
# Effective Graphical Representation of Tumor Data

## Sanjay Matange

Sanjay Matange is an expert in visualization of data using SAS graphics software including the SG procedures and GTL. Sanjay worked at SAS for about 30 years.

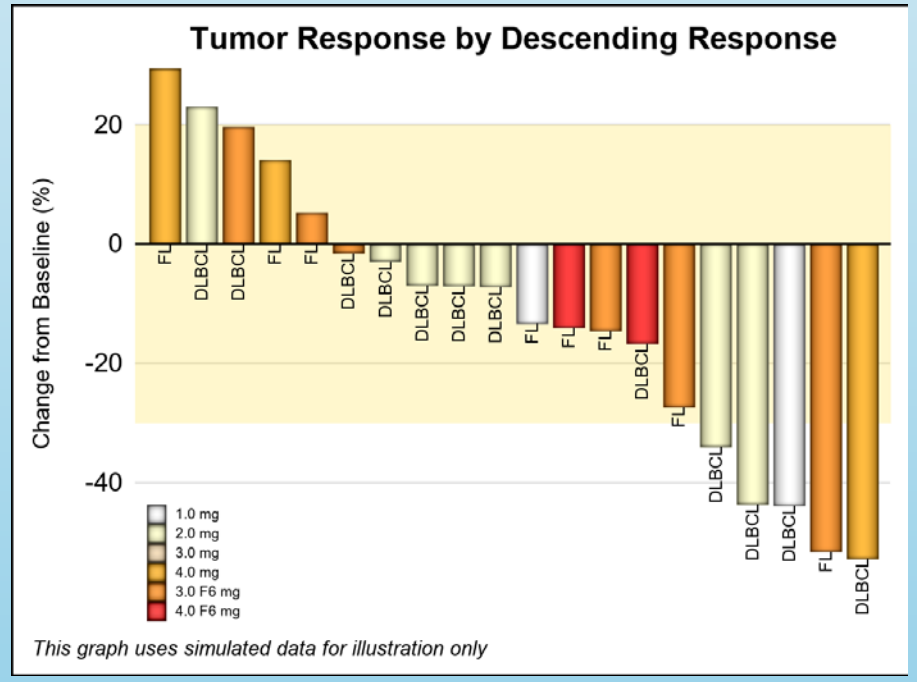
Sanjay is co-author of four patents and the author of four SAS Press graphics books.

Sanjay was the main author of Graphically Speaking SAS blog for 8 years, a blogs on data visualization using ODS Graphics.

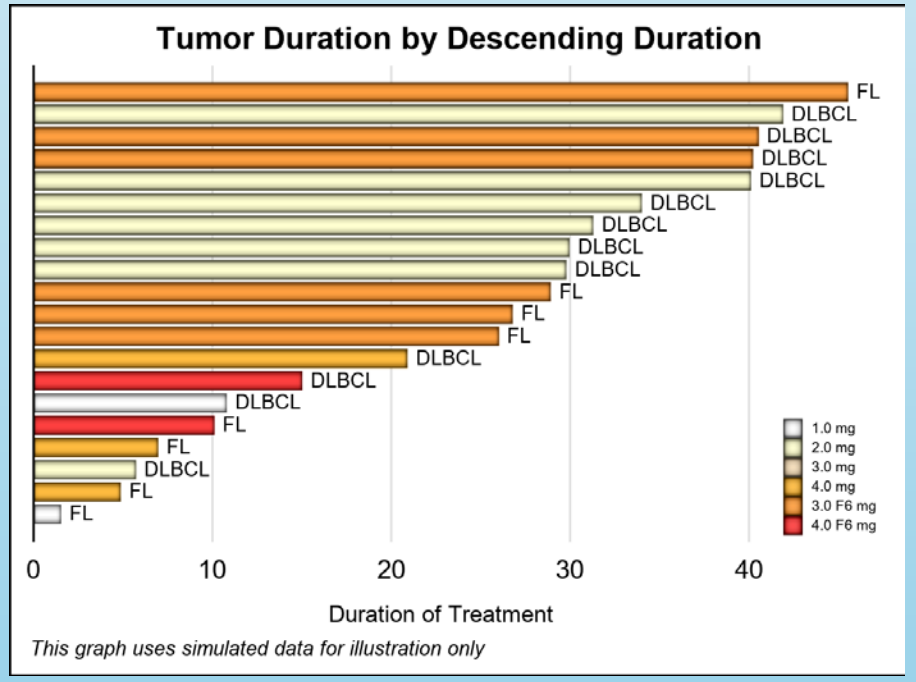
# Tumor Graphs

- Traditionally, there are two popular ways to visualize tumor data.

Waterfall Chart



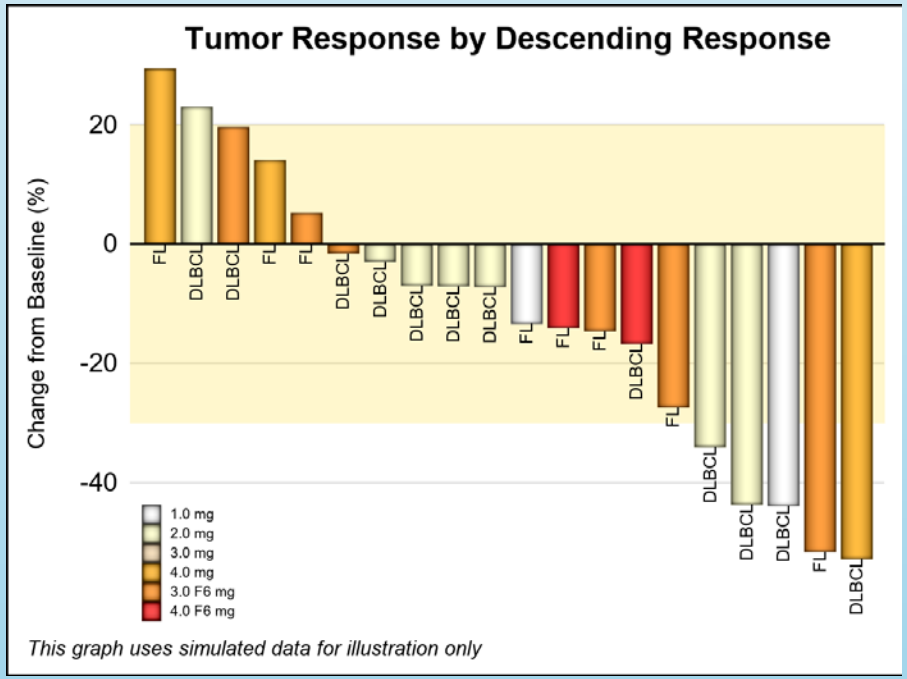

Swimmer Plot



# The Waterfall Chart

- The Tumor Response for each subject in the study is displayed in the Waterfall Chart.
- The data is sorted by increasing response (decrease in tumor size) on the x-axis.
- The percent change in tumor size is displayed on the y-axis.
- Additional data may be shown for each subject such as type of response (RECIST), medication dosage or tumor type using colors or labels.
- A -30% horizontal band is displayed for PR.
- The goal is to show the study population's result.
- The individual subject id is often not displayed on the graph.

