

Advanced Visualization using TIBCO Spotfire® and SAS®

Ajay Gupta, PPD, Morrisville, USA

ABSTRACT

In Pharmaceuticals/CRO industries, you may receive requests from stakeholders for quick access to clinical data to explore the data interactively and to gain a deeper understanding. TIBCO Spotfire is an analytics and business intelligence platform which enables data visualization in an interactive mode. Users can further integrate TIBCO® Spotfire with SAS® (used for data programming) and create visualizations with powerful functionality e.g. data filters, data flags. These visualizations can help the user to self-review the data in multiple ways and will save a significant amount of time. This paper will demonstrate some advanced visualization from Preclarus® Patient Data Dashboard (Preclarus PDD) within PPD® created using TIBCO Spotfire and SAS (for the SDTM database) and share our experiences and challenges while creating this visualization.

INTRODUCTION

TIBCO Spotfire is an analytics and business intelligence platform which enables data visualization in an interactive mode that has grown in popularity within the pharmaceutical industry over the last few years. With its increasing implementation in the field of safety monitoring and data review, TIBCO Spotfire presents its capabilities for exploratory analysis. On the other hand, SAS software gives us efficient, strong statistical analysis and data processing capability.

This paper focuses on how CDISC SDTM standards can be leveraged (using SAS software and TIBCO Spotfire) to create a user friendly, highly visual and interactive environment, where any clinical trial that is generating SDTM data, can quickly access and efficiently review their data on an ongoing basis, for example to support areas such as safety review, identifying data issues, supporting dose escalation and identifying trends. Our Preclarus Patient Data Dashboard (Preclarus PDD) has been deployed on over 80 studies within PPD and growing. This paper is a great opportunity to share our experiences and challenges when using CDISC SDTM standards in this way and the huge advantages that these visualizations are bringing the monitoring and ongoing review of our clinical data.

TECHNIQUE AND MECHANISM

The general process of creating data visualizations in TIBCO Spotfire is as follows:

1. For Data preparation, execute macro %Spotfire_Dataprep on SAS data sets using SAS v9.3 and above. This macro will add modified flags and common variables across the SAS data sets. These variables will be useful for filtering and marking of data.

Important Note: The macro %Spotfire_Dataprep is not part of the Preclarus® Patient Data Dashboard (Preclarus PDD) macro suite. This macro is developed specifically for this paper to serve as an example on how SAS can be leveraged for data preparation.

2. Import SAS data sets in TIBCO Spotfire and create data visualizations as per user specifications.

%SPOTFIRE_DATAPREP

For data preparation, a SAS macro called %Spotfire_Dataprep (see Appendix for detail) was developed for SAS v9.3. This macro will perform the following steps:

- Get the sort keys from the data set or assign the user provided sort keys.
- Exclude sequence based variables from current and previous datasets e.g. sequence or group variables. These variables are subject to change when the sort variables are changed in a data set.

- Compare the previous and current data sets and add the data modified flags. e.g. 'Y' if the row is new or updated.
- Merge common variables e.g. sex, race from demographics or other data sets. These variables are further used for data filters and data marking.
- Delete temporary data sets.

There are nine keyword parameters:

In: Input data set.

Out: Output data set.

PrevDS: Previous data set for comparison.

DropVars: Sequence or group variable to be dropped from comparison.

SortVars: Sort variables in case the sort keys are missing.

MergeDS: Dataset containing the common variables.

MergeVars: List of common variables.

MergeSort: Sort keys to merge the common variables.

