics / reset width=4in height=3in;
proc sgrender template=Waterfall_Plus data=tumorsorted dattrmap=attrmap;
format duration 3.0;
dattrvar drug="Resp";
run;
```

Figure 7 – GTL program for 2D Waterfall with Duration data

This representation has many advantages:

- The 2D data is represented in a 2D visual, without any “chart Junk”.
- The sorted response bars are shown in the lower cell, colored by treatment.
- The Lymphoma type is clearly labeled at the bottom of the bar.
- The duration of treatment for each subject is shown in the upper cell.

- The duration value can be easily displayed at the top of each bar.
- The “dropped” markers are clearly visible.
- Alternate vertical blue bands are added for alignment.
- Multiple legends are easier to place.
- The two y-axes for Duration and Response are clearly marked and readable.
- A band can be placed on the response bars indicating the +20 to -30 region.
- The “information density” is much higher, with better utilization of the space.
- This graph can be easily created as a 2-row Lattice using GTL.

2D WATERFALL GRAPH WITH MORE DATA

A desire was expressed to add more relevant data to this graph. In one such case we add the display of the tumor burden for each subject in the same graph as shown in Figure 8.

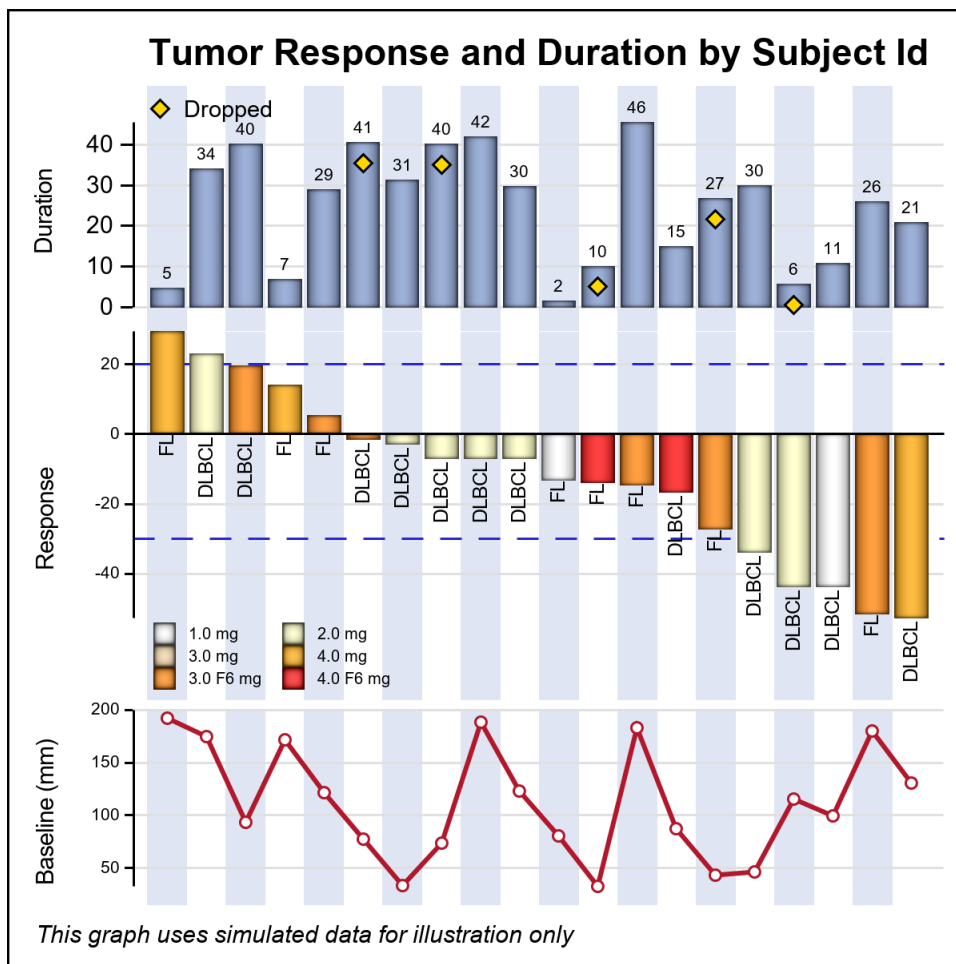