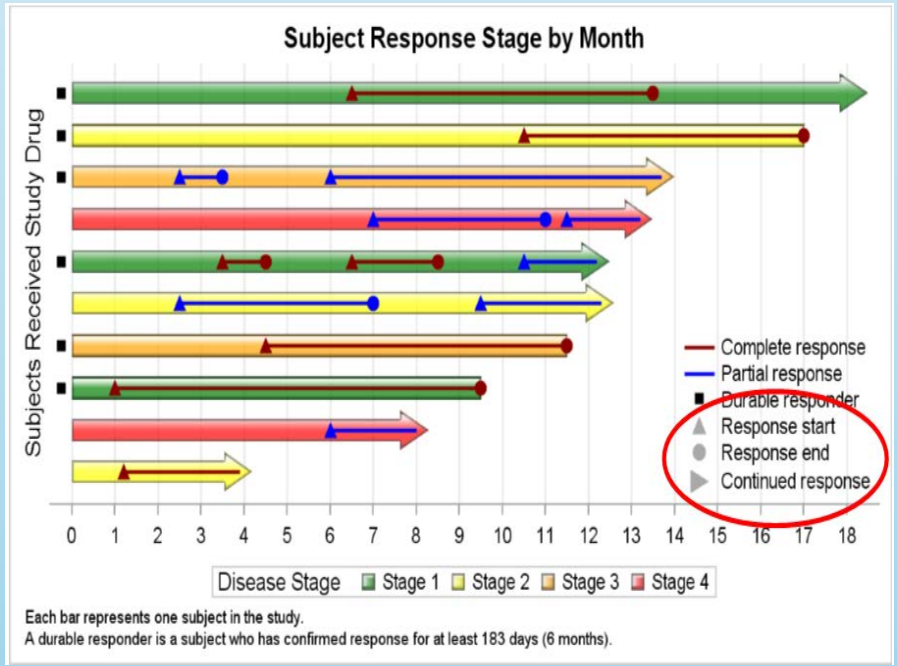
Waterfall Chart



# The Swimmer Plot

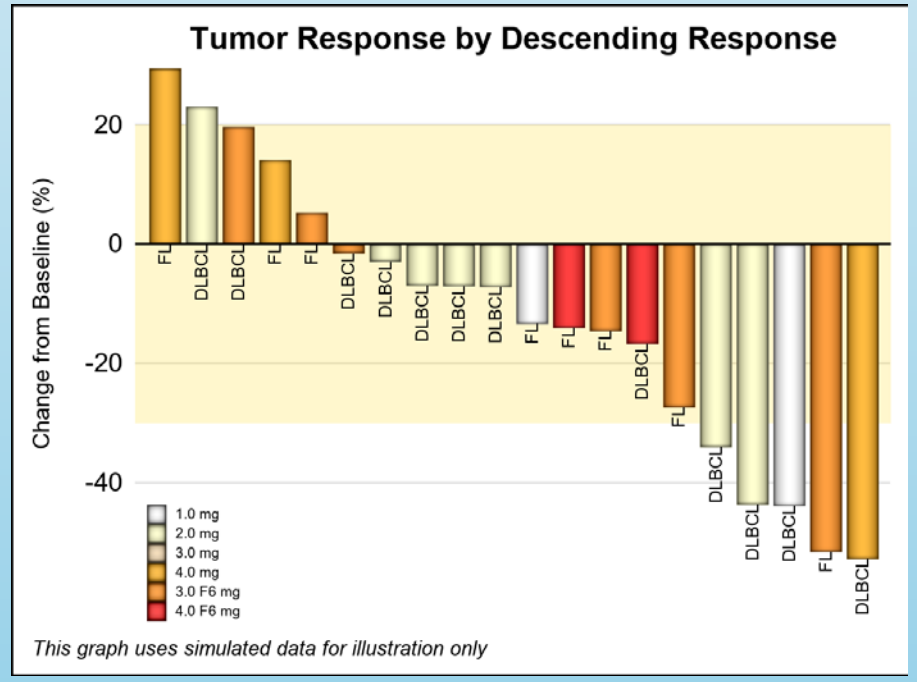
- Duration of treatment for each subject is displayed on the x-axis as a horizontal bar colored by disease stage.
- The data is sorted by duration of treatment on the y-axis.
- Additional data may be shown for each subject such duration of each response, continued response and durable responder.
- The goal is to show the treatment history of each subject.
- The individual subject id is often not displayed on the graph.

Swimmer Plot

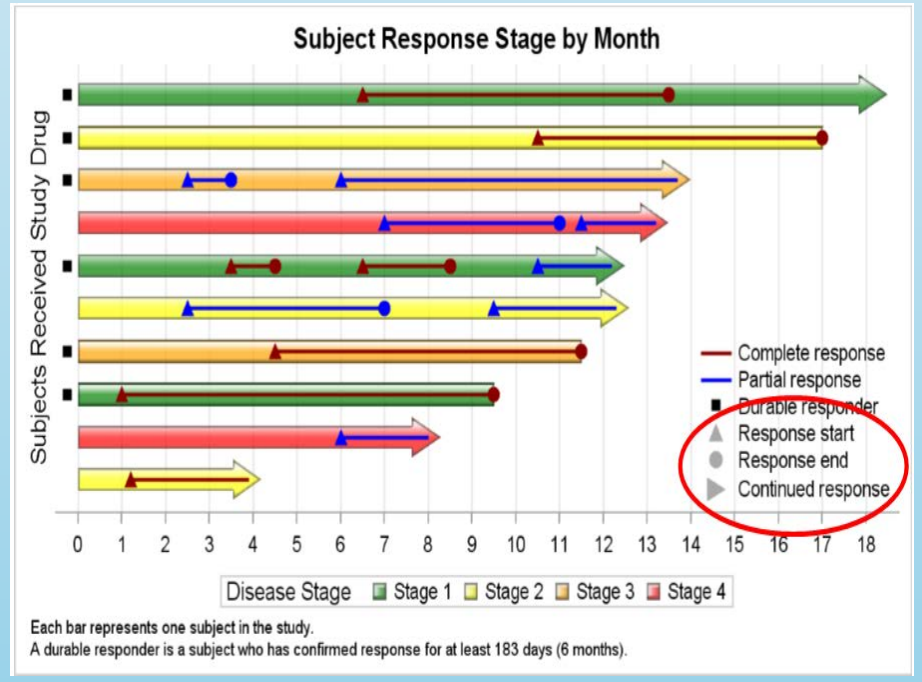


# Understanding the data

- These two graphs show different data sorted by different criteria.
- Subject Id is not shown, so it is harder for investigators to understand the data.
- There is a desire to create graphs that include more data in one visual.




Waterfall Plot

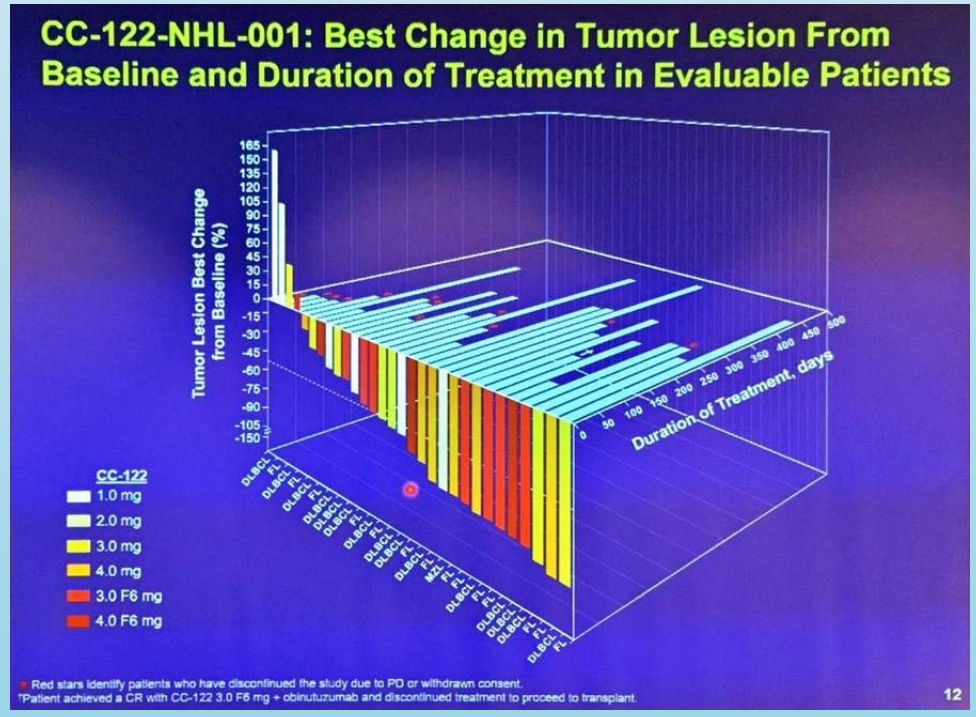



Swimmer Plot

# 3D Waterfall Plot

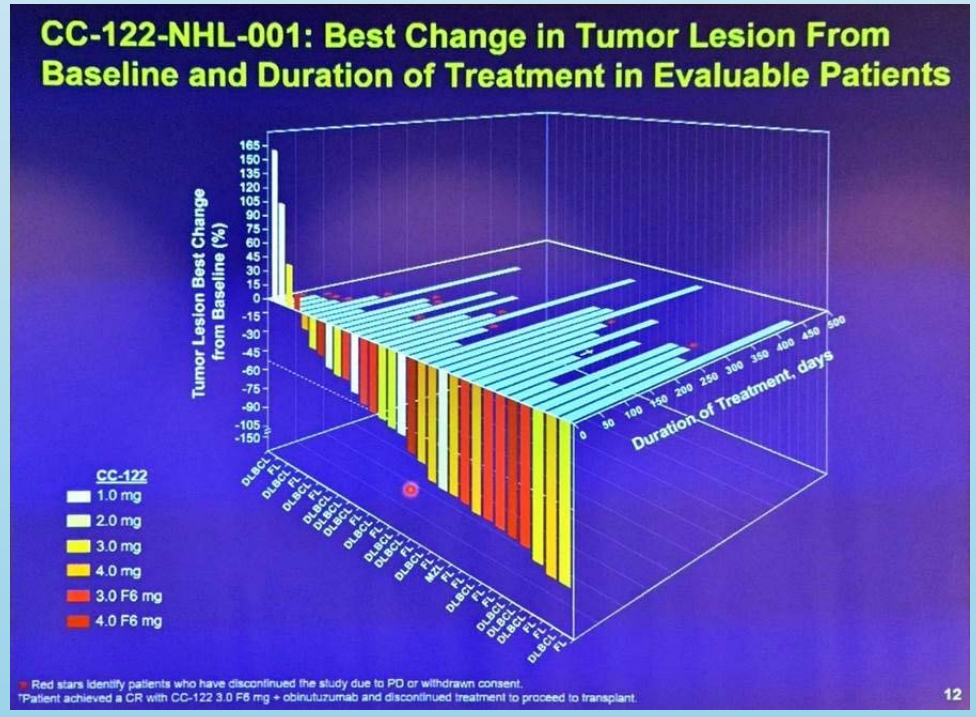
- In recent publications, 3D graphs have been proposed that display more data in one graph.