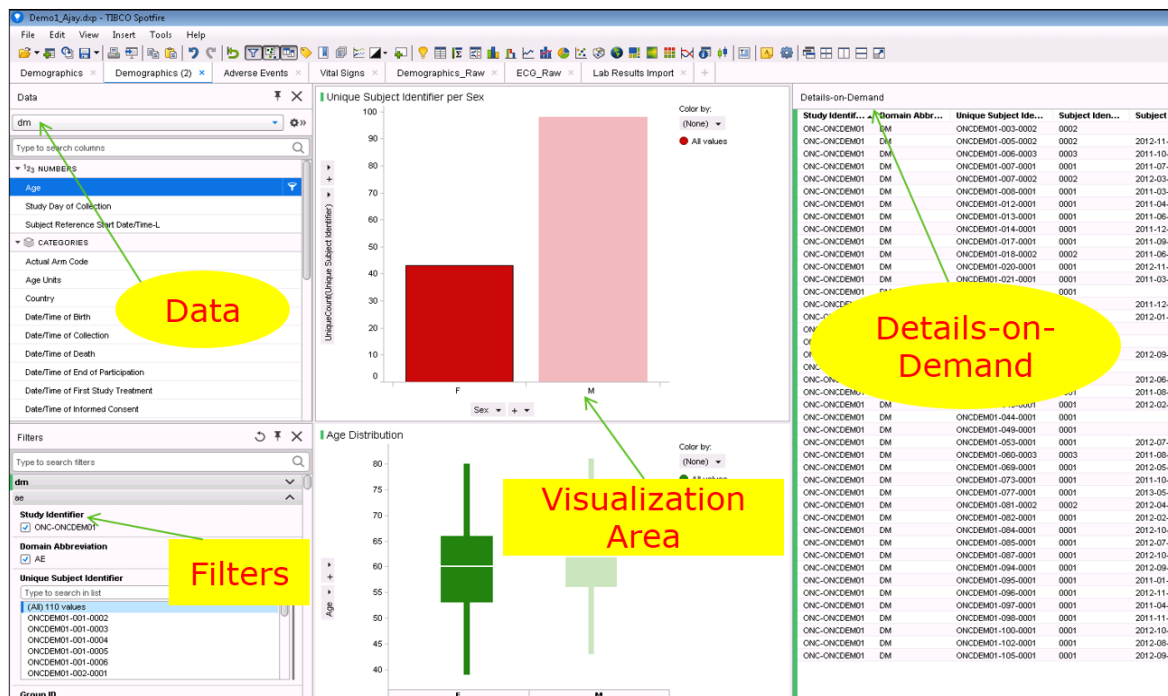
DeleteDS: Delete temporary dataset. Default value is 'Y'.

Below is the simple macro call to macro %Spotfire_Dataprep.

```
%Spotfire_Dataprep(In=AE, Out=AE_Spot, PrevDS=AE_Prev, DropVars=AESEQ,
SortVars=USUBJID AETERM, MergeDS=DM, MergeVars=SEX RACE AGE,
MergeSort=USUBJID, DeleteDS=Y)
```

TIBCO SPOTFIRE OVERALL VIEW

Below display will provide you a brief overview of the TIBCO Spotfire Development area. The development area consists of the following four main windows which can be resized as per need:



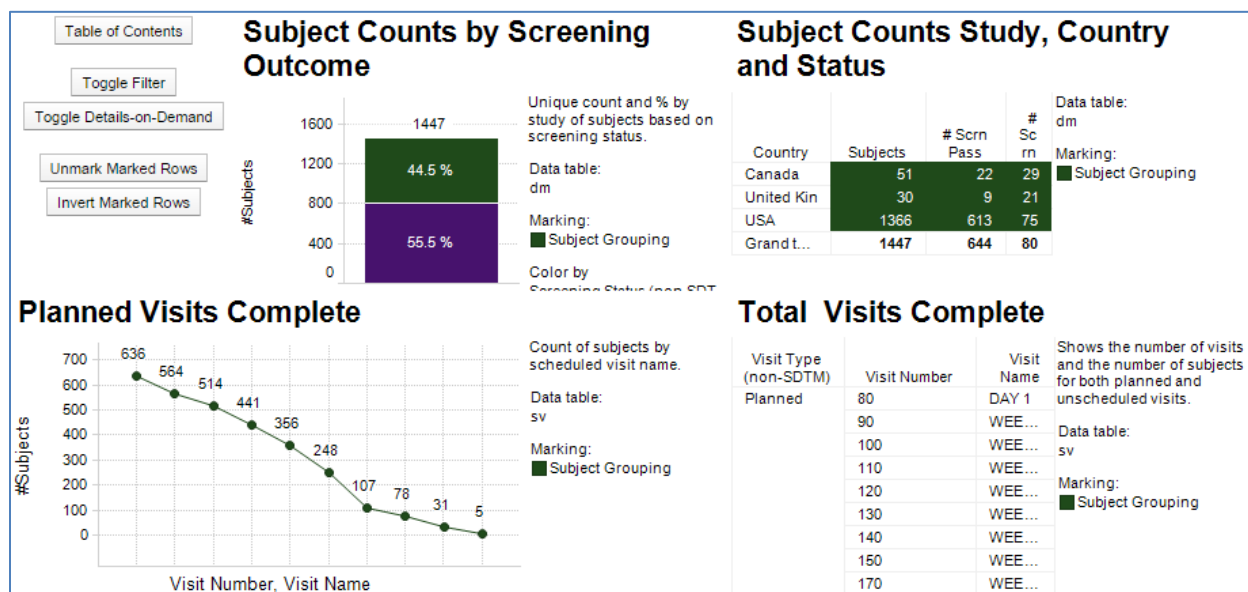
Display 1. TIBCO Spotfire Overall View

1. **Data:** This window will provide the list of all data sets and variables available for the visualization. This window can be closed from the view tab.
2. **Filters:** This window will provide the list of variables available for sub setting the data. This list includes the common variables and modified flags added by the data prep macro. This window can be closed from the view tab.
3. **Details-on-Demand:** This window will provide a data set view of selected data in the visualization. It will provide information about the data set used for the particular visualization and the list of variables available in the data set. This data can be exported into Excel or .CSV files for further evaluation. This window can be closed from the view tab.
4. **Visualization Area:** This area contains all the visualizations. Multiple visualizations (e.g. graphs, bar chart, tree map, pie chart, box plot) can be added in one tab.

DATA VISUALIZATIONS FROM PRECLARUS® PATIENT DATA DASHBOARD

Now we will go through multiple visualizations from Preclarus Patient Data Dashboard (Preclarus PDD) developed using CDISC SDTM. Users can view these visualizations using the web player where they can access the filter and data-on-demand windows but are unable to add new visualizations.

