



## **Data visualization for clinical data using SAS. Visualize clinical data using SAS, such as heatmaps, scatter plots, and interactive dashboards.**

**Arvind Uttiramerur**

ThermoFisher Scientific

---

### **ABSTRACT**

The rise, in data complexity and volume in trials has led to an increase in the creation of Statistical Tables, Figures and Listings (TFLs). This growth presents a challenge for professionals like Statistical Programmers, Biostatisticians, Clinicians, Medical Writers and others involved in producing, reviewing and summarizing trial details in a study report (CSR). To tackle this issue there is a growing demand to present trial results, in an engaging and efficient way. This article suggests a method to achieve this goal by utilizing representations. These visual aids come in formats. Mainly aim to concisely summarize the key information.

**Keywords:** Data visualization, SAS

---

### **INTRODUCTION**

Clinical trials serve purposes like assessing the safety and efficacy of new treatments. With time these trials have evolved in terms of numbers, trial complexity and data volume. The growing data has led to the need for statistical tables, figures and listings that can be time-consuming and challenging to interpret. This paper proposes the use of displays for a typical Phases 2 or 3 trial with a double blind, randomized placebo-controlled design. While the ideas discussed here are specific to trials comparing treatments with placebos they can be adapted for other phases and study designs. It's important to mention that the displays in this paper are not from a trial but are instead examples sourced from various publications and online sources. There are excellent examples of figures displaying clinical trial data available online, from previous PhUSE conferences, PharmaSUG publications and CTSPedia. The main goal of this paper is to showcase sample figures that effectively summarize clinical trial results.

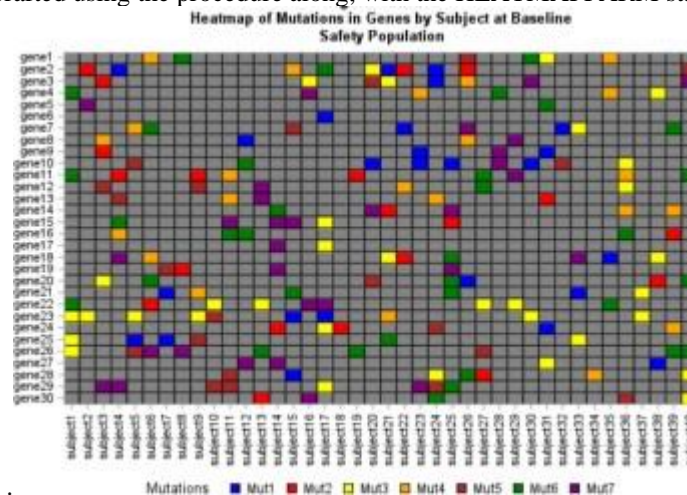
Summarizing the aspects of a trial design and its outcomes visually involves presenting an overview of the trial groups critical evaluation timepoints along with what is assessed at each timepoint participant distribution, demographic details, efficacy, safety, and other significant findings.

Numerous studies in the field aim to assess efficacy by using various endpoints. The recent ICH E9 (R1) guidance, titled "Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials" introduces the estimand framework that can be employed to design a trial incorporating specific efficacy or safety endpoints crucial for regulatory decision making. This framework highlights the significance of defining the estimand and performing sensitivity analyses to gauge result robustness. Regarding safety information, while some displays may exhibit demographics or efficacy data there are now approaches available for presenting critical safety details, such, as evaluating Adverse Events, Laboratory Results, Vital Signs and ECGs.

Heatmaps are a tool for visualizing data using colors to depict information in a two-dimensional image. They are effective for illustrating cancer traits and characteristics. Additionally, they can show treatment related side effects, metabolic irregularities and infection rates among population groups. Heat map visuals aim to simplify data by using color gradients to showcase diverse data values allowing for quick comprehension of intricate information. In contexts they serve as valuable aids, in clinical interpretation.

In trials domain heat maps are handy tools for illustrating adverse effects of treatments highlighting metabolic irregularities and comparing infection rates among different groups. Essentially a heat map graph visually displays data in colors to indicate varying values. The main aim is to offer a summary of information for better understanding. Heat maps are widely used to track weather patterns, geographical elevations and human

migrations. In the context of trials heat map graphs can depict adverse effects based on treatment types metabolic issues, infection rates in various demographics like children versus young adults or males, versus females. This study showcases how treatment groups compare to placebos using examples and SAS® procedures to enhance the graphical representation. Graphs are a way to visually represent data that may be hard to interpret in a table. If you aim to convey a lot of information in a space heat maps can come in handy. These maps are versatile and easy to understand making them popular across industries. They showcase data values through colored and patterned rectangular areas or cells which are usually quite clear. The darker the color, the the value it represents. With these graphs we can plot both numerical and numerical data using solid colors or gradients to show different ranges within the data values. Tools like SAS can be really useful for creating heat maps. While there are methods like Graph Template Language available for creating these visual representations features like HEATMAPPARM and HEATMAP, in SGPLOT and SGPANEL procedures of SAS® 9.4 make creating heat maps relatively straightforward. The structure of HEATMAPPARM and HEATMAP statements is quite similar when using both SGPANEL procedures. This study mainly focuses on SGPLOT. The HEATMAPPARM statement generates a representation that showcases the values of a response variable through various colors. This statement necessitates the inclusion of X, Y and either COLORRESPONSE or COLORGROUP arguments. The COLORRESPONSE argument exclusively works with variables and applies a color gradient to cells continuously. To illustrate heat maps in this study a sample dataset containing 5270 observations with 4 variables such as ID, system organ class of adverse events and treatment group variables was utilized. The frequency distribution of events for each SOC based on treatment groups was computed as depicted below in the SOCFREQ dataset, which includes the treatment group variable TRT system organ class variable AESOC and frequency count variable COUNT. Several instances of heat maps crafted using the procedure along, with the HEATMAPPARM statement are presented.