- Tumor Response data is displayed in the front plane by treatment and sorted in order of increasing response.
- Duration of treatment is displayed in the horizontal plane for the each subject.
- Additional information is displayed at the bottom for each subject.
- Additional information is also displayed on the duration bars.



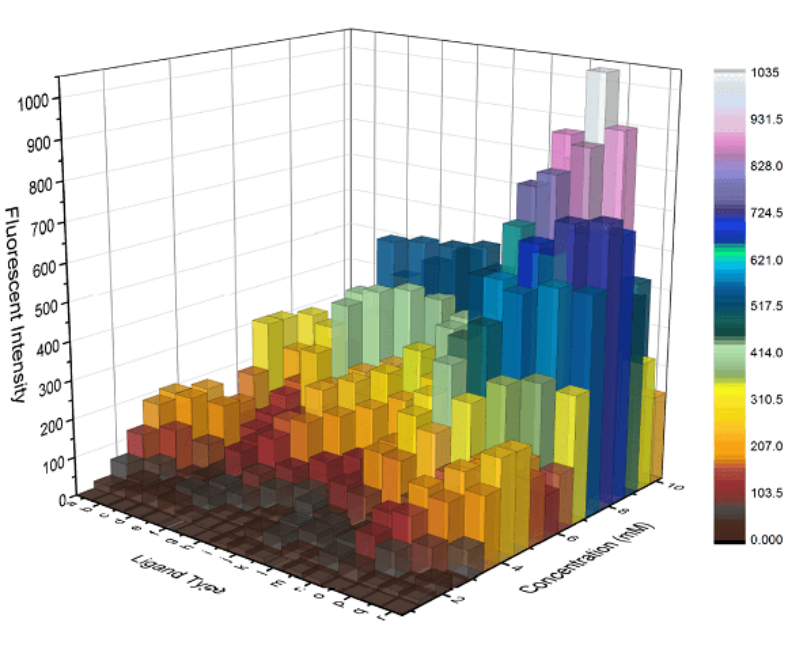
# 3D Waterfall Plot

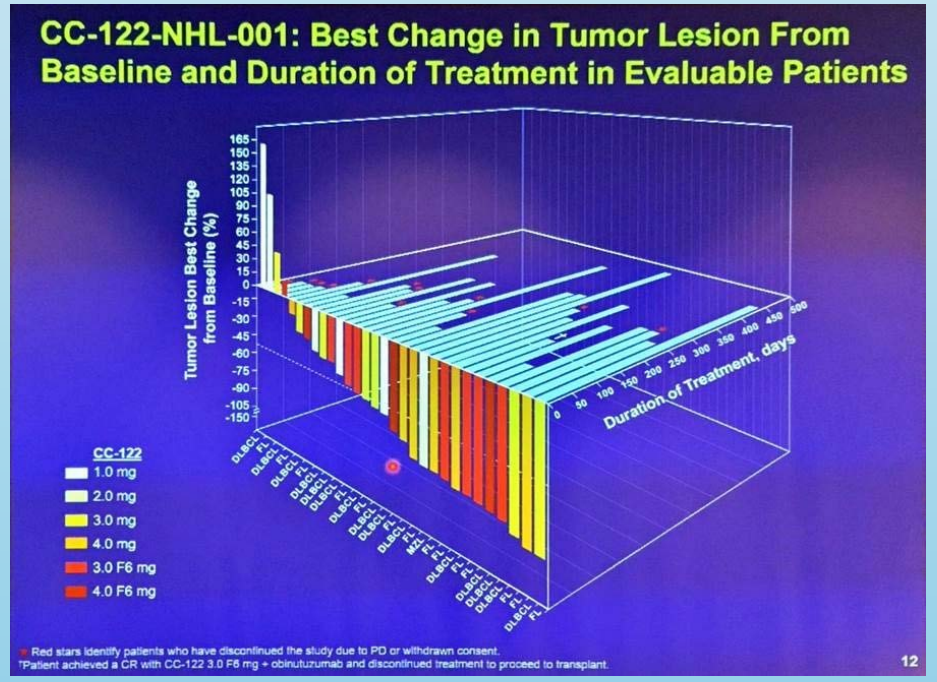
- The main benefit of this layout is the display of both tumor response and duration of treatment data in one graph.
- There are some concerns with this layout.
  - The perspective projection makes it harder to do precise comparisons of bar heights.
  - Some duration bars may be occluded by the tumor response bars.
  - The tumor type at the bottom may be harder to align with the subjects.
  - The markers on the duration bars may be harder to see.
  - The data pixel density is very sparse.



# Traditional 3D Bar Charts.

- Traditional 3D Bar Charts have two independent axes at the base, with one response (height of the bar).
- This 3D Tumor Graph is non-traditional, with one independent x-axis and two dependent responses, the tumor response (y) and duration of treatment (z).

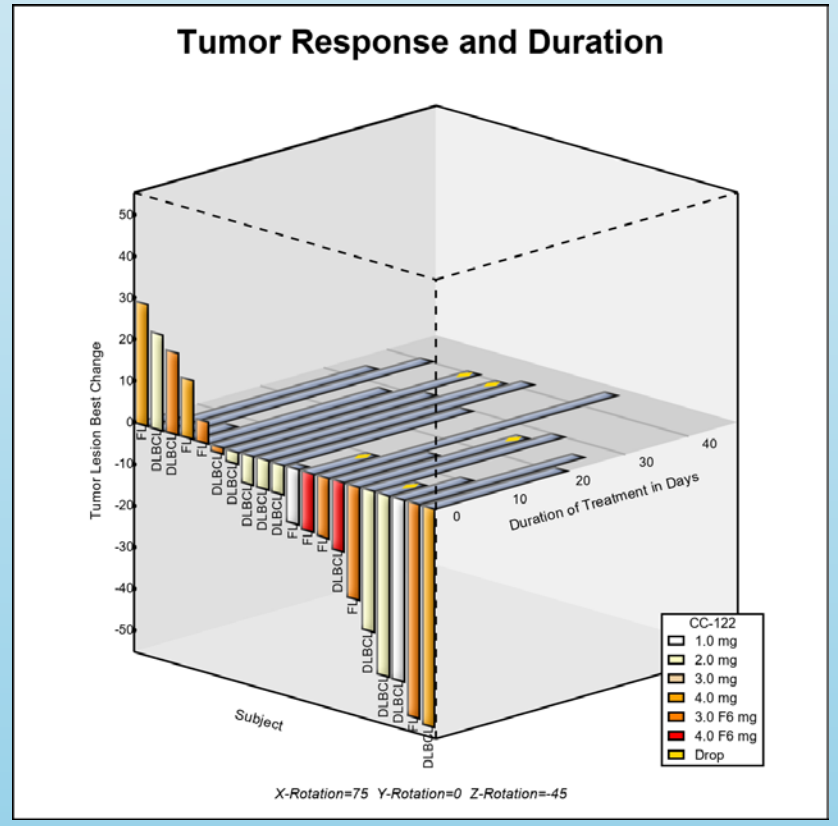


# 3D Waterfall Plot using SAS

- There is no SAS procedure that can create this graph automatically.
- Many users have requested to create such a graph using SAS.
- I have published a macro to render such a 3D graph using the SGPLOT procedure.

