

A Combined Waterfall and Swimmer Graph

Sanjay Matange, SAS® Institute Inc., Cary, USA

ABSTRACT

In recent months, there has been an increasing interest in combining the “Duration of Treatment” and the “Tumor Response” information for subjects in a study in one graph. Traditionally, this information has been displayed in separate graphs where the subjects may be sorted by different criteria. In such a case, the Clinician has to work harder to associate the subject across the 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 cases which may have their pros and cons. This paper will show you how to build a 3D Waterfall Graph in SAS and will also present effective alternatives that allow easy decoding of the data thus allowing more information to be displayed in the graph.

INTRODUCTION

In the paper “3D Waterfall Plots: A better graphical representation of tumor response in oncology” published in Annals of Oncology, March 2017, 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 are often complemented by separate swimmer plots and spider plots. The authors indicate that readers often need to consult different charts to get a complete understanding of the trial results. The authors propose to add the display of treatment duration into a 3D waterfall plot that provides extra information that is useful to the reader.

Figure 1 shows a traditional water fall graph and Figure 2 shows a 3D waterfall chart where the tumor response is shown in the front vertical plane of the graph and the associated duration information for each subject is shown in the horizontal plane.

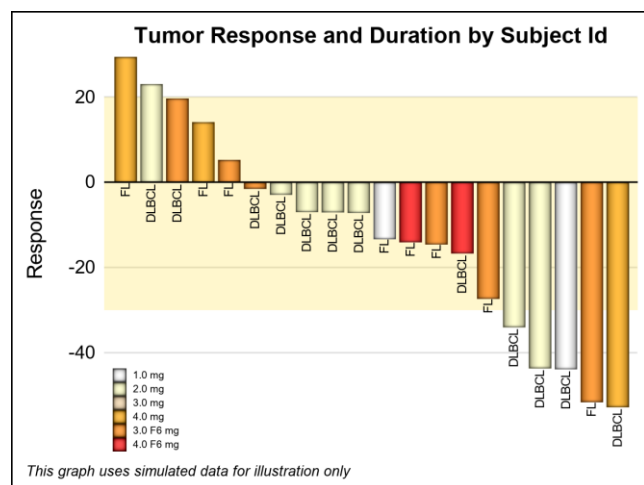


Figure 1 – Traditional Waterfall plot

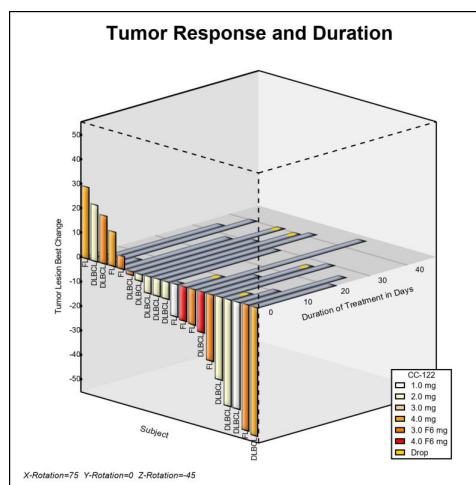


Figure 2 – 3D Waterfall plot

Figure 1 shows the tumor response in a study sorted by increasing response. 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. Note, the data for this graph is simulated for illustration purposes only.

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 horizontal plane of the graph. Note, the data for this graph is 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, whether subject was dropped from study. Codeloc is used to display the code value and baseline in not used in this graph.