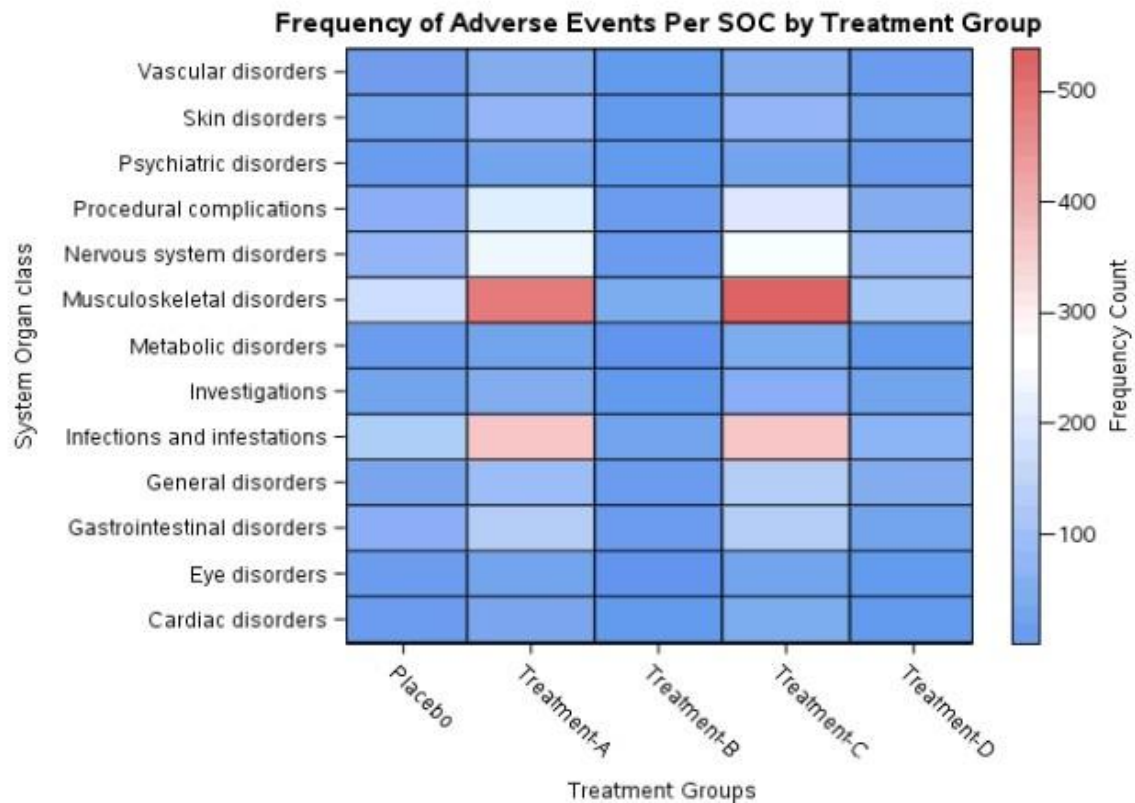


The heat maps display how often negative events occur within each system organ class based on the treatment group. By observing the colors we can determine which organ class experiences fewer adverse events across different treatments. The graph illustrates the relationship between the variables AESOC (Y axis) and TRT (X axis) with cell colors representing the frequency count of the response variable COUNT through the HEATMAPPARM statement. Through the use of the COLORRESPONSE argument a color gradient is applied to indicate frequencies continuously. The legend on the side of the plot showcases frequency levels using varying shades of default colors, where darker blue signifies lower frequency and darker red indicates higher frequency. The OUTLINE option ensures that all cells are outlined in black, for clarity.

The code for the graph is shown below:

```
proc sgplot data=socfreq;
title "Frequency of Adverse Events Per SOC by Treatment Group";
heatmapparm x=trt y=aesoccolorresponse=count / outline;
yaxis label="System Organ class"; xaxis label="Treatment Groups";
run;
```

Figure 1 shows a heat map using all the options mentioned above and default assignment of colors.