  *%WaterFall_3D_Macro (Data=tumor, Duration=duration, Response=response, Dropped=dropped, Group=Drug, Code=Code);*
- See: https://blogs.sas.com/content/graphicallyspeaking/2018/04/24/3d-waterfall-chart-redux/

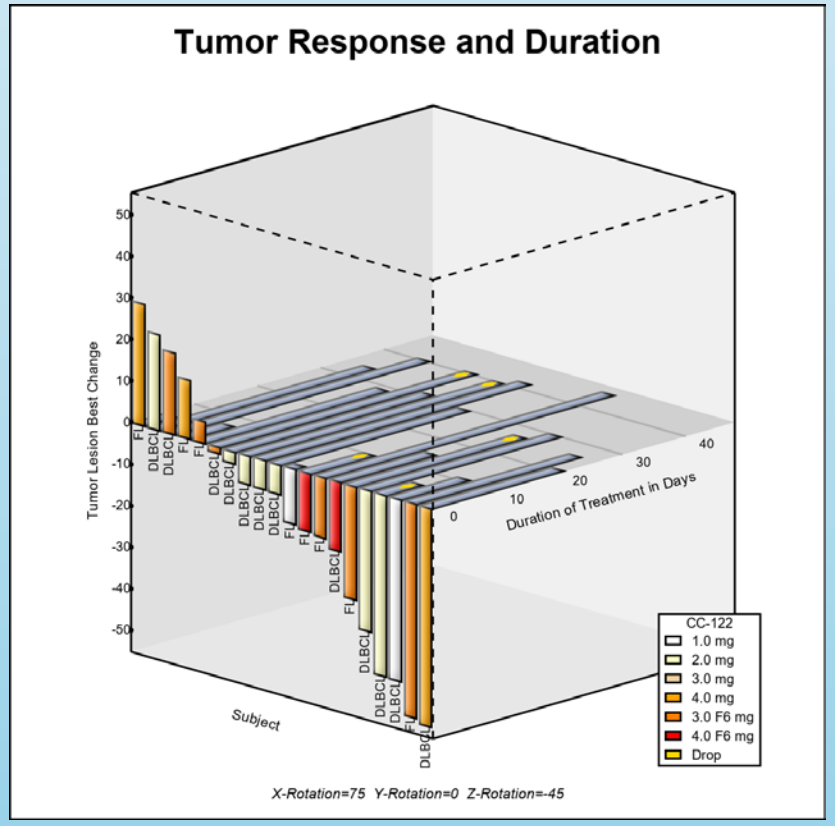


# 3D Waterfall Plot using SAS

- This graph addresses some concerns from the previous version.

- An Orthographic projection is used instead of perspective. This avoids distortion of the data.
- The tumor type is displayed close to the base of the bars for easier association.

- However, later we will see other ways to display the data together as desired using a simpler 2D layout.



# Data for Combined Graph

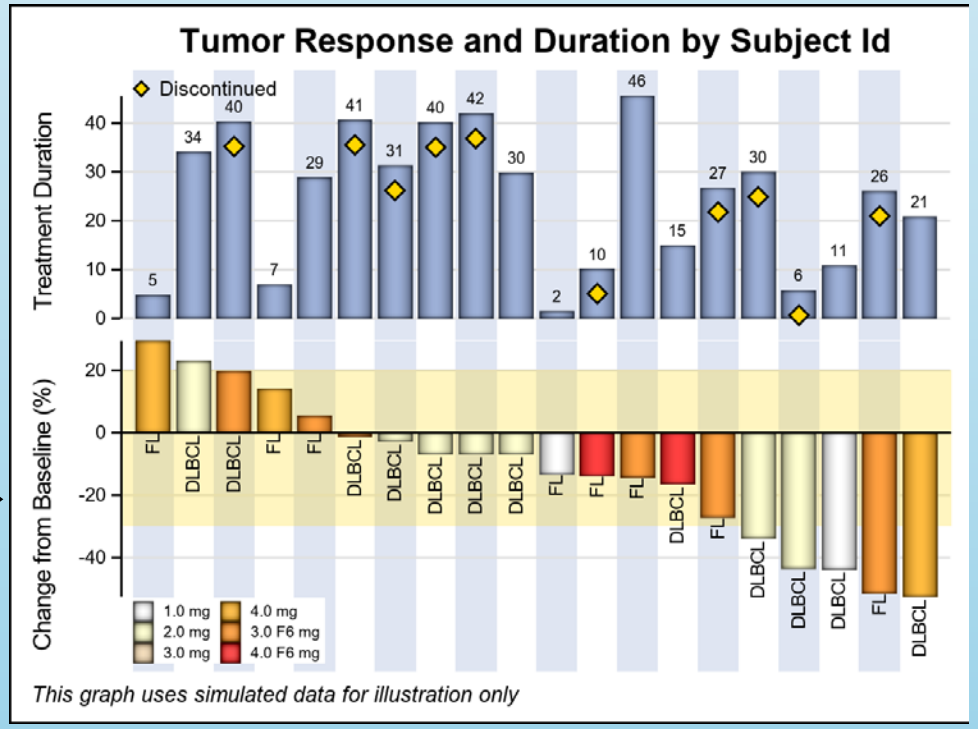
- To create the 3D graph, we start with the data used for a traditional Waterfall plot.
- To that data, we add the column for Duration of Treatment.
- A column is added for display of baseline tumor load.
- Additional data can easily be displayed in a 2D layout.

| 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 |
| 6 | 6 | 3.0 F6 mg | DLBCL | -1.4493 | 40.5163 | 35.5163 | -1.4493 | 77.174 |
| 7 | 7 | 2.0 mg | DLBCL | -2.8468 | 31.2564 | . | -2.8468 | 33.736 |
| 8 | 8 | 2.0 mg | DLBCL | -6.9270 | 40.0830 | 35.0830 | -6.9270 | 73.605 |
| 9 | 9 | 2.0 mg | DLBCL | -6.9796 | 41.8860 | . | -6.9796 | 188.649 |
| 10 | 10 | 2.0 mg | DLBCL | -7.1062 | 29.7538 | . | -7.1062 | 122.890 |
| 11 | 11 | 1.0 mg | FL | -13.2853 | 1.5048 | . | -13.2853 | 80.836 |
| 12 | 12 | 4.0 F6 mg | FL | -13.9817 | 10.0777 | 5.0777 | -13.9817 | 32.562 |
| 13 | 13 | 3.0 F6 mg | FL | -14.5112 | 45.5133 | . | -14.5112 | 183.410 |
| 14 | 14 | 4.0 F6 mg | DLBCL | -16.6046 | 14.9807 | . | -16.6046 | 87.256 |
| 15 | 15 | 3.0 F6 mg | FL | -27.2632 | 26.7480 | 21.7480 | -27.2632 | 43.587 |

# Combined 2D Graph

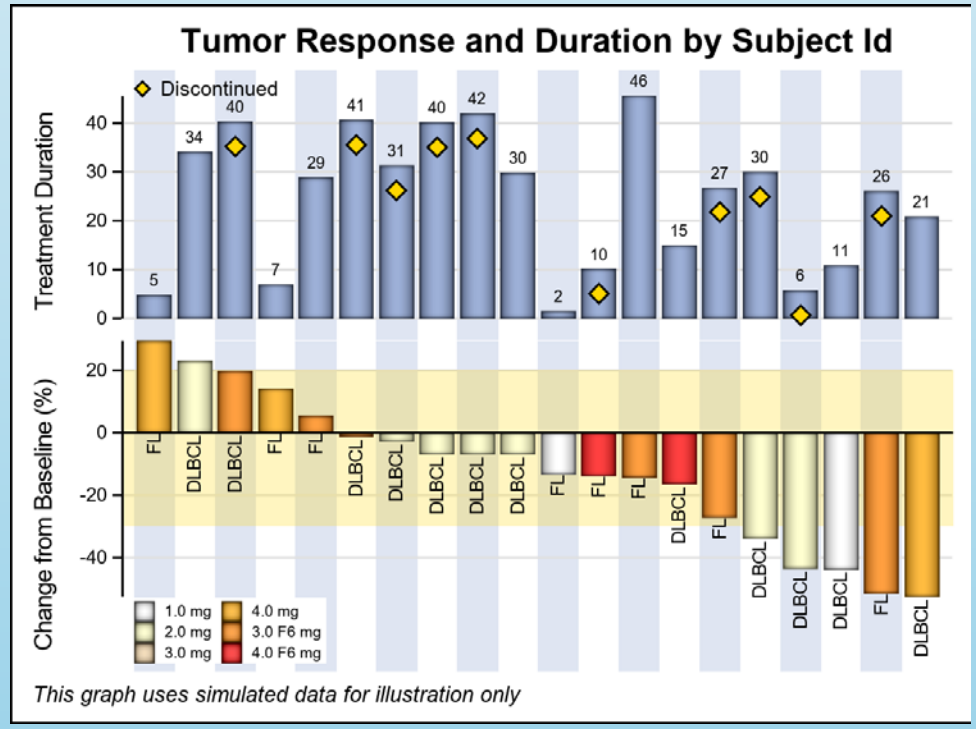
Here is an effective display of the same data in 2D

- This graph uses a two-row GTL layout.
- The bottom row displays the traditional Waterfall plot of tumor response by subject sorted by increasing response (reducing size).
- The bars are classified by treatment.
- The tumor type is displayed at the bottom of the bar.