Figure 8 – GTL program for 2D Waterfall with Duration and Baseline data

There is one independent variable (subject id), and it is easy to extend the graph in Figure 6 as follows:

- A third cell is added below the “Response” cell in the graph.
- Tumor baseline values by subject id are clearly displayed.
- The baseline values could be displayed as needles, scatter markers or line (as shown).
- Y-axis for baseline is clearly displayed.
- Note, the +20 to -30 band is displayed with outlines only.
- This process can be easily extended to accommodate more data.

In Figure 8, the baseline tumor load is displayed as a series plot, with the y-axis increasing from bottom to top. It may not be immediately clear due to lack of a baseline. It is also possible to display the tumor burden as a needle or bar plot with minor modification to the GTL template as shown in Figure 9.

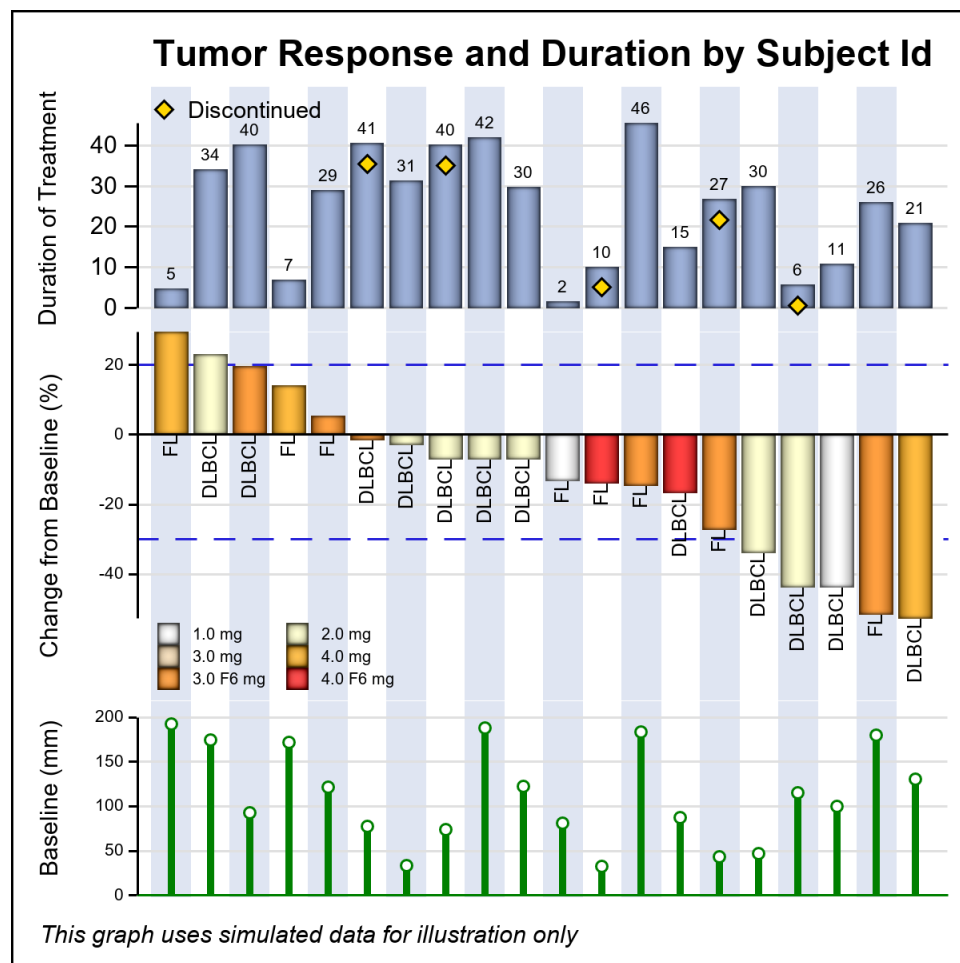


Figure 9 – GTL program for 2D Waterfall with Duration and Baseline data

SWIMMER PLOT WITH TUMOR DATA

In the previous examples, we have extended the tumor response data by adding the duration of treatment data for each subject. Additional columns can be added to display other related information as shown in Figure 8 and Figure 9. Now, let us look at an alternate visualization of tumor data. We will start with the data for creating a Swimmer Plot and extend it with tumor response information to create a combined Waterfall + Swimmer plot.