### CREATING A SCATTER PLOT WITH SGSCATTER

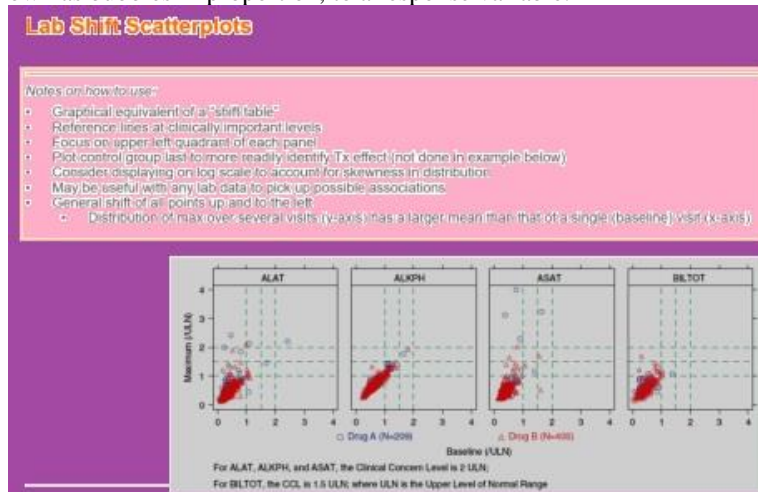
Creating a Scatter Plot involves plotting data points on a graph with vertical axes to explore potential relationships between variables. Scatter plots are useful for visualizing how one variable is influenced by another. Are commonly used to assess correlation between two variables. When you click on a data point in a scatter plot it triggers an action that generates a query containing information about that point. This query can also be generated using tools like dropdown menus. Interactive data visualizations are always linked to the data source ensuring that any queries made update the data and subsequently affect visuals accordingly. This allows for updates, in graphics to either emphasize or filter points based on the results obtained.

The scatter plot results are generated using the SGPLOT method along with a scatter statement. Scatter plots in SGScatter show data points on both vertical axes to assess if there is a potential relationship between two variables, often known as their correlation. In addition to visualizing repeated measures data scatter plots can also illustrate XY relationships, such as the connection between estimated changes and actual changes in vital signs ECG readings or other continuous safety data. The output of the scatter plot is created using the SGSCATTER method with a plot statement. Scatter matrix plots data points on horizontal and vertical axes to determine potential relationships, between variables. The scatter plot output is generated using the SGSCATTER method with a MATRIX statement.

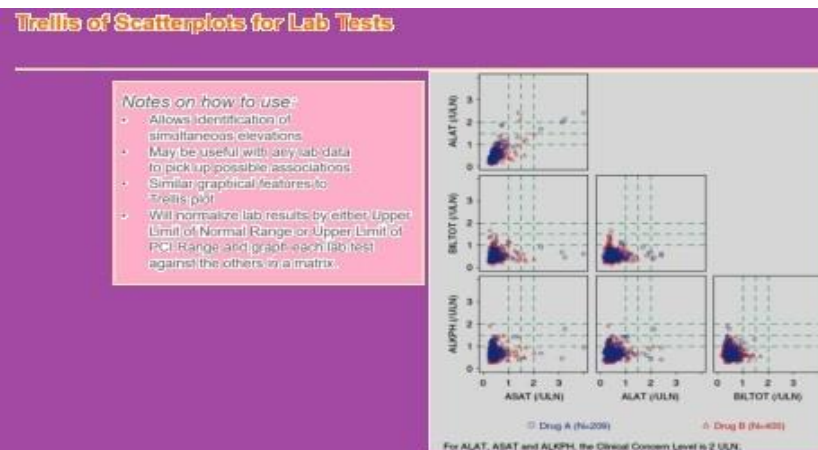
```
TITLE 'Basic Scatter Plot';
PROC SGPLOT DATA= dataset ;
SCATTER X=variable 1 Y=variable 2 ;
RUN ;
PROC SGSCATTER < DATA=data-set-name >> options > ;
Plot variable1 * variable2 / < options > ;
Compare X=(variable(s)) Y=(variable(s)) < / options > ;
Matrix variable(s) < / option(s) > ;
RUN ;
TITLE 'SGScatter Plot Matrix' ;
```

PROC SGSCATTER DATA= Dataset ;  
 MATRIX X variable Y variable ;  
 RUN ;

A volcano plot is a scatter plot that shows statistical significance like from a Fisher's Exact test compared to the extent of change indicated by the Odds ratio for numerous tests, such as for each adverse event term preferred between the groups receiving treatment and those on placebos. On the hand a bubble plot is a distinct scatter plot where dot sizes are shown as bubbles in proportion, to a response variable.

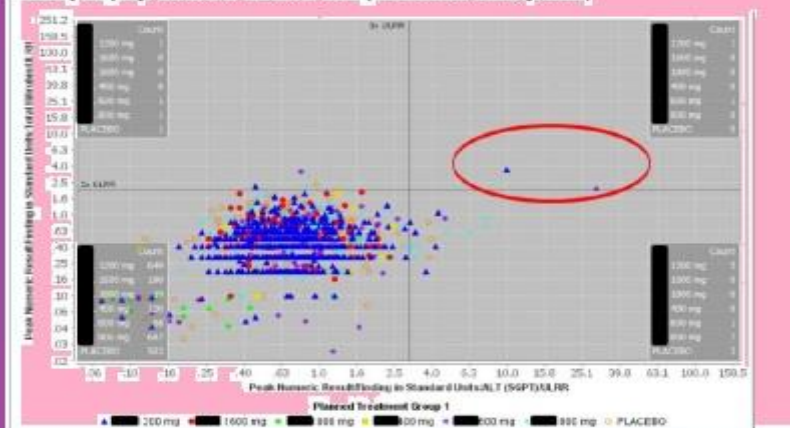


A different kind of scatter plot can be utilized to assess liver function irregularities based on Hy's law as shown in the illustration.