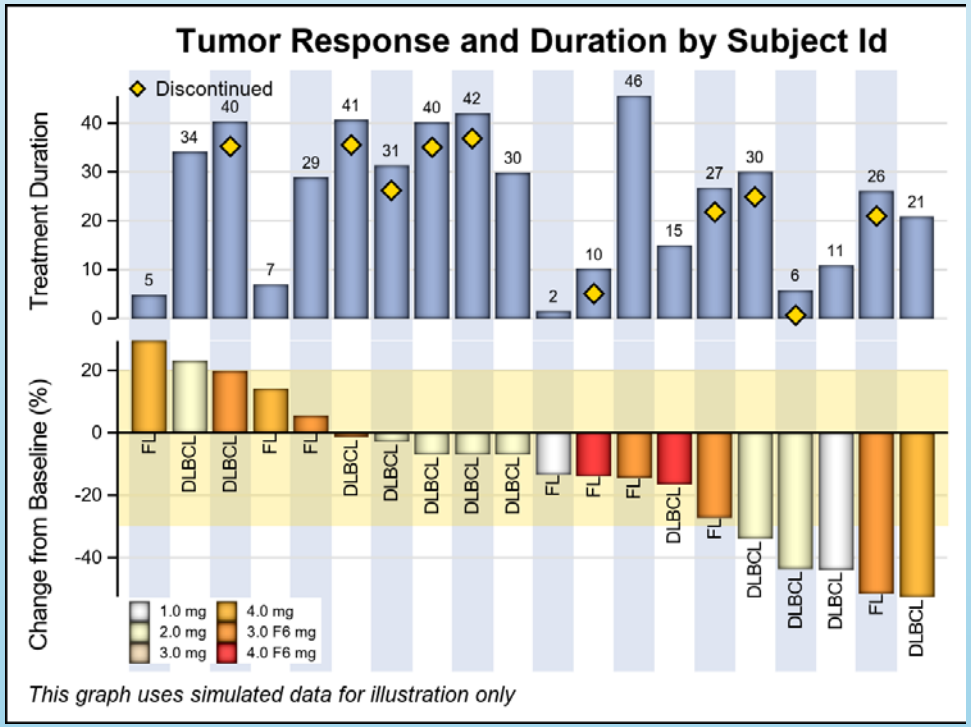
# Combined 2D Graph

- The top row displays the duration of treatment data for each subject.
- The duration value is displayed at the top of each bar.
- A marker is displayed for subjects that have discontinued.
- Alternate blue bands are help line up the tumor response and duration bars.



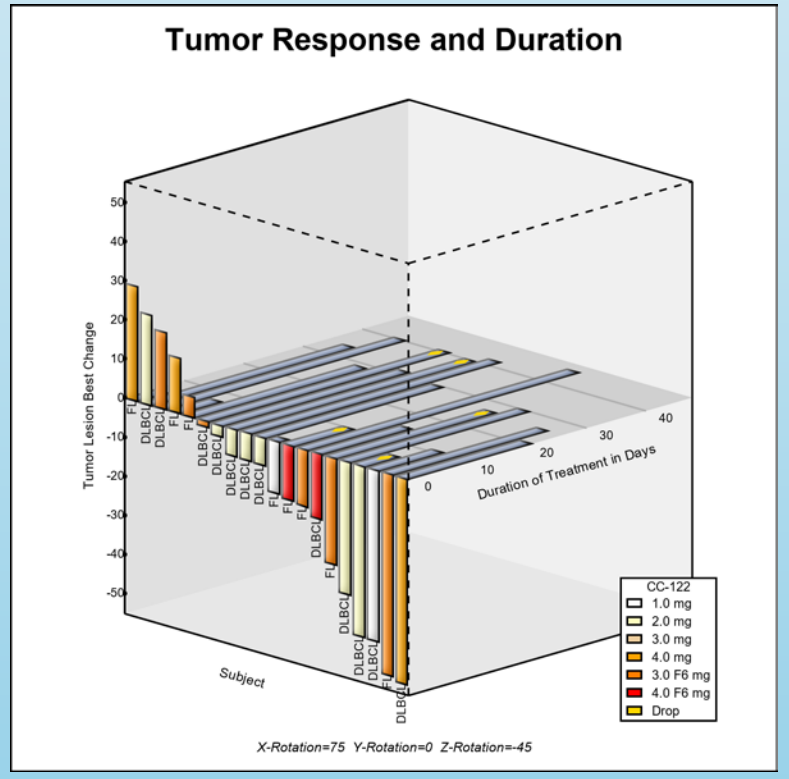
# Benefits of the 2D Layout

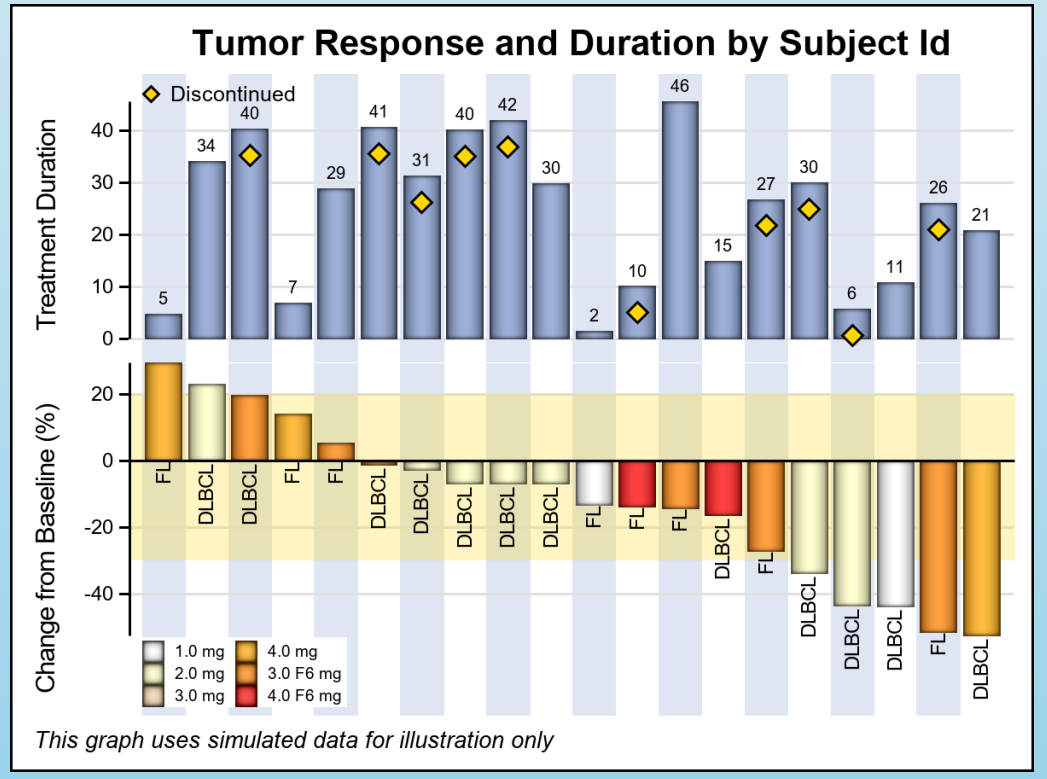
- Whenever possible, a 2D visual is easer to understand.
- Here, the tumor response and duration of treatment are vertically aligned.
- Alternating blue bands further help in alignment of the bars.
- 2D bars from a common baseline are widely accepted as the best visual for comparison of magnitude.
- See the paper for the full code.



# Comparison of Layouts

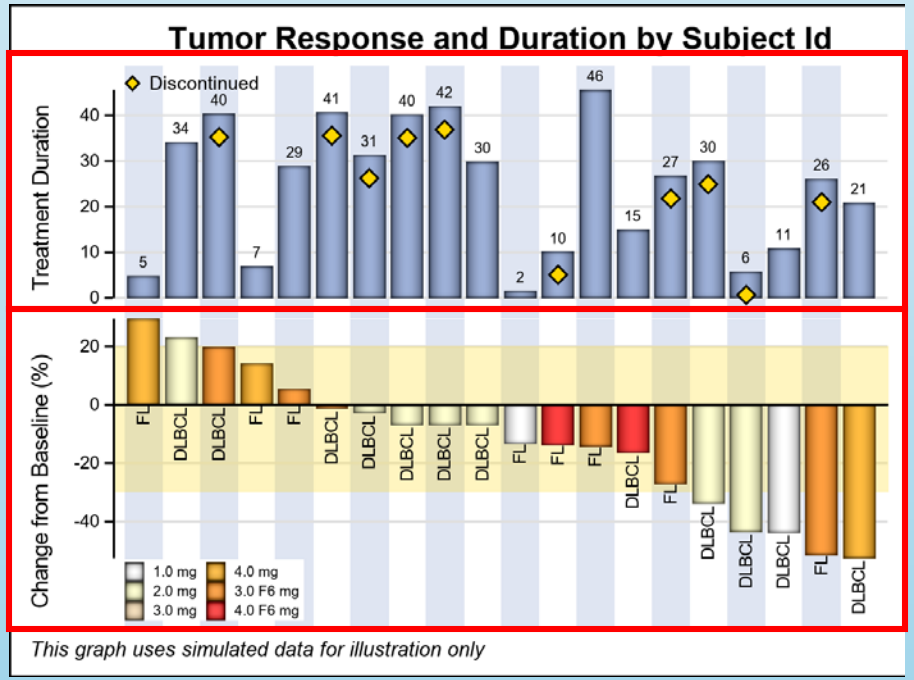
Here is a side-by-side comparison of the two visuals.