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 and requires the following parameters:

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

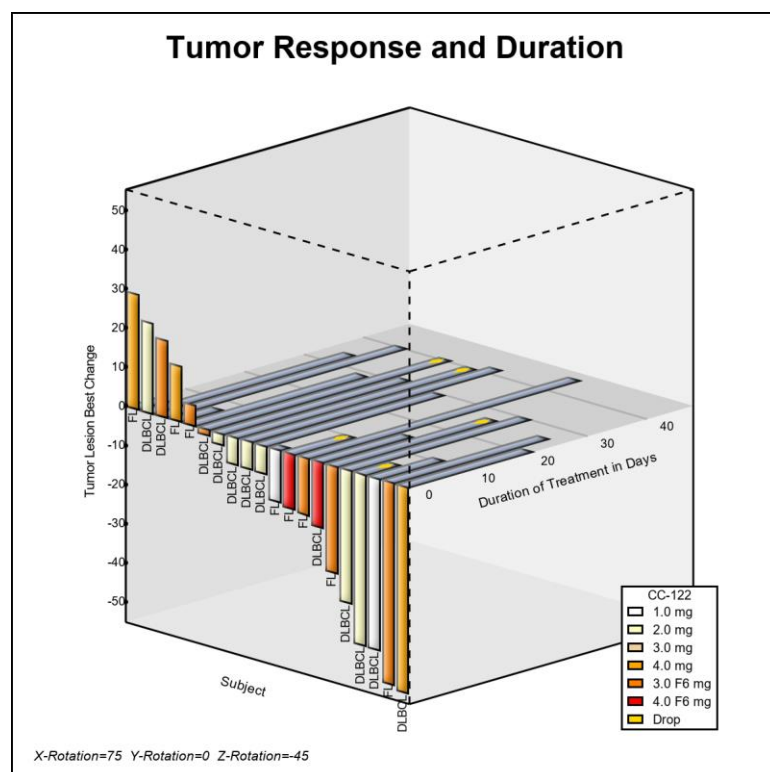


Figure 5 – 3D Waterfall graph

PhUSE EU Connect 2018

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 bar data is coded as polygons in 3D coordinate space.
- 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.

While the 3D graph shown in Figure 5 has a modern and slick appearance, there are some shortcomings of this visualization of the data.

- 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 is not easy.
- This data is really 2D in nature with one independent variable (the sorted subject id). All response, duration and dropped information are dependent only the subject id.

The only 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.

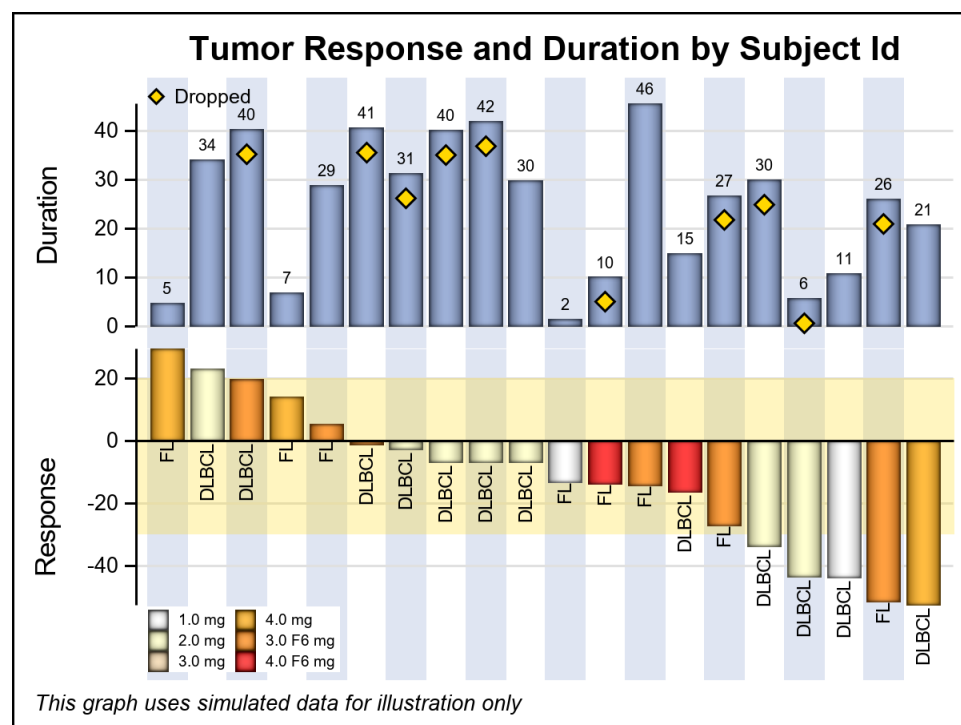


Figure 6 – 2D Waterfall Graph with duration data.