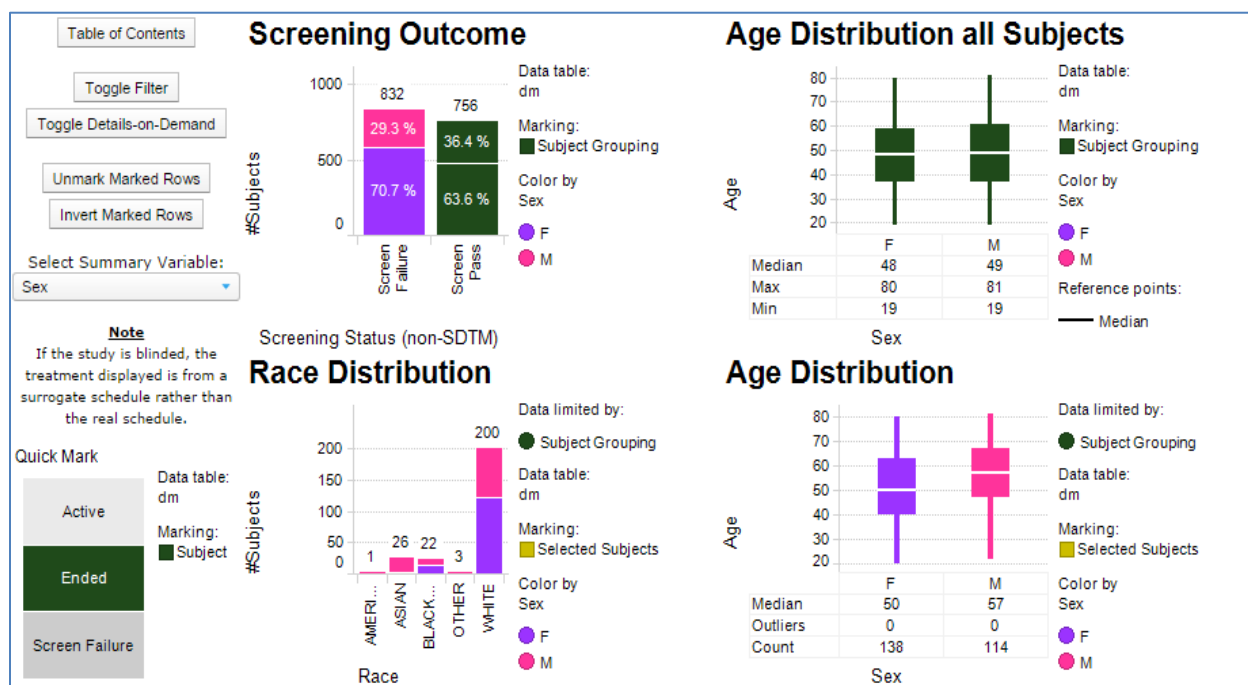
EXAMPLE 1: STUDY SUMMARY



Display 2. Study Summary Visualizations

Description:

The above interactive visualization is created using the Demographics (DM) and Subject Visits (SV) domains from the SDTM database. It consists of bar chart, line chart, and multiple tables. This visualization will give user an overall study summary e.g. planned visits complete, screening outcome, subjects counts by country and total visits complete. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab. All plots using similar datasets are interlinked together. So, if you mark/select a particular area on a particular plot then area containing the selected data on the other plot will be highlighted. For example, if you select only screen failure subjects within the bar chart then the row containing the screen failure subjects will be highlighted in the subject counts table. This visualization is very useful to understand overall study status.

EXAMPLE 2: DEMOGRAPHICS**Display 3. Visualizations for demographics data****Description:**

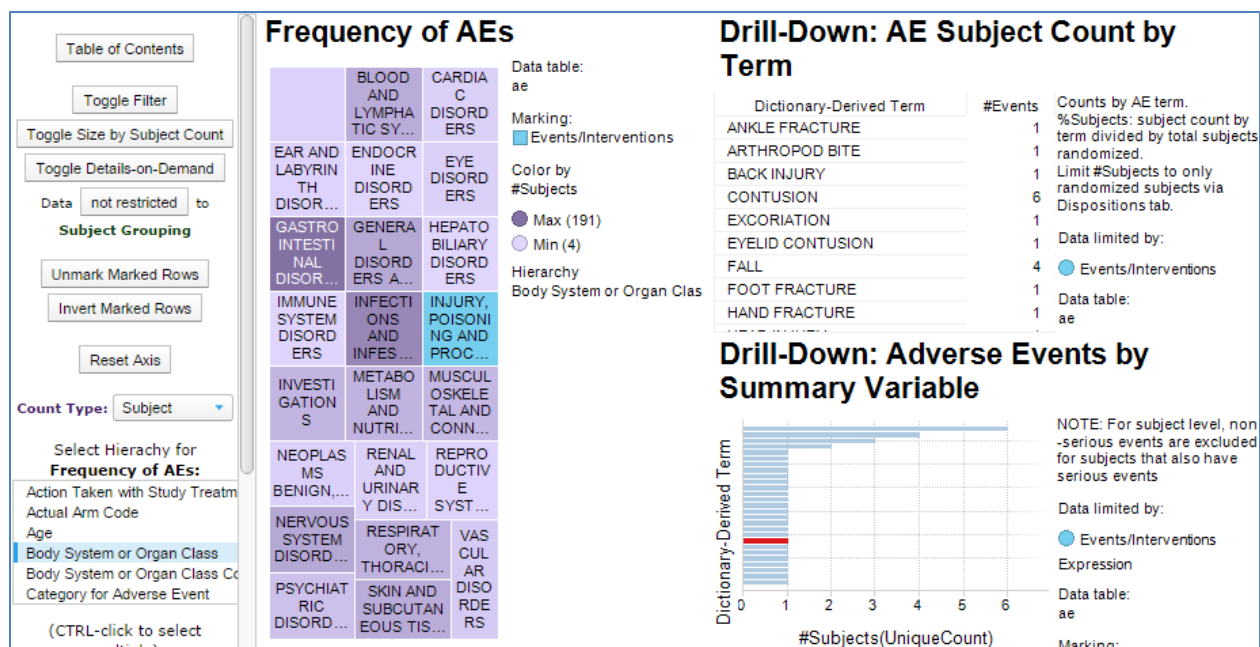
The above interactive visualization is created using the Demographics (DM) data set from the SDTM database. It consists of a box plots and multiple bar charts. This visualization will provide useful study level information about overall age distribution by gender, race distribution by gender, and screening outcome by gender. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab. The filter option within visualization or summary variable can be used to select a specific subgroup. Also, user have ability to quickly mark or unmarked the data within the visualization.

