

Choosing Your Tile: Quick Clinical Data Summary Visualizations using SAS Enterprise Guide 7.1 Graph Tasks (Tile Charts)

Amber Frazier, PPD, part of Thermo Fisher
NC, USA

ABSTRACT

Graphing can be a great way to review hundreds of data points in just a few minutes but can be intimidating or complex to program. SAS Enterprise Guide 7.1 Tasks offers point-and-click visual representations of many different types without writing any code. Tile charts is one of these options and is often overlooked but can be an efficient way to quickly review categorical representations of numerical data. This paper will briefly discuss choosing the tile chart and sorting the data to create meaningful interpretations of clinical data in addition to filtering, axis modifications, color levels, and other options to aid in tile chart creation. Following the discussion will be some realistic examples of using tile charts to review data, including demographic summaries, visit collection windows, and simplistic visualizations of outlying data points.

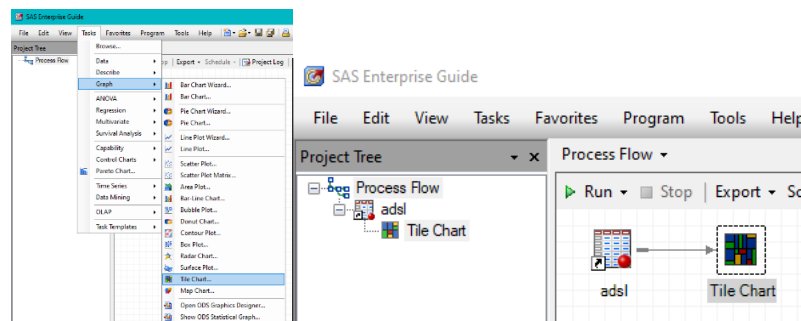
INTRODUCTION

This paper will briefly discuss choosing the tile chart and the data to create meaningful visual summaries of clinical data using filtering, axis modifications, colors, and layout to aid in tile chart creation without writing any code. What can tile charts display?

PIECES OF THE TOOL

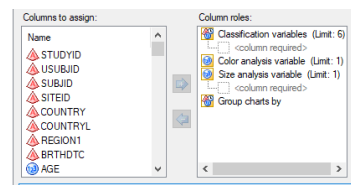
NAVIGATING TO START THE SETUP OF A TILE CHART

Open SAS Enterprise Guide 7.1 and choose Tasks > Graph > Tile Chart. SAS EG 7.1 will guide the user to choose the input data and create a new Process Flow.



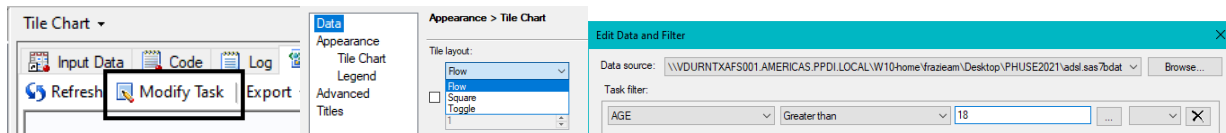
CREATING THE TASK

Drag and drop the variables into the column roles. Blue icons denote numeric fields and red icons denote character fields. While classification and grouping variables can be numeric or character, the size and color variables (used to determine the size and color of the tiles) are required to be numeric.



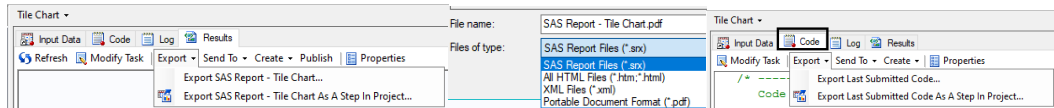
MODIFYING THE TASK

Task modification includes three types of tile charts, changing input variables, title/footnote options, legend appearance and range, and the addition of external web links to the chart output.



EXPORTING THE TASK

SAS EG 7.1 can export the created task in several different formats such as HTML, XML, PDF, and SAS Report Files.