Figure 10 shows some simulated data used to create a Swimmer plot, displaying the Tumor Response “story” for subjects in a study. The data includes one observation per subject showing the stage of the disease and one response interval. For subjects having multiple response durations, additional observations are included for the same subject “Id”. Disease stage and response status is included per episode.

Figure 11 shows a Swimmer Plot as proposed by Stacey Phillips, created from the data in Figure 10 using the SGPLOT procedure.

Obs	id	item	stage	low	high	highcap	status	start	end	durable	response	startline	endline
1	1	9	Stage 4	0	8.3	FilledArrow	Partial response	6.0	.	.	20	6.0	8.0
2	2	6	Stage 2	0	12.6	FilledArrow	Partial response	2.5	7.0	.	10	2.5	7.0
3	2	6		0	12.6	FilledArrow	Partial response	9.5	.	.	10	9.5	12.3
4	3	5	Stage 1	0	12.5	FilledArrow	Complete response	3.5	4.5	-0.5	-15	3.5	4.5
5	3	5		0	12.5	FilledArrow	Complete response	6.5	8.5	.	-15	6.5	8.5
6	3	5		0	12.5	FilledArrow	Partial response	10.5	.	.	-15	10.5	12.2
7	4	4	Stage 4	0	13.5	FilledArrow	Partial response	7.0	10.0	.	-20	7.0	10.0
8	4	4		0	13.5	FilledArrow	Partial response	11.5	.	.	-20	11.5	13.2