EXAMPLE 3: ADVERSE EVENTS

Adverse events (AEs) are a key focus of any safety review, for example the ability to quickly locate and isolate relevant AEs.

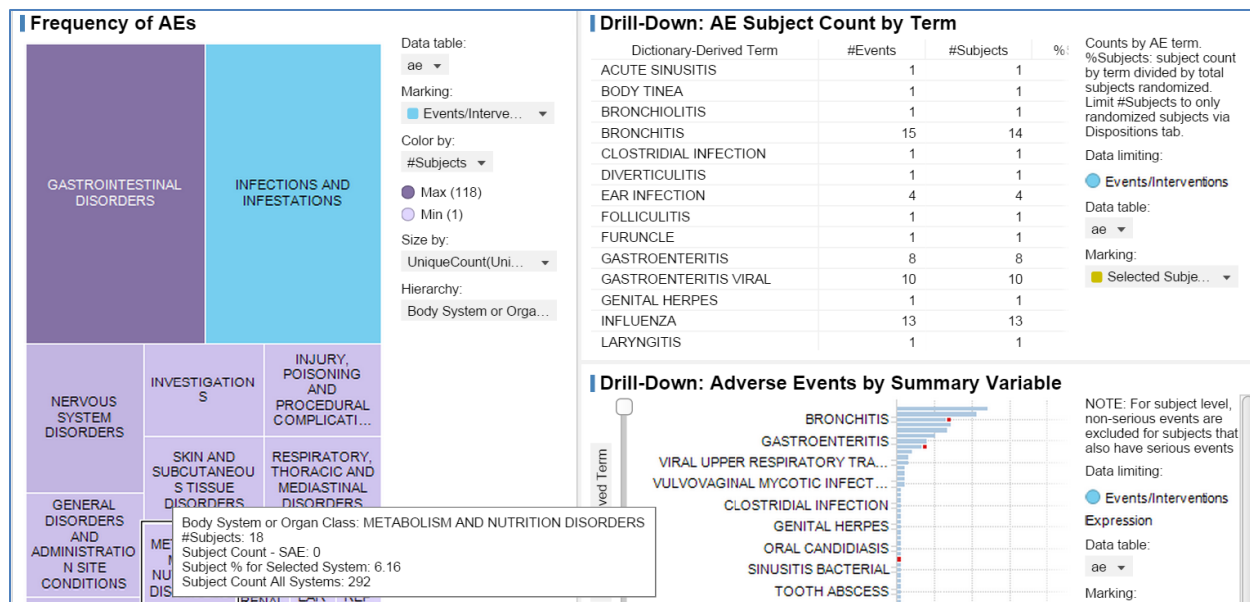
Description:

In this example, we show how a very comprehensive set of visualizations can be designed to suit any clinical trial, before the data is collected. Below is one of our AE displays, where users can select SDTM labels to populate the tree map (as in this example, grouped by "Body System or Organ Class"). In this way, users can focus on individual AE groups that are of interest to them. This exploratory graphic allows a user to start at high-level overview of adverse events and then drill-down to a patient level view.