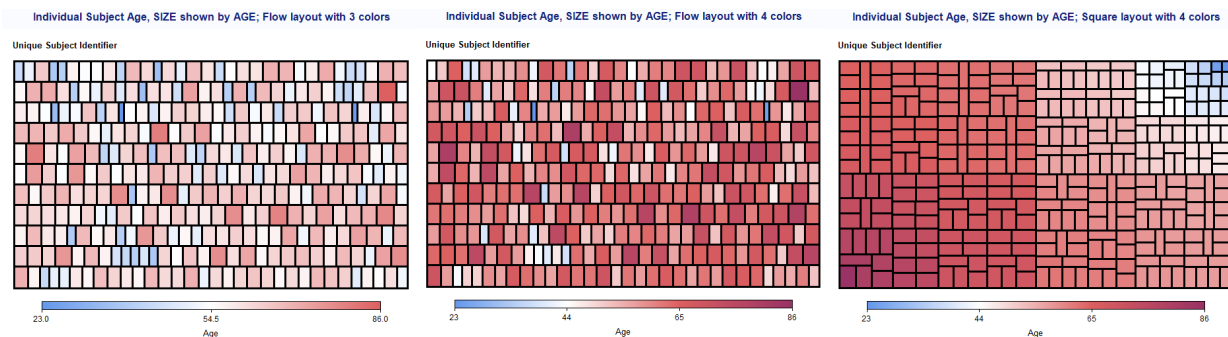


CREATING BASIC VISUAL REPRESENTATIONS

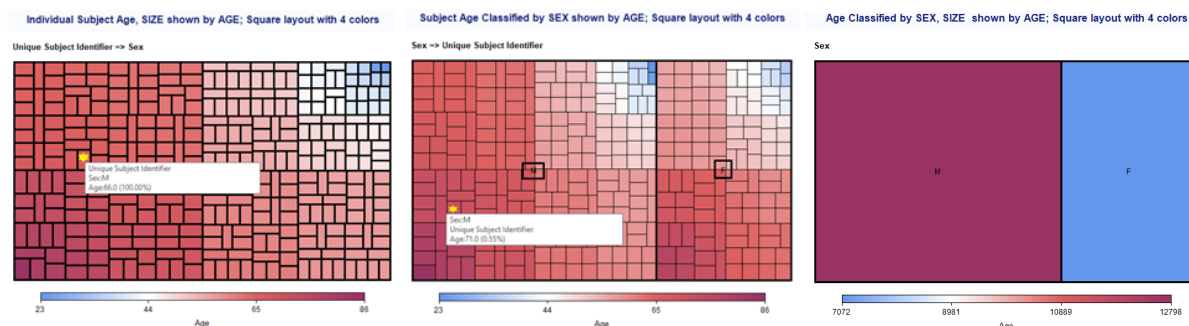
Basic data interpretations useful for tile charts include visually summarizing proportions of population characteristics, showing possible site-level data collection trends, and spotting data outliers.

DEMOGRAPHICS - DEMONSTRATION

At first glance the first tile chart shows the ages (denoted by the color) appear well-distributed; it appears that there are a similar number of dark blue and dark red squares, but it is not obvious what is going on in the middle. Addition of another color to the axis helps gain more specific insight. The second tile chart is shown with 4 colors/levels of axis values. It is easier to see here that most of the subjects are over 44 years old, and there are more subjects at the upper end of the age scale than the lower. Still with the scattered colors it can be tedious to observe with confidence. As shown in the third diagram, the appearance of the tile chart can be improved by opting for the SQUARE layout over the FLOW layout as this will organize the data from bottom left to top right. The ages being grouped into like colors in squares helps us see the proportion of ages with more order.



When choosing a second level of classification, SAS EG will offer the data value in a text box which shows when the cursor is hovering over the observation, but it will not change the overall appearance of the chart. This new chart (shown below) is the same as the third chart above but now the information is available within each data point. If sex should be the primary classification, reversing the order of classification variables in the modification window can show by-sex age distribution. The benefit of leaving the unique subject identifier (subject number - USUBJID) as the second classification variable allows the axis to maintain the subject-level values instead of a summation of classified values and allows display of the number of boxes that make up each level of primary classification. SEX can be added as the GROUP variable instead of the CLASSIFICATION variable to split this data into two separate charts, Male and Female. The sixth chart shows the removal of USUBJID from the classification. Removing the subject-level classification creates a summation of the analysis variable. In this case, this is not a meaningful display and only reflects the sum of the ages of all subjects within each classification.

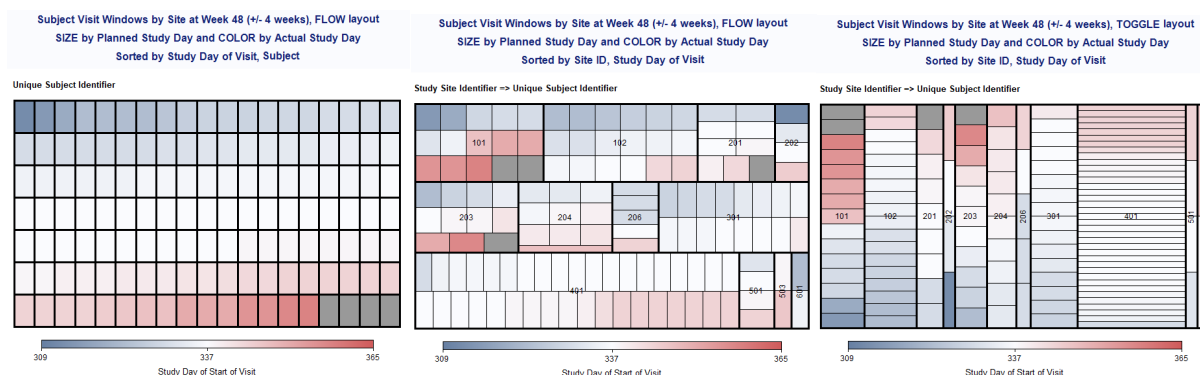


Since AGE is used as a single response variable, the lower ages are smaller boxes and therefore can skew the visual inference if not taken into consideration. In this case there is not much variation between the sizes, but this feature of the tile plot can be tricky if not utilized properly. Tile charts are strongly utilized to show 2 levels of response.