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 for each subject is shown in the upper cell, clearly aligned with the bars in the lower cell.
- The duration value can be easily displayed at the top of each bar.

PhUSE EU Connect 2018

- The “dropped” markers are clearly visible.
- Vertical blue bands provide clear “connection” between the bars for each subjected.
- 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-cell Lattice using GTL.

```
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=preserved displaybaseline=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=preserved
          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

2D WATERFALL GRAPH WITH MORE DATA

Already there have been calls to add more data displays to this graph. One such case was the need to add the display of the tumor burden for each subject in the same graph. Since once again there is only one independent variable (subject id), it is easy to extend this 2D representation to accommodate more data as shown in Figure 8.

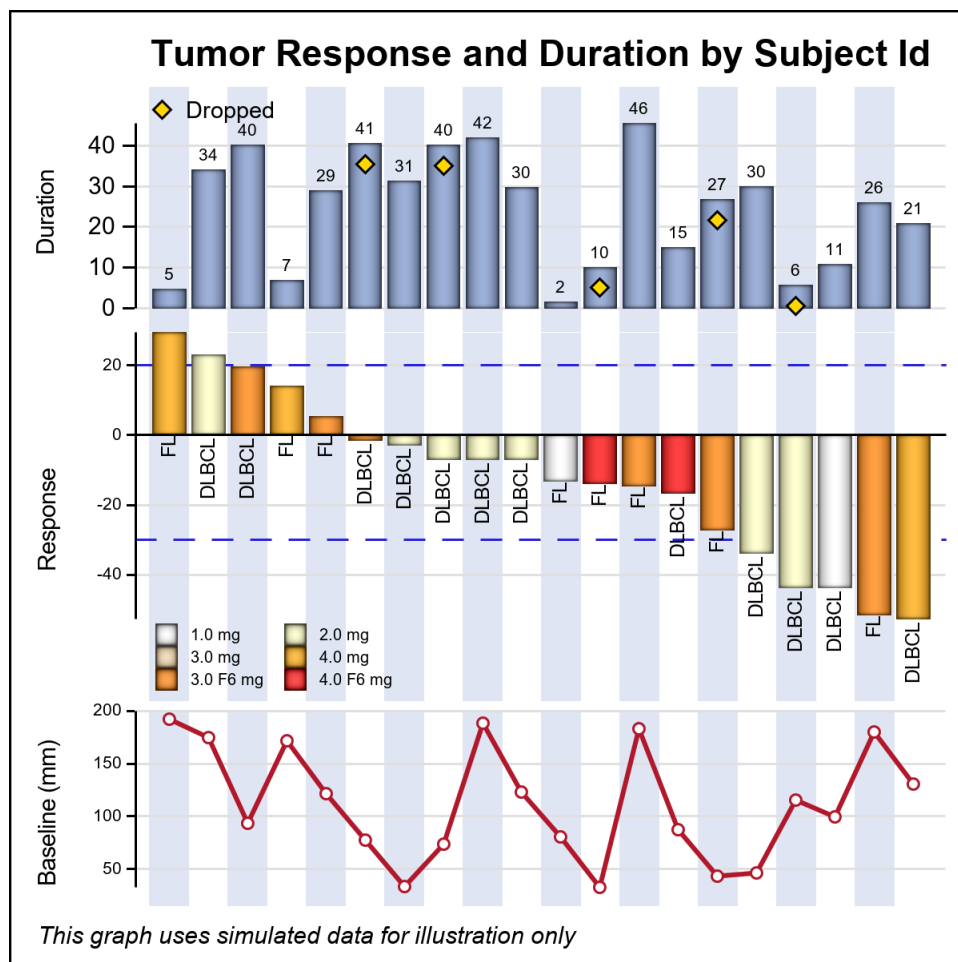