9	5	3	Stage 3	0	14.0	FilledArrow	Partial response	2.5	3.5	-0.5	-22	2.5	3.5
10	5	3		0	14.0	FilledArrow	Partial response	6.0	.	.	-22	6.0	13.7
11	6	8	Stage 1	0	9.5		Complete response	1.0	9.5	-0.5	-25	1.0	9.5
12	7	7	Stage 3	0	11.5		Complete response	4.5	11.5	-0.5	-30	4.5	11.5
13	8	10	Stage 2	0	7.2	FilledArrow	Complete response	1.2	.	.	-42	1.2	6.9
14	9	2	Stage 2	0	17.0		Complete response	10.5	17.0	-0.5	-45	10.5	17.0

Figure 10 – Data for Swimmer Plot with added columns for tumor response.

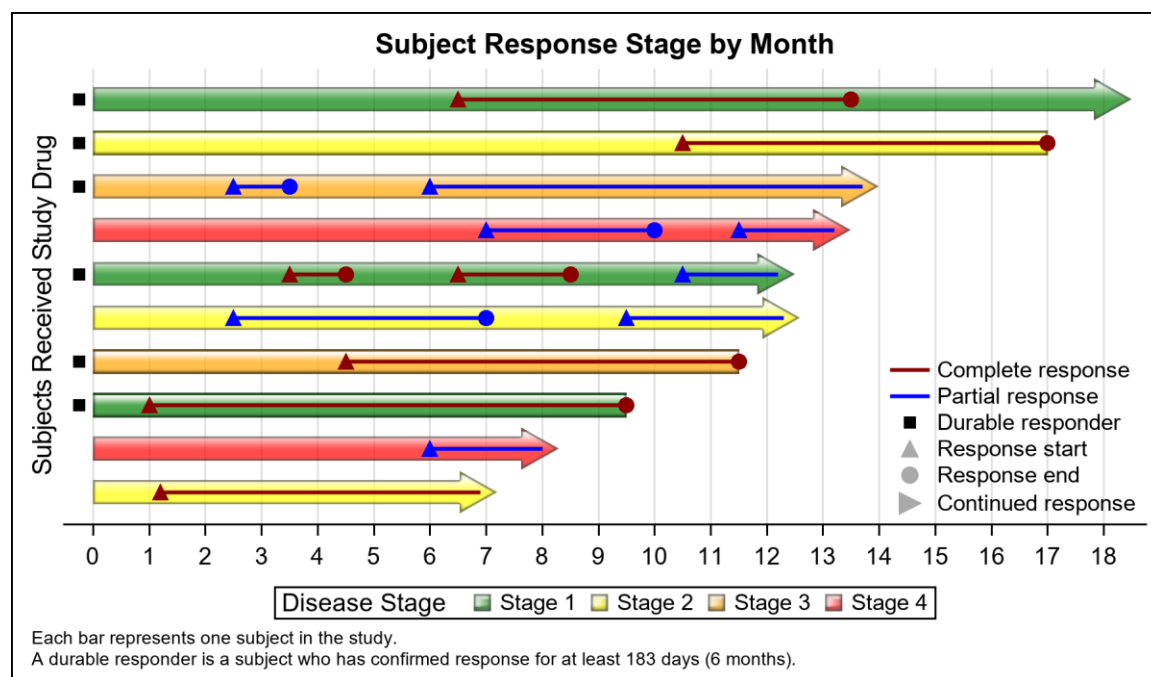


Figure 11 – Swimmer plot of subjects receiving treatment by drug.

The graph in Figure 11 shows the following details:

- One bar is displayed for each subject in the study colored by the disease stage.
- Continuing response is indicated by the arrow at the right end of the bar.
- A discrete attributes map is used to color the disease stage with custom colors.
- Treatment episodes are displayed as overlaid lines with start and end markers.
- The treatment episode is color coded by the status of the response.
- Durable responders are indicated by the markers at the left end of the bars.
- A legend of the disease stage is displayed at the bottom
- A legend of the response status and various markers is displayed in the graph at bottom right.

The SGPLOT Procedure program for this Swimmer graph is shown in Figure 12.