**Plots to Identify Potential Hy's Law Cases**

**Figure 1. Hy's Law Plot: ALT>3xULN vs TBili>2xULN with ALP <2xULN Nonpilepsy Double-blind Pool (includes Study 206)**



**INTERACTIVE DASHBOARDS USING SAS VISUAL ANALYTICS.**

This paper introduces the use of six template reports giving an overview and detailed explanations about the data objects utilized and potential interactive features. It covers the analysis of a sample assignment outlining the data preparation and adjustments for report creation highlighting the conversion of data formats for visualization purposes. Tailored for individuals lacking experience in web app development or JAVA SAS Visual Analytics facilitates the creation of web based reports through user friendly drag and drop and point and click methods. SAS Visual Analytics improves reporting by introducing interactive controls and graphs allowing for dynamic exploration of data via web interfaces. Data visualization is commonly associated with data where billions of records and millions of variables are analyzed concurrently. However this doesn't imply that data scientists analyzing clinical trial data containing thousands of records and hundreds of variables cannot benefit from data visualization techniques. The document outlines a process, for importing standard clinical trial data into SAS® Visual Analytics, an interactive analytics tool. The process incorporates template reporting to enable a range of point and click visualizations.

For a century data has traditionally been shared through printed reports like journal articles or regulatory submissions containing tables, graphs and written explanations. This method has only recently been replaced by PDF files, which are closely related to paper. Comparatively SAS Visual Analytics is in a league of its own when it comes to data representation. For those accustomed to creating printed reports or PowerPoint slides with tables and graphs using SAS Visual Analytics will introduce them to a new way of reporting. In some aspects working with SAS Visual Analytics may feel like software development rather than traditional data analysis. Despite this perception utilizing SAS Visual Analytics does not demand expertise, in web app development or JAVA programming. The reporting capabilities offered by SAS Visual Analytics can be accessed through drag and drop and point and click methods – allowing statisticians and clinical trials programmers to create dynamic web based reports with ease. While some analysts may argue that they are statisticians and not web designers based on their job titles set by Human Resources they should not underestimate their ability to craft impactful reports. SAS Visual Analytics provides analysts with the chance to develop web based reports that showcase the impact of key variables established connections trustworthy relationships and other findings from statistical models. Without involvement from analysts during the report development process report authors may resort to using convenient variables instead of focusing on those that are truly significant. This could result in reports that simply present hierarchies and relationships in a mechanical manner. Typically the initial dropdown selection would represent the condition, followed by a demographic variable and so forth. While there is nothing incorrect about such reports the challenge lies in the fact that in many instances the actual impact of treatment could be influenced by interactions between the experimental condition, various demographic factors and other variables. The impact might be specific to an age group or dependent on subjects who did not encounter a specific severe side effect within the first month of the study. The real world is intricate and multifaceted making it challenging to capture this complexity through traditional printed reports. However SAS Visual Analytics offers a straightforward approach, to showcasing this complexity. The template reports outlined here serve as examples of what can be achieved with this tool.

Extensive training on SAS Visual Analytics goes beyond the scope of this document. Providing a brief overview of the software and its terminology can be beneficial. SAS Visual Analytics is a server based application that users can access similarly to how they access SAS® Drug Development. Each user has an ID and password with their access rights managed by the System Administrator. The System Administrator determines whether a user can load data view data or create and view reports. The source data for the reports mentioned in this paper consists of SAS® datasets. As SAS Visual Analytics operates as an in memory software solution data needs to be loaded into the SAS® LASR™ Server, which's a RAM based analytics engine. This loading process can be carried out in ways as detailed in the User Manual. Once data is loaded into the SAS LASR Server it becomes accessible for analysis and reporting purposes. Within SAS Visual Analytics there are two interfaces, for analysis and reporting; SAS® Visual Analytics Explorer and SAS® Visual Analytics Designer. It's worth noting that these terms may not entirely reflect their functions. You can create a report as a way to explore data. You can also analyze data within a designed report. The templates showcased in this document were constructed using the design tool. One benefit of the design tool is that it allows a research team to offer reporting access without compromising the security and stability of the underlying data. Analysts who are worried that experienced users might misinterpret data or generate inaccurate reports can maintain more control over the reporting process by simply assigning other users to view reports in Read Only mode. Once data is uploaded into SAS Visual Analytics report creators have an array of objects at their disposal for presenting, visualizing and managing interactions among visual elements, in a report. Variables found under the "Data" section can be dragged onto these objects, which can then be used to produce a report. But the true value of SAS Visual Analytics is the interactivity between these controls and graphs. In traditional scientific reporting, a graph is a fixed-format, static representation of a relationship. The X variable is on the horizontal axis and the Y variable is on the vertical axis on paper, this graph would probably appear with a table of the actual values. This basic graph, surfaced as html on a web page, could have additional functionality, enabling users to scroll over bars to see the actual counts.