Figure 8 – GTL program for 2D Waterfall with Duration and Baseline data

Figure 8 extends the previous graph 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.

WATERFALL GRAPH WITH SWIMMER 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. Now, let us see how we might start with the data for creating a Swimmer Plot and extend it with tumor response information to create a combined Waterfall + Swimmer plot.

Figure 9 shows the 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.

PhUSE EU Connect 2018

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 9 – Data for Swimmer Plot with added columns for tumor response.

The graph in Figure 10 shows a Swimmer Plot created from the data in Figure 9 using the SGPLOT procedure.

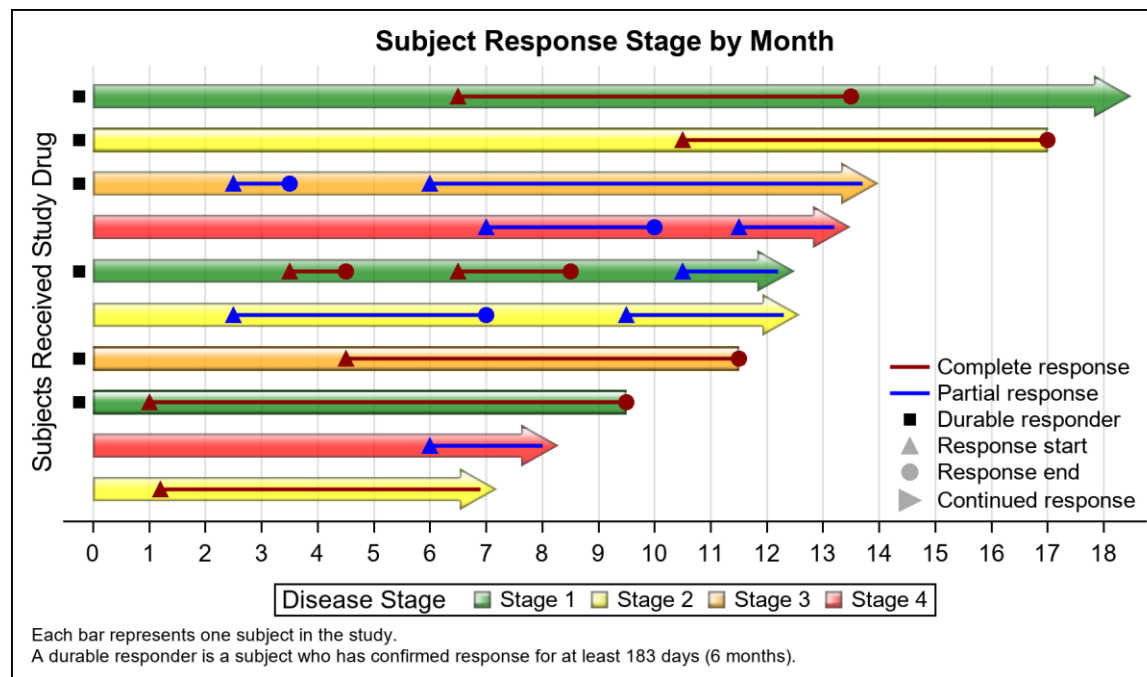


Figure 10 – Swimmer plot of subjects receiving treatment by drug.

The graph in Figure 10 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.
- For each subject, the 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.

PhUSE EU Connect 2018

The SGPLOT Procedure program for this Swimmer graph is shown in Figure 11.