```

title 'Subject Response Stage by Month';
footnote J=1 h=0.8 'Each bar represents one subject in the study.';
footnote2 J=1 h=0.8 'A durable responder is a subject who has confirmed
response for at least 183 days (6 months).';
proc sgplot data= swimmer dattrmap=attrmap nocycleattrs noborder;
  legenditem type=marker name='ResStart' /
    markerattrs=(symbol=trianglefilled
      color=darkgray size=9) label='Response start';
  legenditem type=marker name='ResEnd' / markerattrs=(symbol=circlefilled
    color=darkgray size=9) label='Response end';
  legenditem type=marker name='RightArrow' /
    markerattrs=(symbol=trianglerightfilled
      color=darkgray size=12) label='Continued response';
  highlow y=item low=low high=high / highcap=highcap type=bar
    group=stage fill nooutline dataskin=gloss
    lineattrs=(color=black) name='stage' barwidth=1
    nomissinggroup transparency=0.3 attrid=stage;
  highlow y=item low=startline high=endline / group=status
    lineattrs=(thickness=2 pattern=solid) name='status' nomissinggroup
    attrid=status;
  scatter y=item x=durable / name='Durable' legendlabel='Durable responder'
    markerattrs=(symbol=squarefilled size=6 color=black);
  scatter y=item x=start / group=status attrid=status
    markerattrs=(symbol=trianglefilled size=8);
  scatter y=item x=end / markerattrs=(symbol=circlefilled size=8)
    group=status attrid=status;
  xaxis display=(nolabel) values=(0 to 20 by 1) valueshint grid;
  yaxis reverse display=(noticks novalues noline)
    label='Subjects Received Study Drug' min=1;
  keylegend 'stage' / title='Disease Stage';
  keylegend 'status' 'Durable' 'ResStart' 'ResEnd' 'RightArrow' /
    noborder location=inside position=bottomright
    across=1 linelength=20;
run;

```

Figure 12 – SGPLOT Procedure code for Swimmer Plot.

The data set in Figure 10 already has the tumor response information, we can create a combined Waterfall + Swimmer plot as shown in Figure 13. Note, this 2-Row graph cannot be made using the SGPLOT procedure. We will use GTL to create this graph using a 2-row Lattice.

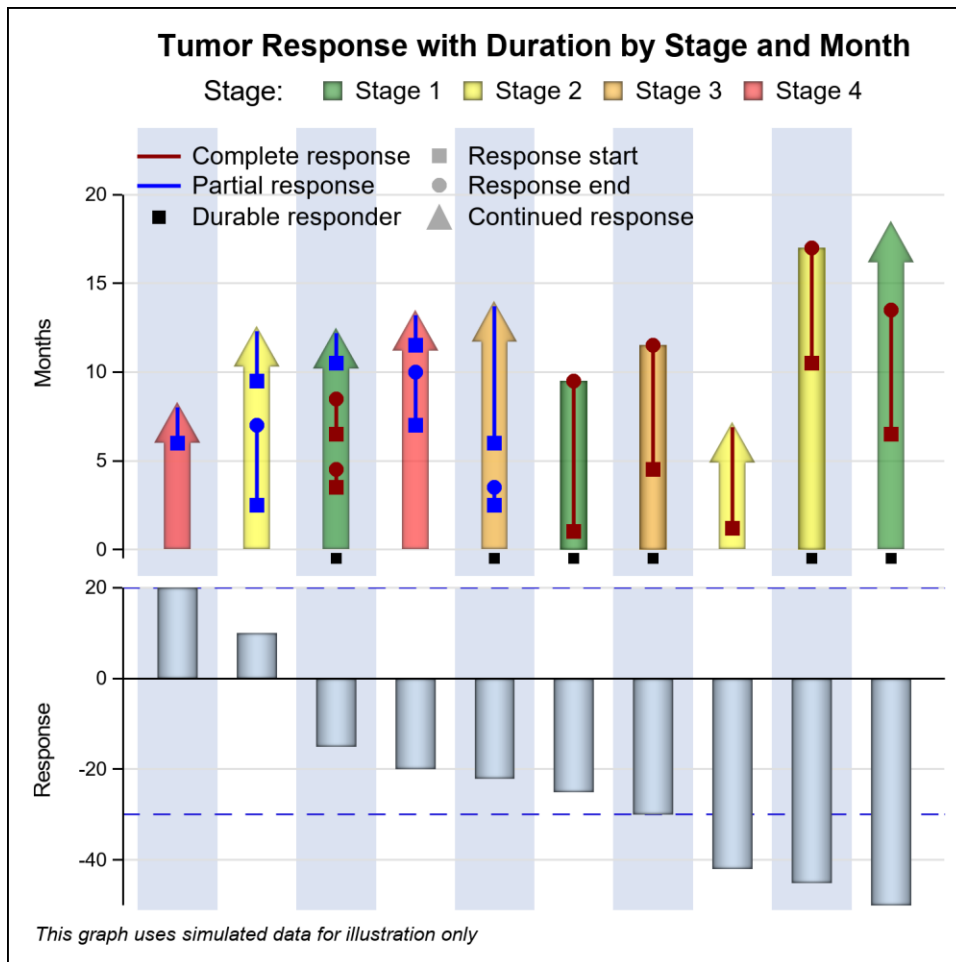


Figure 13 – Combined Waterfall + Swimmer plot from data shown in Figure 10

Let us review the details of this graph.

- The tumor response is shown in the lower cell sorted by increasing reduction in size.
- The subject response history is shown in the upper cell, aligned with the bar for tumor response.
- The graph in the upper cell is essentially a rotated version of the graph in Figure 11.
- The upper cell shows a vertical bar for each subject for the duration of the response by stage.
- Continued response is indicated by the arrow head.
- Individual treatment episodes are displayed by the line with start and end markers by status.
- Durable responders are indicated by the markers at the bottom of each bar.
- A legend for the disease stage is shown at the top.
- A legend for the status and other markers is shown inside the upper cell.
- Vertical alternate blue bands help align the related bars.
- Additional information can be added to this graph in cells above or below.

The program GTL code for this graph is shown below.