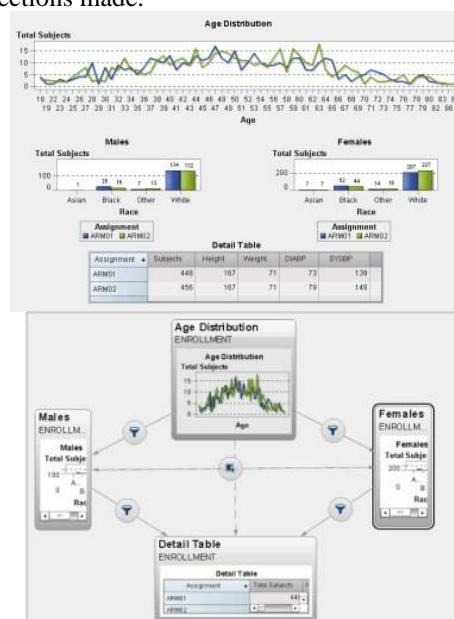
### OVERVIEW OF THE TEMPLATE REPORTS

In this paper the upcoming parts detail how six template reports were put into practice. A summary of each report is presented in the table below. Following this summary is the Report Template Gallery showcasing each report along with details, on the data objects utilized and potential interactivity.

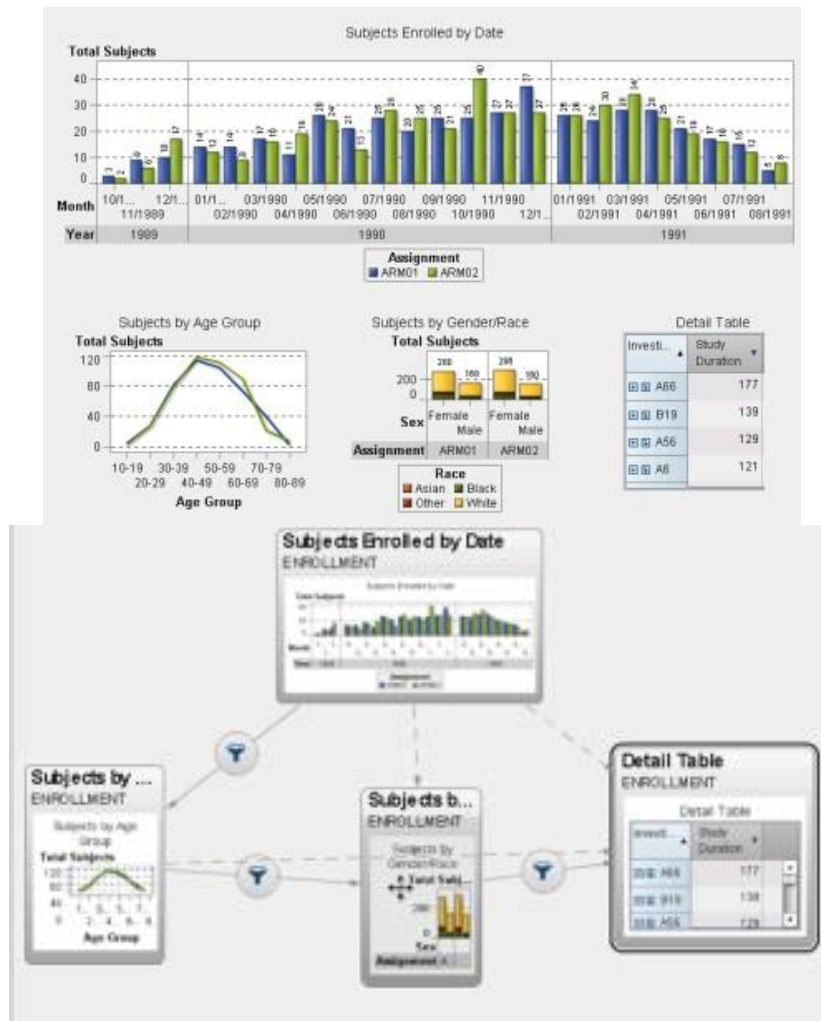
During the initial assessment, we examined a demographic dataset containing information on experimental assignments and the initial recording of height, weight, and blood pressure for each patient. The data is organized at the individual patient level per row, serving as a key dataset for future analyses and reports in visual analytics.

The report starts by looking at the age distribution graph, which then influences the bar charts for males and females, as well as the detail table. It's important to note that when selecting race in the bar charts for males and females, changes affect both sides of the chart and the detail table.

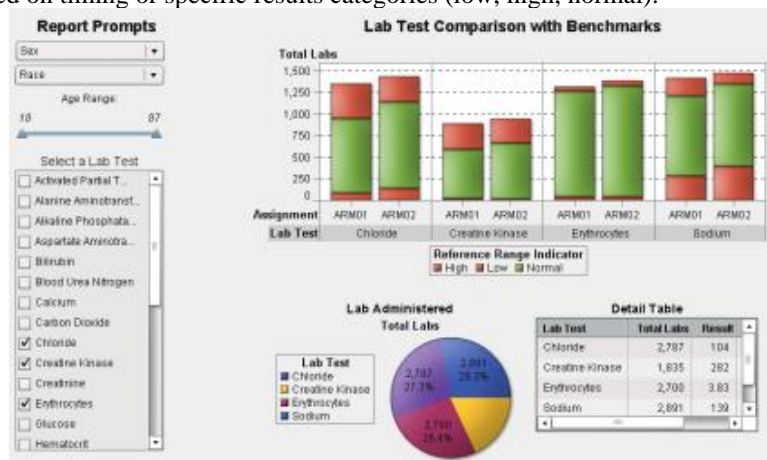
The enrollment history report is based on the same data as the trial assignment analysis, but it includes additional variables like the Investigator. The data is organized at a one patient per row level. Visual analytics elements in this report include a bar chart showing subjects enrolled by date, another bar chart with stacked options displaying subjects by gender and race, a line plot depicting subjects by age group and a crosstab detailing table. The report allows for interactivity starting from the top with the enrollment graph acting as a control for selecting different groups (such as arm01/arm02, specific months or individual bars). These selections influence the graph's display and can be interacted with from left to right. The detail table on the far right provides more data that can be sorted and explored based on previous selections made.