```

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 / markerattrs=(symbol=squarefilled size=6 color=black)
    name='Durable' legendlabel='Durable responder';
  scatter y=item x=start / markerattrs=(symbol=trianglefilled size=8) group=status
    attrid=status;
  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 11 – SGPLOT Procedure code for Swimmer Plot.

Since we added tumor response data for each subject in the data shown in Figure 9, we can create a traditional waterfall graph as shown in Figure 12.

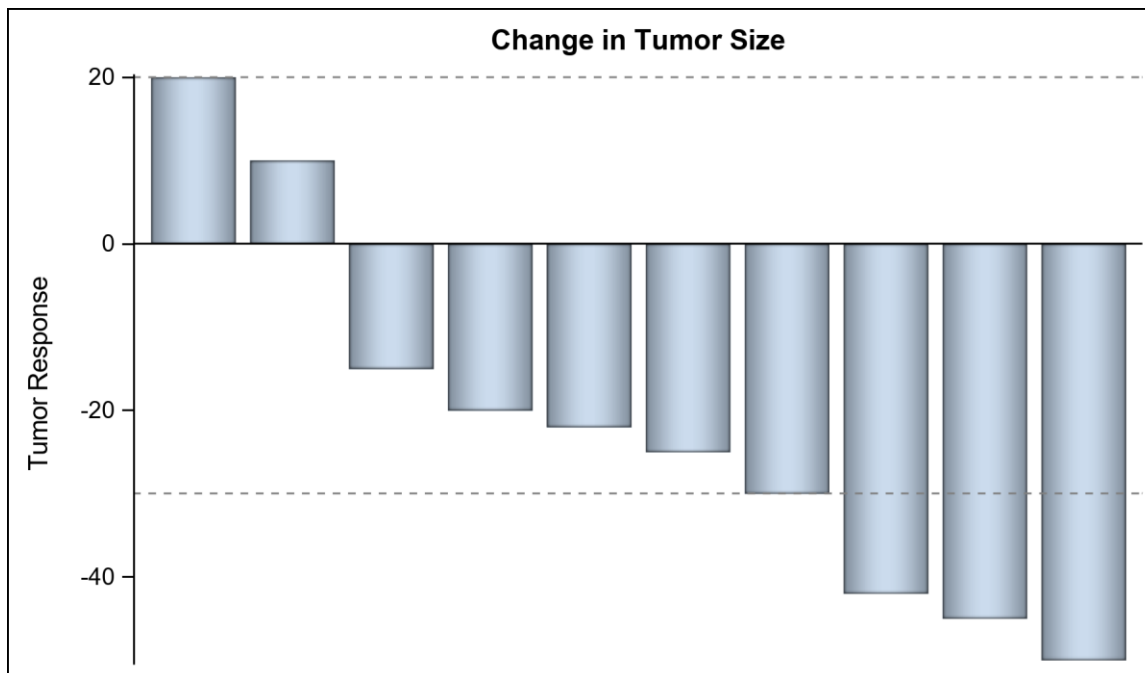


Figure 12 – Waterfall plot from data in Figure 9

PhUSE EU Connect 2018

The SGPLOT Procedure program for this Waterfall graph is shown in Figure 13.

```
title 'Change in Tumor Size';
proc sgplot data= swimmer_sort_2 noborder;
  vbarparm category=id response=response / dataskin=preserved;
  refline 20 -30 / lineattrs=(pattern=shortdash) ;
  xaxis display=none;
  yaxis label='Tumor Response';
run;
```

Figure 13 – SGPLOT Procedure code for Swimmer Plot.

Now, we can create a combined Waterfall + Swimmer plot from the data shown in Figure 9 as shown in Figure 14.