Display 4. Visualizations for Adverse Events data

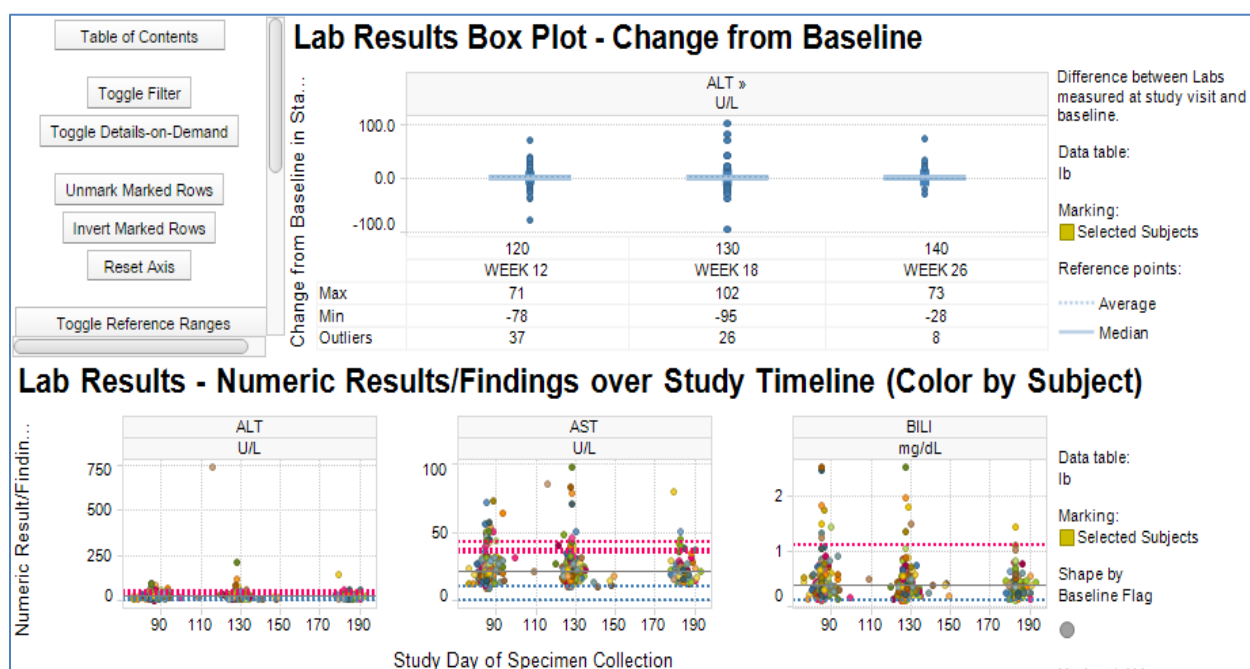
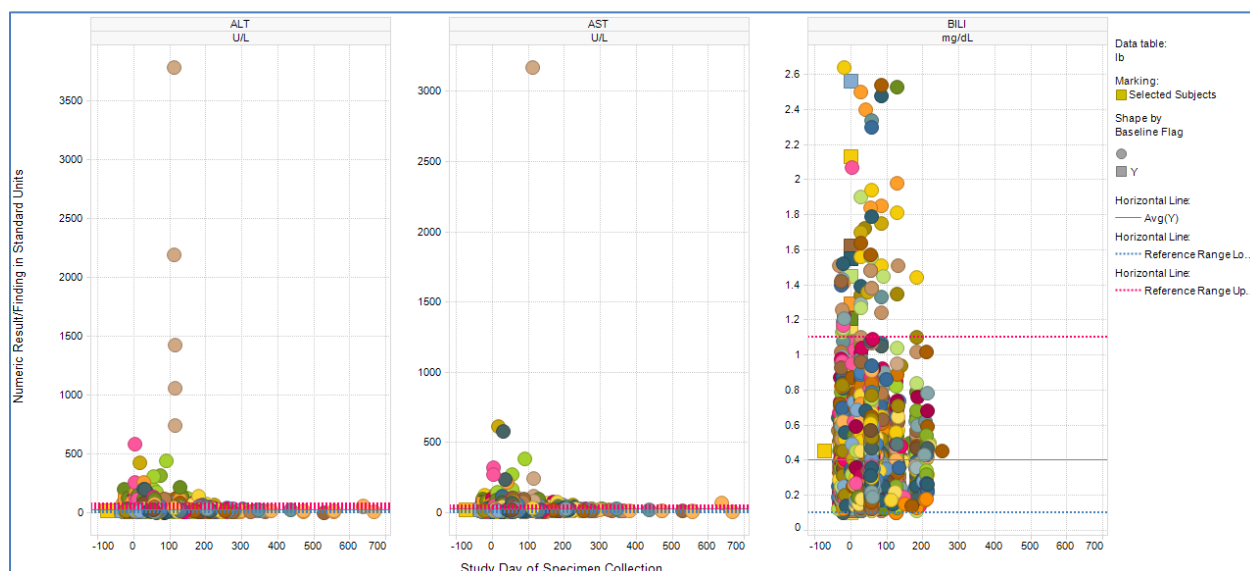
The treemap graph shown below is an interactive graph which displays a hierarchy ordered by body system or organ class. The size of the rectangles corresponds to the number of patients in a particular body system and organ class. The tooltip displays some summary information about the respective area. Further, this display also gives AE subject count by AETERM or by summary variables.