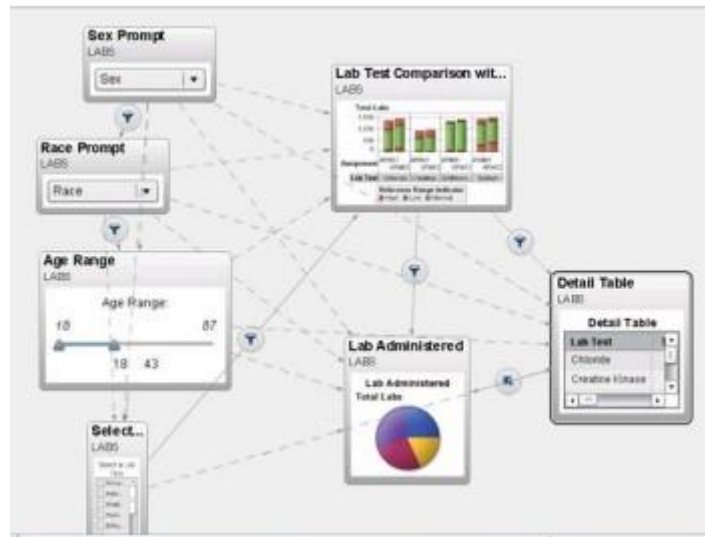


Comparing Lab Test Data Sources; The data structure in this analysis combines information from the Enrollment History report with lab data, creating a dataset that lists one lab per row, with patients and labs repeated to show all tests given to each patient. The visual analytics tools include a dropdown menu for selecting sex and race, a slider for choosing age ranges, a list of available lab tests, a bar chart with stacking options and column lattices, a pie chart displaying total labs and a detailed table. Users can interact with the report by choosing sex and race from dropdown menus, selecting age ranges using the slider, picking one or more lab tests from the list provided. An interesting feature is that selections are interconnected. Choosing one option influences the choices available in subsequent selections. The bar chart reflects all chosen options and can also impact the data shown in the pie chart and detailed table. For a comprehensive view of these interactive features, refer to Figure 9 which illustrates how all selections are linked together.



The patient's medical records contain information sourced from the same data used for the Lab Test Comparison report. The data is structured at the level of one lab per row, with details of patients and labs repeated to encompass all tests conducted on all patients. The Visual Analytics Objects in this report include a Text Input Field (located under "Search for a Patient"), a List Table displaying Subject Details, a Bar Chart presenting data in clustered form (depicting Lab Administration Timeline) and another List Table showing Lab Results Details with applied display rules. This report follows a conventional top-down approach and is designed to be interactive. By entering a patient ID into the text input field, specific patient data can be accessed. The interactive bar chart allows selection of labs based on timing or specific results categories (low, high, normal).

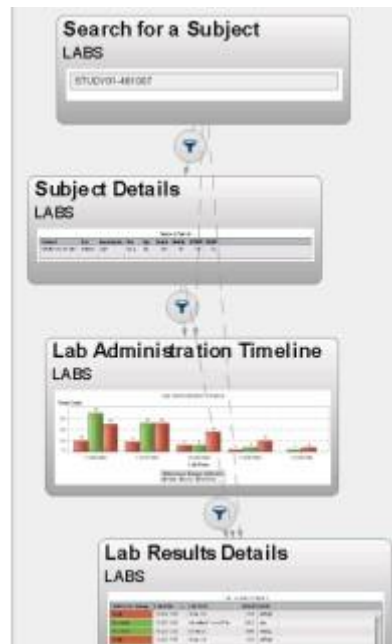




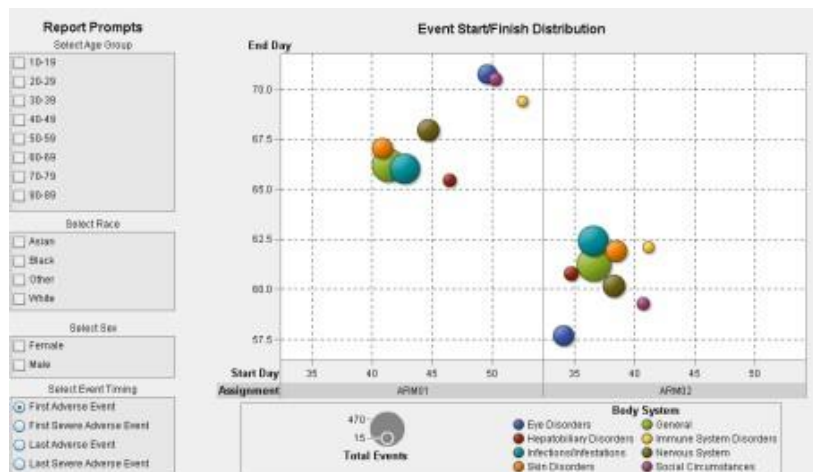
The information about adverse events duration is sourced from data that has been integrated into the Trial Assignment/Enrollment History data structure. This dataset does not include laboratory results. The data is organized at a one event per row level, with both patients and events being repeated to account for all occurrences experienced by each patient. It's important to note the typical repetition of the data explained below. The Visual Analytics Objects consist of List Box Controls for Age, Race and Sex, each requiring a single selection only. Additionally, there is a Bubble Plot with Hierarchy assigned to the Group variable and a lattice that separates the two experimental conditions. Regarding interactivity, all reporting is influenced by selecting options in the bottom left of the screen (the radial box that filters one of four events in the data). For this report, making a selection here is mandatory. The default setting is "First Adverse Event." Once selected, users can choose age groups, race and sex of patients which are then reflected in the bubble plot. Although it's not visible here, the bubble plot allows drilling down for further details on each grouping by clicking on a bubble.