```
proc template;
  define statgraph Swimmer_With_Response;
    begingraph / axislineextent=data;
      entrytitle 'Tumor Response with Duration by Stage and Month';
      entryfootnote halign=left 'This graph uses simulated data for illustration' /
        textattrs=(size=7pt style=italic);
```

```

legenditem type=marker name='ResStart' / label='Response start'
    markerattrs=(symbol=squarefilled color=darkgray size=7);
legenditem type=marker name='ResEnd' / label='Response end'
    markerattrs=(symbol=circlefilled color=darkgray size=7);
legenditem type=marker name='RightArrow' / label='Continued response'
    markerattrs=(symbol=trianglefilled color=darkgray size=12);

layout lattice / columndatarange=union rowweights=(0.6 0.4) rowgutter=0;
    columnaxes;
        columnaxis / display=none type=discrete
            discreteopts=(colorbands=odd colorbandsattrs=(transparency=0.1));
    endcolumnaxes;

    layout overlay / yaxisopts=(griddisplay=on offsetmax=0.15
        tickvalueattrs=(size=7) labelattrs=(size=8) label='Months')
        walldisplay=none;

        highlowplot x=id low=low high=high / highcap=highcap type=bar group=stage
            dataskin=pressedd lineattrs=(color=black) name='stage' barwidth=0.6
            includemissinggroup=false datatransparency=0.3;
        highlowplot x=id low=startline high=endline / group=status
            lineattrs=(thickness=2 pattern=solid) name='status'
            includemissinggroup=false;
        scatterplot x=id y=durable / name='Durable'
            legendlabel='Durable responder'
            markerattrs=(symbol=squarefilled size=6 color=black);
        scatterplot x=id y=start /
            markerattrs=(symbol=squarefilled size=8) group=status;
        scatterplot x=id y=end /
            markerattrs=(symbol=circlefilled size=8) group=status;
        discretelegend 'stage' / title='Disease Stage' valign=top border=false
            title='Stage: ';
        discretelegend 'status' 'Durable' 'ResStart' 'ResEnd' 'RightArrow' /
            order=columnmajor halign=left valign=top border=false
            location=inside down=3 itemsize=(linelength=20);
    endlayout;
    layout overlay / yaxisopts=(griddisplay=on tickvalueattrs=(size=7)
        labelattrs=(size=8) offsetmax=0
        linearopts=(tickvaluepriority=true)) walldisplay=none;
        bandplot x=id limitupper=20 limitlower=-30 / extend=true
            display=(outline)
            outlineattrs=graphdatal(pattern=dash thickness=1);
        barchartparm category=id response=response / barwidth=0.5
            datalabelattrs=(size=5) dataskin=pressedd;
    endlayout;
endlayout;
endgraph;
end;
run;