Display 5. Visualizations for Adverse Events data

EXAMPLE 4: VITAL SIGNS**Display 6. Visualizations for Vital Signs data****Display 7. Visualizations for Vital Signs data****Description:**

The above interactive visualization is created using the Vital Signs (VS) data set from SDTM database. It consists of a scatter plot of vital signs numeric results/findings over study timeline (color by subject) panel by vital signs test and units. This visualization also provides information about subject count by test and visit, average numerical test results by visit, and box plot showing the change from baseline by visits. Users can mark a particular area on the graph and access the data using the Details-on-Demand tab. Also, filters in Spotfire can be used to select a specific subgroup. This scatter plot is very useful to observe the overall picture of vital sign test results and select the outliers from the visualization for further evaluation. The tooltip displays some summary information about the respective area.

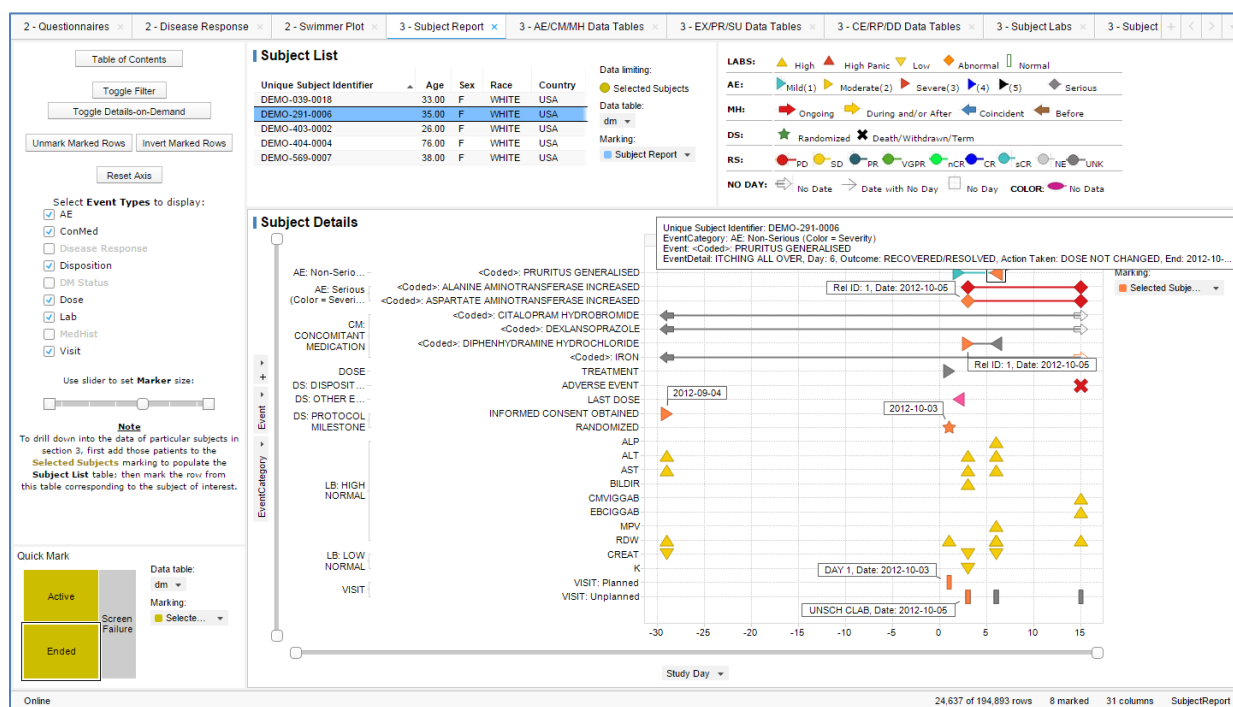
EXAMPLE 5: LABS**Display 8. Visualizations for study Lab data****Display 9. Visualizations for study Lab data****Description:**