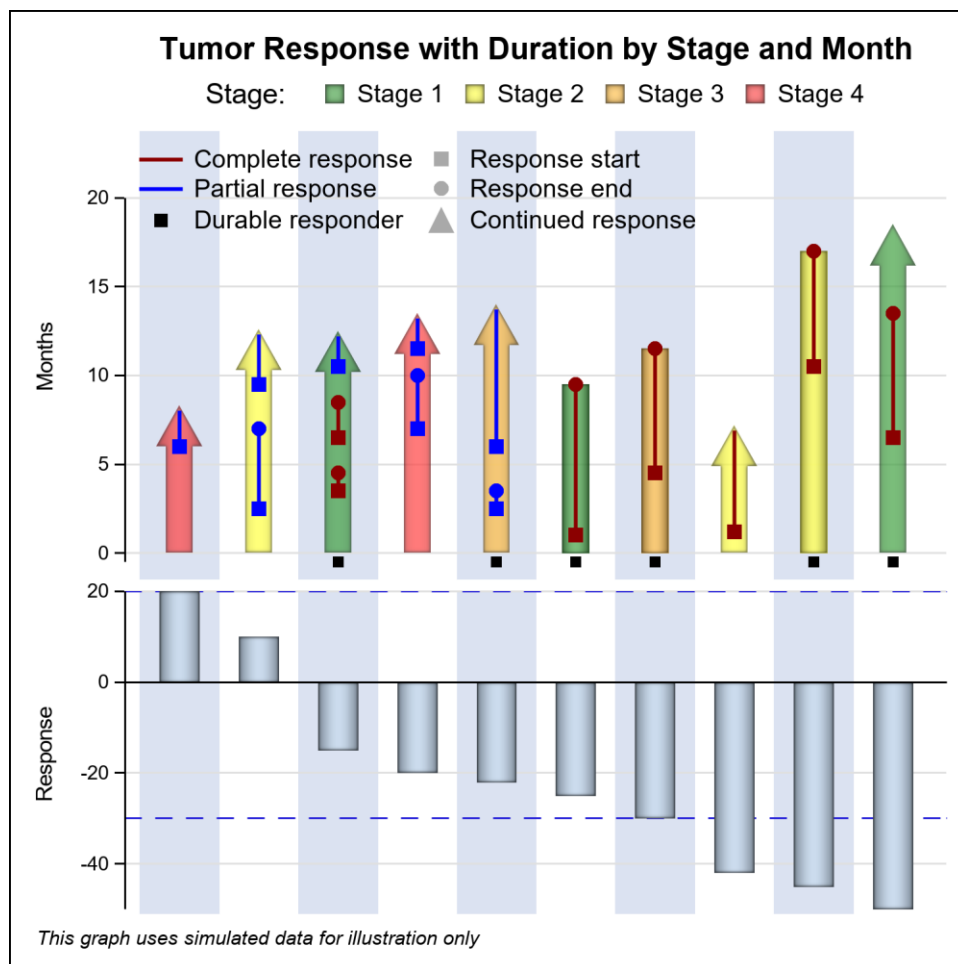


Figure 14 – Combined Waterfall + Swimmer plot from data shown in Figure 9

Let us review the details of this graph.

- The tumor response is shown in the lower cell, very similar to the graph in Figure 12.
- 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 10.
- The subject response 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.

PhUSE EU Connect 2018

```

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=presse 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=graphdata1(pattern=dash thickness=1);
          barchartparm category=id response=response / barwidth=0.5
            datalabelattrs=(size=5) dataskin=presse;
        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 dattmap=attrmap;
  label response='Response';
  dattvar stage='stage' status='status';
run;

```

Figure 15 – GTL program for combined Waterfall + Swimmer plot.

CONCLUSION

There is an increasing interest in plotting tumor response and duration of treatment data together in one graph so it is easier for the reader of the graph to understand the information. To view duration of treatment together with tumor response, 3D visuals have been suggested. These 3D visuals of 2D data are hard to read and make inefficient use of the space available in the graph. Adding more data makes these graphs more cluttered.

This paper presents alternative 2D visuals that present the same data in a clean and understandable visual with better usage of the space available. Comparisons are easier as we use linear comparisons from common baselines. Information can be easily aligned and additional information can be displayed without clutter.

Swimmer plots present more information regarding a subject's treatment history. This data can be combined with tumor response data and plotted together as a combined Waterfall + Swimmer plot for better understanding of the data.

REFERENCES

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

RECOMMENDED READING

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

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)

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

Author Name	Sanjay Matange
Company	SAS Institute Inc.
Address	SAS Campus Dr.
City / Postcode	Cary, NC 27513
Work Phone:	
Fax:	
Email:	Sanjay.Matange@sas.com
Web:	https://blogs.sas.com/content/graphicallyspeaking

Brand and product names are trademarks of their respective companies.