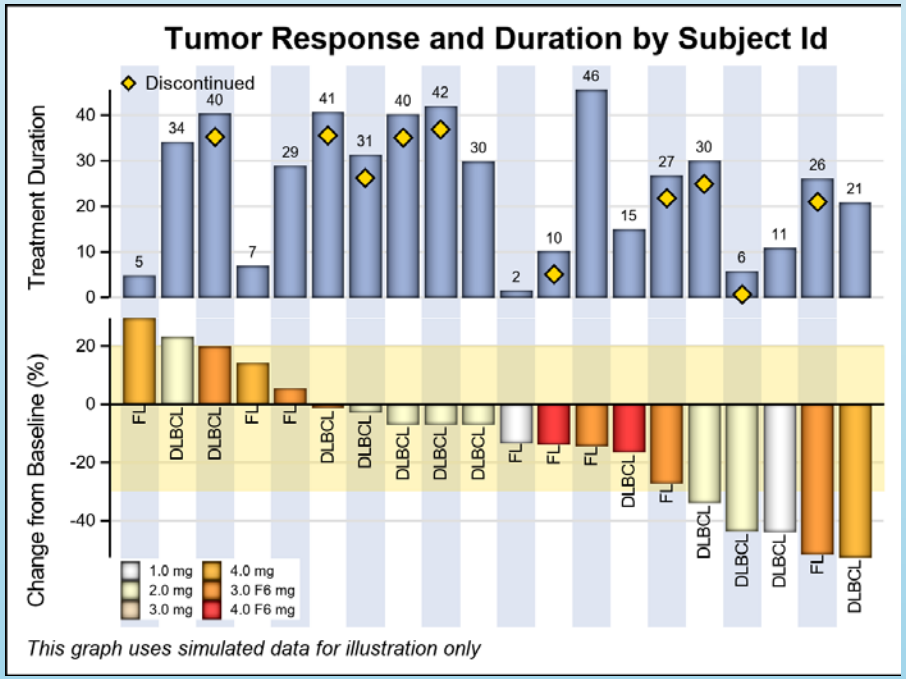
# Structure of this Graph

- This graph has a two-row layout.
- To make this graph, we have used the Graph Template Language (GTL).
- The GTL Layout Lattice syntax allows building of multi-row graphs.
- Each column of cells  can have separate multi-layered plots with uniform x-axes.

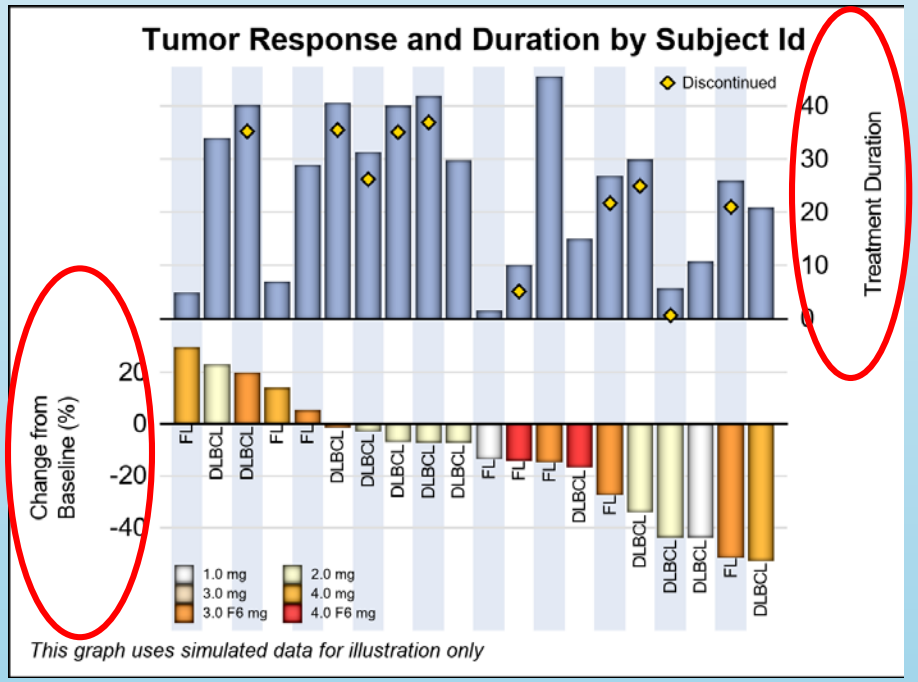


# Comparison of Layouts

This graph can be made using SGPLOT, but with limits.




Using GTL

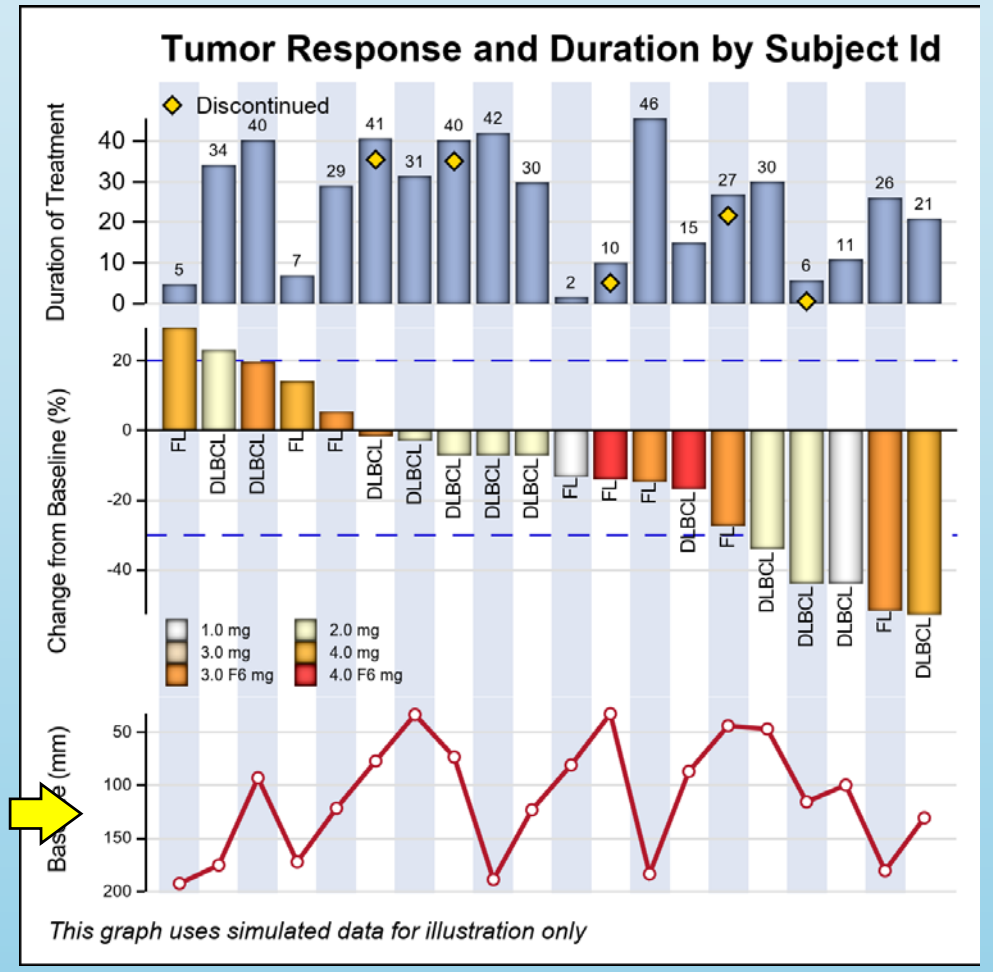



Using SGPLOT

https://blogs.sas.com/content/graphicallyspeaking/2017/07/30/clinical-graphs-waterfall-plot/