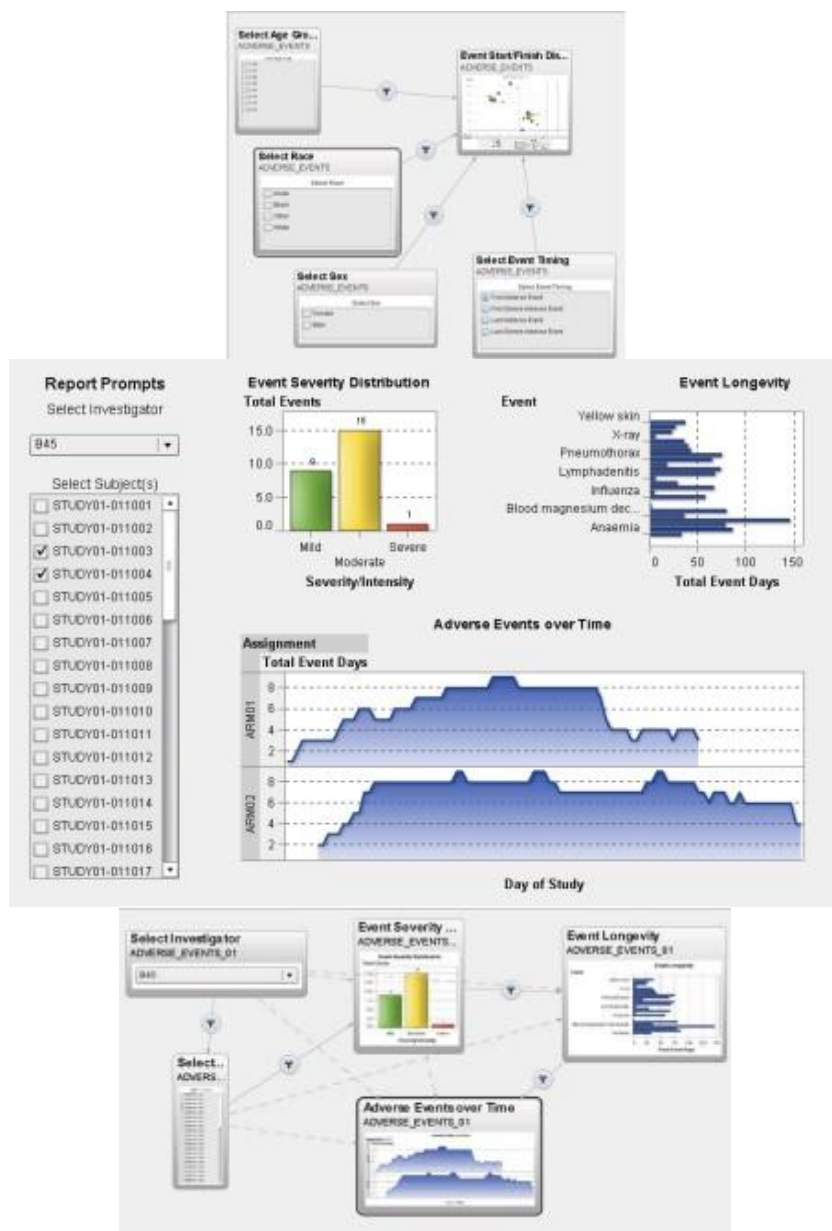






Adverse Events Over Time Data Source This dataset contains information similar to what is found in the Adverse Event Duration report, detailing individual events on a row by row basis. Both patients and events are listed multiple times to account for all occurrences experienced by each patient. In contrast to the previous report, the data structure presented here is not stacked. Visual Tools for Analysis. Dropdown Menu (Select Investigator). List Box (Subject). Vertical Bar Chart (Event Severity Duration). Horizontal Bar Chart (Event Longevity). Line Plot with Overlay Filled feature Interactivity When an investigator is chosen from the dropdown menu, the list of subjects narrows down to only show patients associated with that investigator in the list box. These selections are reflected in the graphs, with each graph influencing the next one as you move from left to right and top to bottom. To focus on events of longest duration, users can select the right end of bars on the Event Longevity chart, which will then adjust the bottom graph to display only these long lasting events and indicate when they began under two different experimental conditions.