ods graphics / reset width=5in height=5in imagename='Swimmer_Plus';
proc sgrender template=Swimmer_With_Response data=swimmer_sort_2 dattrmap=attrmap;
label response='Response';
dattrvar stage='stage' status='status';
run;

```

Figure 14 – GTL program for combined Waterfall + Swimmer plot.

With some modifications, the graph in Figure 12 can be arranged horizontally as shown in Figure 15. The changes primarily involve the following:

- Set the number of LATTICE columns to two (columns=2).
- Change ROWWEIGHTS to COLUMNWEIGHTS.
- Change COLUMNAXES to ROWAXES.
- Change YAXISOPTS to XAXISOPTS.
- Change variable assignments to plot roles from X to Y.

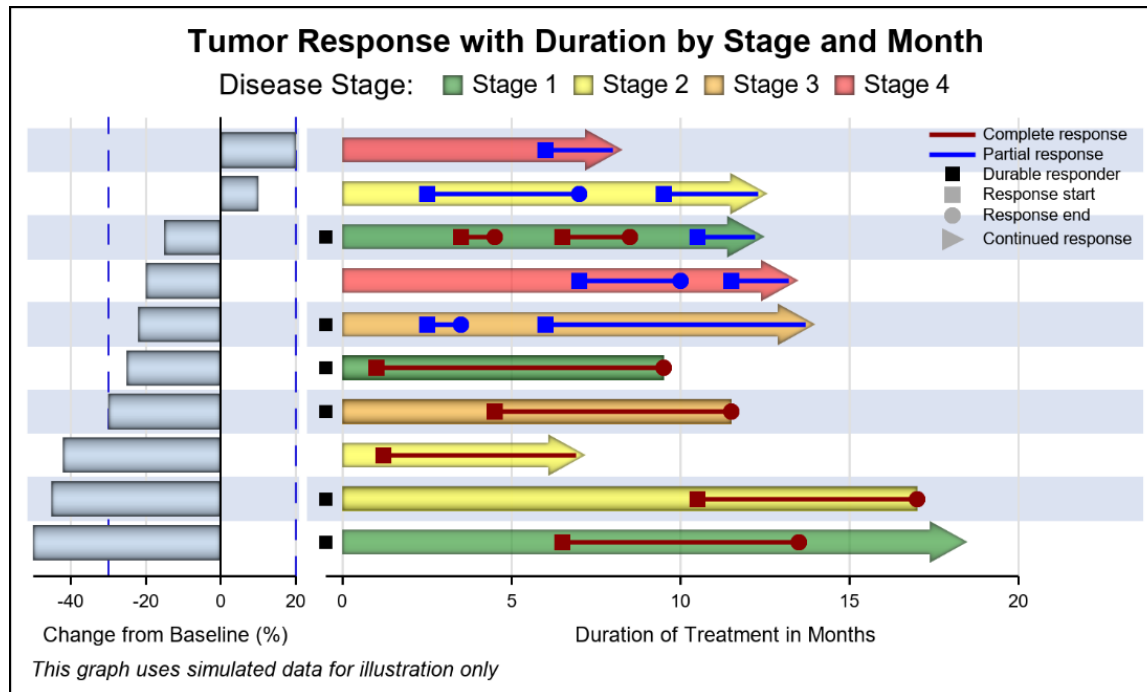


Figure 15 – Combined Waterfall + Swimmer plot in 2-column layout

The template code for Figure 15 is shown in Figure 16.

```
proc template;
  define statgraph Swimmer_Plus_Horizontal;
    begingraph / axislineextent=data;
      entrytitle 'Tumor Response with Duration by Stage and Month';
      entryfootnote halight=left
        'This graph uses simulated data for illustration only' /
        textattrs=(size=7pt style=italic);
      legenditem type=marker name='ResStart' /
        markerattrs=(symbol=squarefilled color=darkgray size=7)
        label='Response start';
      legenditem type=marker name='ResEnd' /
        markerattrs=(symbol=circlefilled color=darkgray size=7)
        label='Response end';
      legenditem type=marker name='RightArrow' /
        markerattrs=(symbol=trianglerightfilled color=darkgray
          size=9) label='Continued response';
      layout lattice / rowdatarange=union columnweights=(0.25 0.75)
        columngutter=0 columns=2;
      rowaxes;
      rowaxis / display=none type=discrete reverse=true
        discreteopts=(colorbands=odd
```