VISIT WINDOWS – DEMONSTRATION

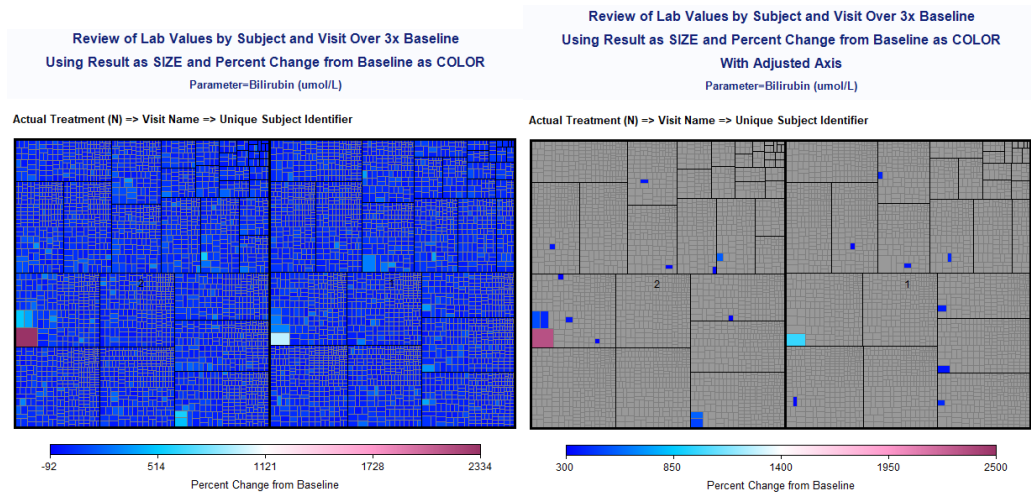
Visit windowing is often a point of interest with a sponsor. Clinical trials commonly have allowable visit windows, which are generally based on dosing intervals and the length of time acceptable between dosing and testing accurately for analysis. A meaningful usage is to review the data points on a subject/visit level, and comparison of the actual visit day to the planned visit day to review possible trending between sites or individual subjects. Below is shown using subject as the classification variable and visit as the grouping variable, the planned visit day as the size and the study day as the analysis variable, adjusting the ranges supported by the axis values and colors. Sorting the data allows the coloring to be easy to interpret. Since USUBJID is a single occurrence, the data is best sorted by classification variables other than the color variable.

For this example, the review will be of the endpoint of a study at Week 48, where the visit window is +/- 4 weeks. This uses a FLOW layout and adjusts the axis to mimic the visit window. By default, anything not within the axis range appears gray. Since the number of colors in the axis is minimal, anything blue is on the lower end of the window, red is the upper end of the window, the whites are close to the planned study day and the grays are out of window. Reviewing using SITE as a classification variable can show trends on that level. When comparing within a stratification a TOGGLE layout displays the site-level data into horizontal columns and adds another level of organization for comparison between sites, including the number of subjects and any windowing trends.



OUTLIERS - DEMONSTRATION

The number of colors used in the axis can be useful for creating more specific summaries and it can also be used to help spot outliers. This example uses lab values and toxicity grading. Bilirubin at 3x baseline is a Grade 3 toxicity level, but more than 10x baseline can be a serious Grade 4 toxicity level. Adjusting the axis values to reflect the percentage range of interest simplifies looking for trends related to treatment and visit. Below are two examples of the same tile chart with different axis ranges. The first does show some extreme values as compared to the rest of the data, but the range is so wide it is hard to distinguish any differences between -92% and 1121%. Adjusting the axis to include test results between 3 and 10 times over baseline, more easily seen are the instances which might be considered safety concerns. In the second chart, it is notable that most of the extreme values do fall within about 300-1000% of the baseline value. However, there is one notable value in pink. This one stands out as it is over 20 times the baseline value and should be investigated. Using interactive SAS EG, the subject number and visit can be noted by hovering the cursor over the pink block.



CONCLUSION

Learning how to create SAS Graph tasks can aid in the visualization and review of data points collected without the aid of any manual programming. Multiple levels of classification create drill-down mouse-over access to the actual data values. Legend and axis manipulation can highlight trending and data outliers. Have fun choosing your tile!

CONTACT INFORMATION

Contact the author at:
 Amber Frazier
 PPD, part of Thermo Fisher
 Wilmington, NC
 Email: a.frazier.stats@gmail.com

Brand and product names are trademarks of their respective companies.