**CONCLUSION**

Patterns and distinctions in data may not always be immediately clear when presented in tables and reports. Utilizing visualization techniques can offer distinct advantages over traditional communication methods such as tables and reports. By employing visual methods one can enhance the understanding of data, a requirement that resonates with data analysts, statisticians and others. The SAS user community benefits from the availability of ODS Statistical Graphics tools that empower users to create quality graphical outputs including charts, figures and visuals for data exploration, analysis and statistical purposes. This paper outlines ten guidelines for leveraging ODS Statistical Graphics highlighting its features, capabilities and syntax to enhance the production of superior charts and visuals. Using SAS to generate heat map plots proves to be an approach for visually representing large volumes of complex data. Procedures like SGPLOT in SAS® 9.4 along with statements, like HEATMAPARM and HEATMAP enable users to create graphical representations as illustrated in this research paper. The paper showcases the generation of heat map graphs using used options on a sample dataset.

**REFERENCES**

[1]. “Seeing Is Believing! Good Graphic Design Principles for Medical Research”, Susan Duke et. al at [https://www.ctsmedia.org/wiki/pub/CTSmedia/GraphicsPresentationArchive/DIA2014\\_Susan\\_Duke\\_Graphi cs.pdf](https://www.ctsmedia.org/wiki/pub/CTSmedia/GraphicsPresentationArchive/DIA2014_Susan_Duke_Graphi cs.pdf)

- [2]. “Graphical and Quantitative Literacy: Empowerment Begins with Naming and Describing What We Do”, Joint Statistical Meetings, Vancouver, BC, July 29, 2018, Susan Duke, Mathematical Statistician, Office of Biostatistics, CDER, FDA.
- [3]. “A Picture is Worth a Thousand Tables, Graphics in Life Sciences”, Editors Andreas Krause, Michael O’Connell, Springer, 2012
- [4]. Pocock SJ, Trivison TG, Wruck LM. Figures in clinical trial reports: current practice & scope for improvement. *Trials*. 2007 Nov 19;8:36. doi: 10.1186/1745-6215-8-36. *Trials*. 2007 Nov 19;8:36.
- [5]. Murali Kanakenahalli, Avani Kaja. Graphical Presentation of Clinical Data in Oncology Trials. *PharmaSUG 2015 - Paper DV10*
- [6]. Chia PL, Gedye C, Boutros PC, Wheatley-Price P, John T. Current and Evolving Methods to Visualize Biological Data in Cancer Research. *J Natl Cancer Inst*. 2016 May 31;108(8):djw031. doi: 10.1093/jnci/djw031. Print 2016 Aug. PMID: 27245079
- [7]. Whiting, P. F., Sterne, J. A., Westwood, M. E., Bachmann, L. M., Harbord, R., Egger, M., & Deeks, J. J. (2008). Graphical presentation of diagnostic information. *BMC medical research methodology*, 8, 1-15.
- [8]. Tsymbal, A., Huber, M., Zillner, S., Hauer, T., & Zhou, S. K. (2007). Visualizing Patient Similarity in Clinical Decision Support. In *LWA* (pp. 304-311).
- [9]. Plotly. “plotly.js | JavaScript Graphing Library.” Accessed April 29, 2018. <https://plot.ly/javascript/>.
- [10]. Square, Inc. “Crossfilter.” Accessed April 29, 2018. <https://square.github.io/crossfilter/>.
- [11]. SAS Institute Inc. “SAS® Help Center: Syntax: PROC STREAM.” Accessed April 29, 2018. <https://go.documentation.sas.com/?docsetId=proc&docsetTarget=p06pqn7v5nkv02n0zpkp7832j1yp.htm&docsetVersion=9.4&locale=en>.
- [12]. SAS Institute Inc. “SAS® Help Center: Syntax: PROC JSON.” Accessed April 29, 2018. <https://documentation.sas.com/?docsetId=proc&docsetTarget=p06hstivs0b3hsn1cb4zclxukkut.htm&docsetVersion=9.4&locale=en>.
- [13]. SAS Institute Inc. “Clinical Data Analysis Software | JMP Clinical.” Accessed April 29, 2018. [https://www.jmp.com/en\\_my/software/clinical-data-analysis-software.html](https://www.jmp.com/en_my/software/clinical-data-analysis-software.html). SAS Institute Inc. “Findings Shift Plots | JMP Clinical.” Accessed April 29, 2018. [https://www.jmp.com/support/downloads/JMPC71\\_documentation/Content/JMPCUserGuide/OP\\_C\\_FI\\_0003.htm](https://www.jmp.com/support/downloads/JMPC71_documentation/Content/JMPCUserGuide/OP_C_FI_0003.htm).
- [14]. SAS Institute Inc. “AE Incidence Screen | JMP Clinical.” Accessed April 29, 2018. [https://www.jmp.com/support/downloads/JMPC71\\_documentation/Content/JMPCUserGuide/OP\\_C\\_EV\\_0004.htm](https://www.jmp.com/support/downloads/JMPC71_documentation/Content/JMPCUserGuide/OP_C_EV_0004.htm).
- [15]. “Plotly.js + Crossfilter.js.” Accessed April 29, 2018. <https://plotly.github.io/plotly.js-crossfilter.js/>.
- [16]. Prashant Hebbar and Sanjay Matange. 2016. “Graph A Million with the SGPLOT Procedure.” *Proceedings of the SAS Global 2016 Conference*. Available at <https://support.sas.com/resources/papers/proceedings16/SAS4341-2016.pdf>
- [17]. Full documentation of SAS Visual Analytics is available here: <http://support.sas.com/software/products/va/>