```

                                colorbandsattrs=(transparency=0.1));
endrowaxes;

layout overlay / xaxisopts=(griddisplay=on tickvalueattrs=(size=6)
                                labelattrs=(size=7) walldisplay=none
                                linearopts=(tickvaluepriority=true)
                                label='Change from Baseline (%)');
    bandplot y=id limitupper=20 limitlower=-30 /
                                extend=true display=(outline)
                                outlineattrs=graphdata1(pattern=dash thickness=1);
    barchartparm category=id response=response /
                                barwidth=0.8 orient=horizontal
                                datalabelattrs=(size=5) dataskin=preserved;
endlayout;

layout overlay / xaxisopts=(griddisplay=on offsetmax=0.15
                                tickvalueattrs=(size=6) labelattrs=(size=7)
                                label='Duration of Treatment in Months')
                                walldisplay=none;
    highlowplot y=id low=low high=high / highcap=highcap type=bar
                                group=stage dataskin=preserved lineattrs=(color=black)
                                name='stage' barwidth=1.0
                                includemissinggroup=false datatransparency=0.3;
    highlowplot y=id low=startline high=endline / group=status
                                lineattrs=(thickness=2 pattern=solid)
                                name='status' includemissinggroup=false;
    scatterplot y=id x=durable / markerattrs=(symbol=squarefilled
                                size=6 color=black) name='Durable'
                                legendlabel='Durable responder';
    scatterplot y=id x=start / markerattrs=(symbol=squarefilled
                                size=8) group=status;
    scatterplot y=id x=end / markerattrs=(symbol=circlefilled size=8)
                                group=status;
    discretelegend 'status' 'Durable' 'ResStart' 'ResEnd'
                                'RightArrow' / valueattrs=(size=6)
                                halign=right valign=top border=false
                                location=inside across=1 itemsize=(linelength=20);
endlayout;
sidebar / align=top spacefill=true;
    discretelegend 'stage' / title='Disease Stage:' border=false;
endsidebar;
endlayout;
endgraph;
end;
run;

ods graphics / reset width=5in height=3in
                                imagedata='Swimmer_Plus_Horizontal';
proc sgrender template=Swimmer_Plus_Horizontal
                                data=swimmer_sort_2
                                datrmap=attrmap;
    label response='Response';
    datrvar stage='stage' status='status';
run;

```

Figure 16 – GTL program for horizontal combined Waterfall + Swimmer plot.

CONCLUSION

Data from a study is easier to decode when the related information is displayed together in one graph sorted by a common classifier. 3D visuals have been suggested to display the duration of treatment in the same graph with the tumor response. This paper provides the method to create such 3D graphs using SAS.

This paper also presents alternative 2D visuals that present the same data in a clean and understandable visual with efficient usage of the space available. The duration of treatment information can be added in a separate cell aligned with the tumor response data. Comparisons are easier as we use linear comparisons from common baselines. A third row can be added to display the tumor burden. Additional rows can be added to the graph to display additional related information without clutter.

For an alternate layout for the visual presentation of the data, we can combine the tumor response data with the data about subject history for a Swimmer plot. In this case, we can extend the Swimmer plot with aligned tumor response data either a 2-Row or 2-Column layout.

REFERENCES

Alvarez, et. al. 2017. "3D Waterfall Plots: A better graphical representation of tumor response in oncology" *Annals of Oncology*, March 2017.

Phillips, Stacey 2014. "Swimmer Plot: Tell a Graphical Story of Your Time to Response Data Using PROC SGPLOT" PharmaSUG 2014.

RECOMMENDED READING

3D WaterFall Macro: <https://blogs.sas.com/content/graphicallyspeaking>. Sanjay Matange.

CONTACT INFORMATION

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

Author Name	Sanjay Matange
Address	109 Flying Leaf Ct.
City / Postcode	Cary, NC 27513
Email:	SanjayMatange@gmail.com
Web:	https://make-a-graph.com/

Brand and product names are trademarks of their respective companies.