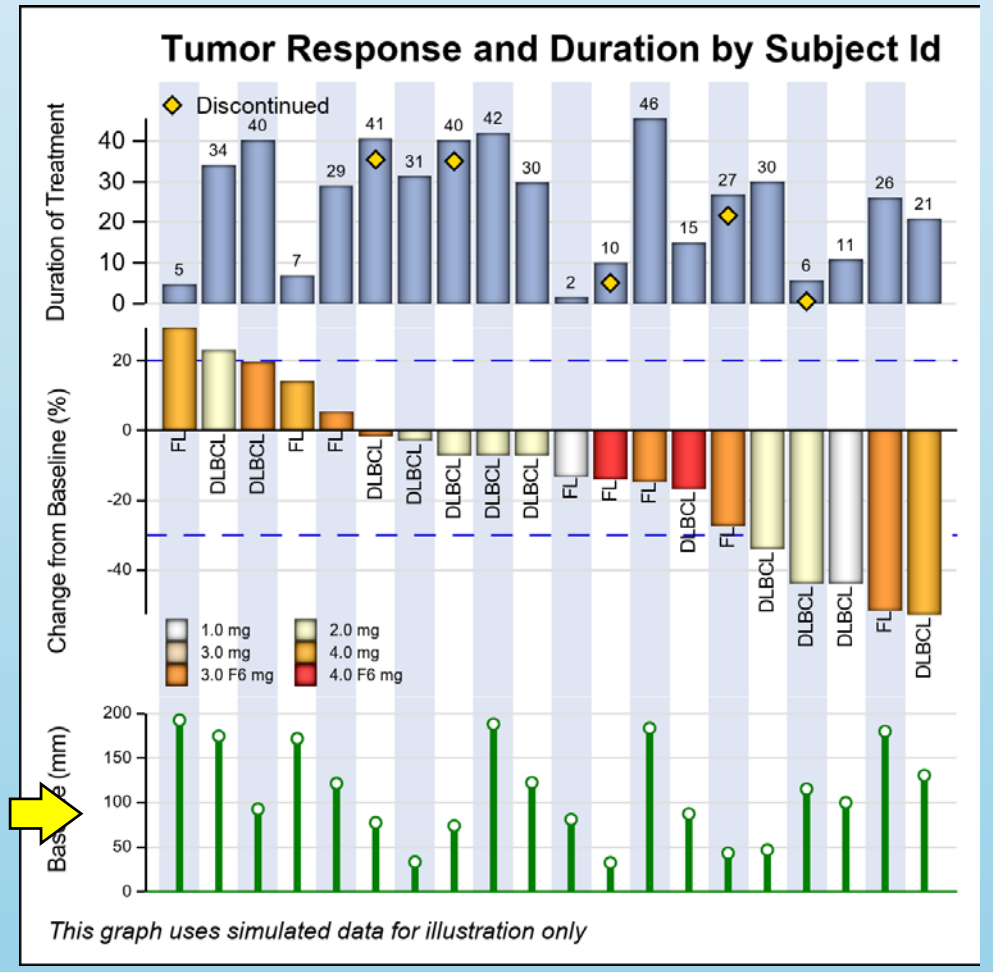
# Displaying more Data is Easier

- Recall Slide #9 displayed the data set used to create these graphs.
- The “Baseline” column contains the initial tumor load.
- Display of “Baseline” can be easily added to this layout as a 3rd row.
- Now, we have a 3 row graph, with baseline tumor load at the bottom displayed using a series plot.
- Note, the axis for this cell is increasing downwards.



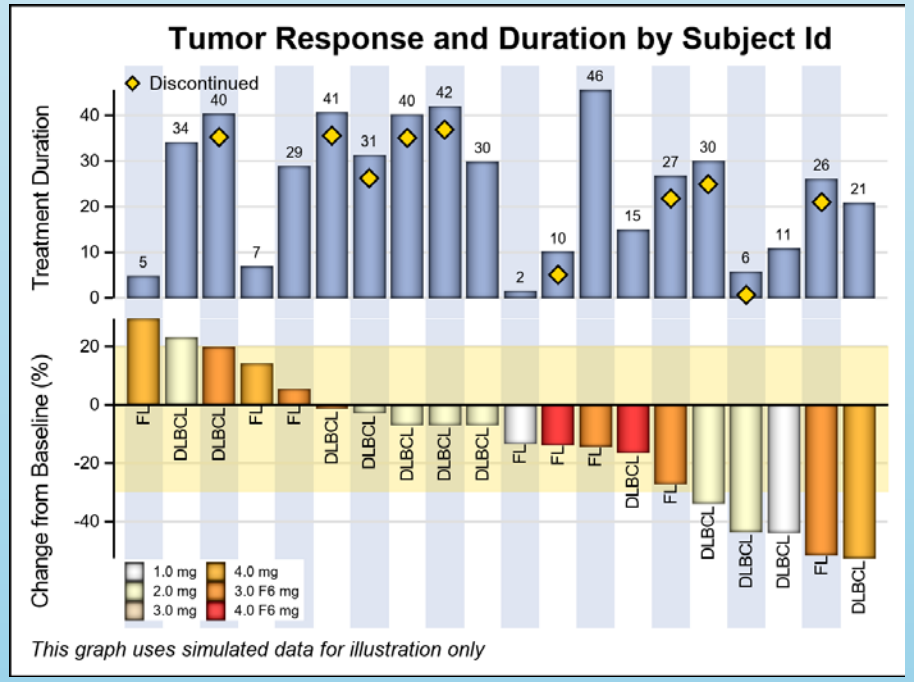
# Displaying more Data

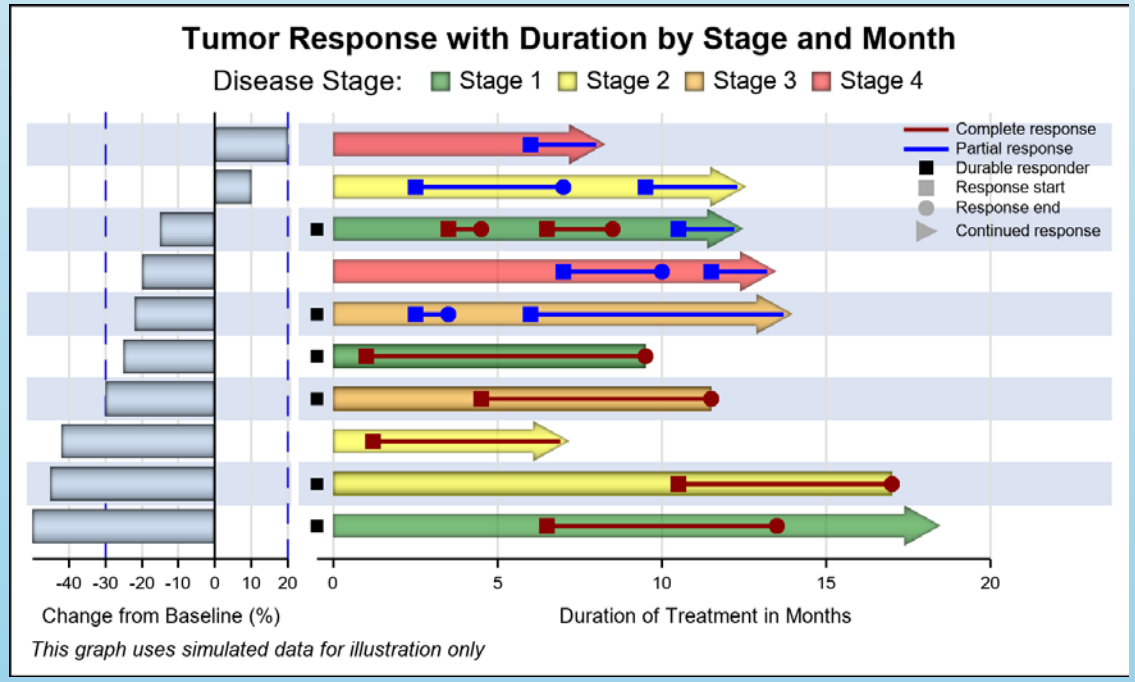
- In this graph, the representation of the baseline tumor load is displayed using a needle plot with markers.
- We could easily use a bar chart.
- The needle with baseline at the bottom is a better visual, as the y-axis direction is clear.
- Display of more data by Subject-Id can be easily added in this layout.



# Alternate Visual for Tumor Data

- So far, we have added duration data to a traditional Waterfall Plot.
- Now, let us review adding Tumor Response data to a Swimmer Plot.



# Data for Extended Swimmer Plot

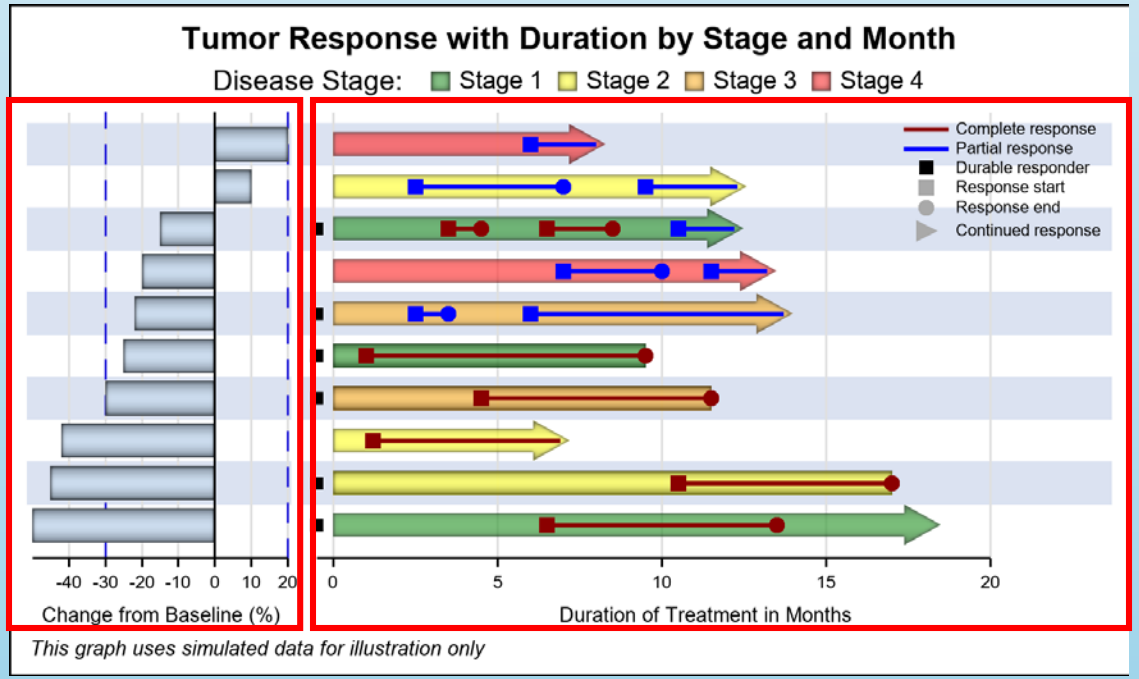
- This is a data set used to create the traditional Swimmer Plot.
- A new column "response" is added for each Subject-Id.
- Note, there are multiple observations per Subject-Id, one for each event.
- The data can be used to create the traditional Swimmer Plot as shown in the article below.

| 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 |
| 15 | 10 | 1 | Stage 1 | 0 | 18.5 | FilledArrow | Complete response | 6.5 | 13.5 | -0.5 | -50 | 6.5 | 13.5 |

- https://blogs.sas.com/content/graphicallyspeaking/2017/10/27/legend-items/#prettyPhoto

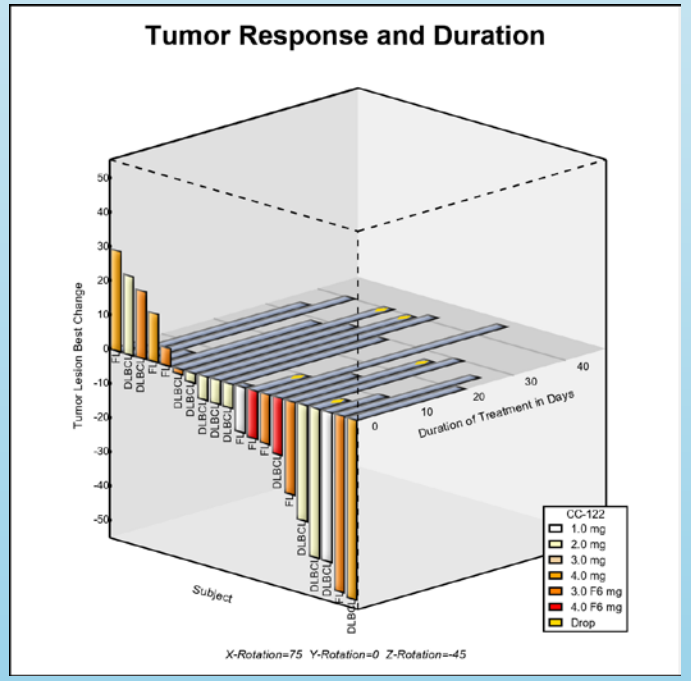
# Extended Swimmer Plot

- Here we have extended the Swimmer plot by adding the display of the tumor response data on the left.
- This graph has a two-column layout.
- To make this graph, we have used the Graph Template Language (GTL).
- The GTL Layout Lattice syntax allows building of multi-column graphs.
- Each cells can have separate multi-layered plots with uniform y-axes.

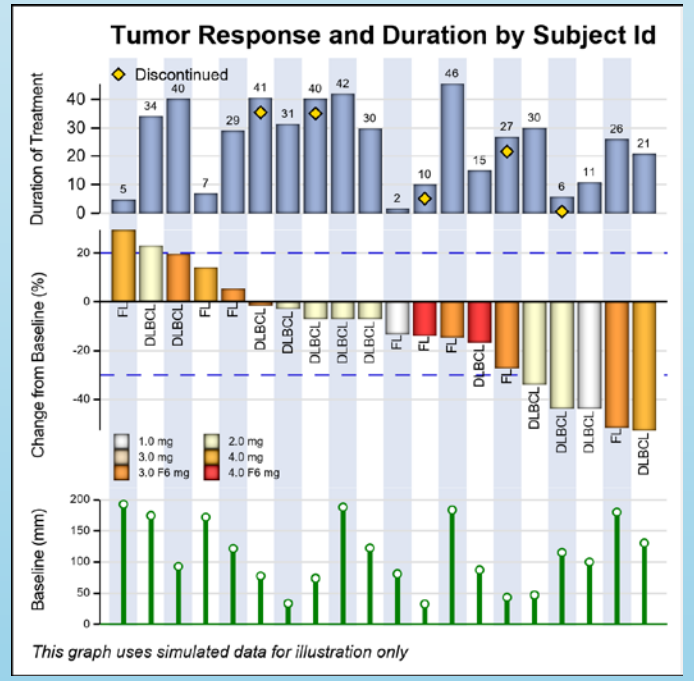


# Summary

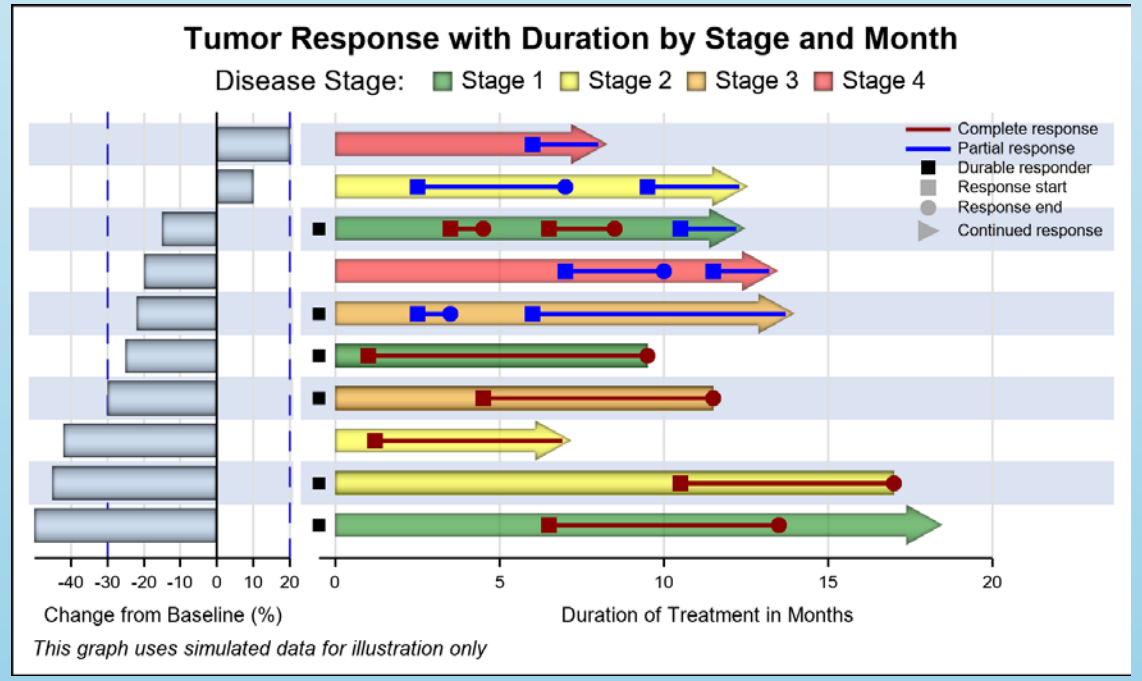
- Displaying additional data in the same graph has its benefits.
- A 2D layout is effective and extensible for display of the data.




3D Waterfall Plot




Multi-Data 2D Waterfall Plot




Multi-Data 2D Swimmer Plot

Name: Sanjay Matange
E-mail: sanjaymatange@gmail.com
Web: https://make-a-graph.com/
Twitter: @matange