The above interactive visualization is created using the Lab data (LB) data set from SDTM database. It consists of a scatter plot of lab numeric results/findings over study timeline (color by subject) panel by lab test and units. This visualization also provides box plot showing the change from baseline by visits. Users can mark a particular area on the graph and access the data using the Details-on-Demand tab. Also, filters in Spotfire can be used to select a specific subgroup. This scatter plot is very useful to observe the overall picture of lab data test results and select the outliers from the visualization for further evaluation. The tooltip displays some summary information about the respective area.

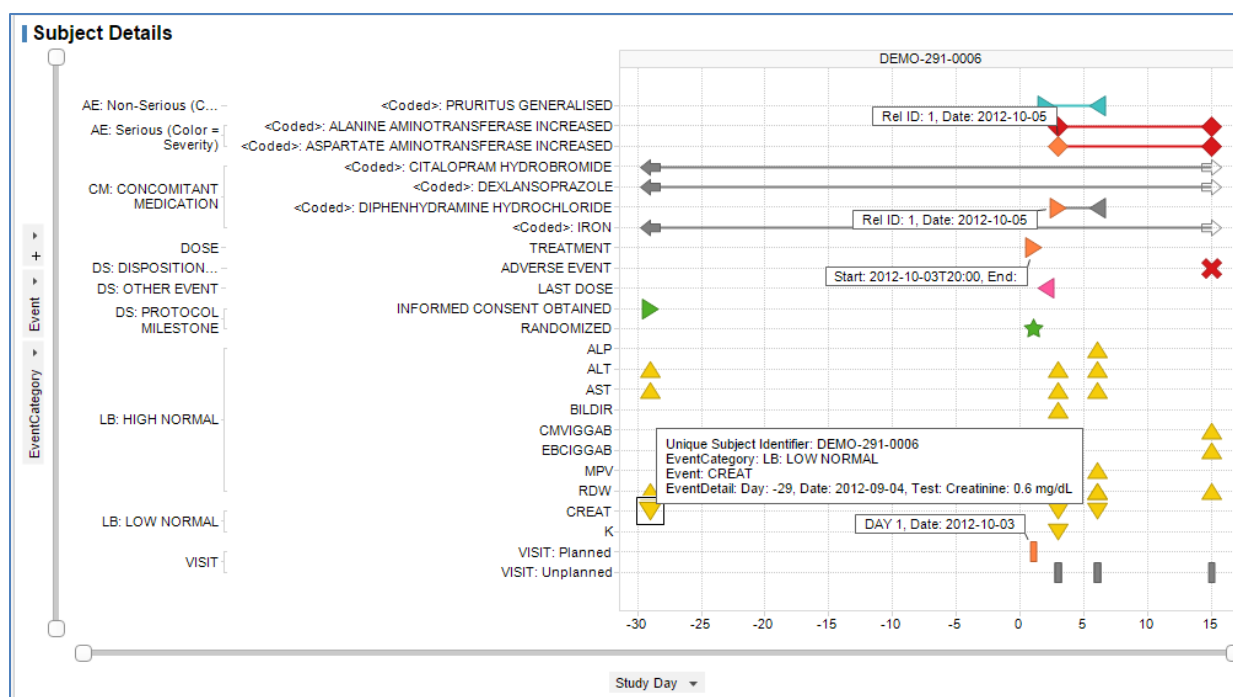
EXAMPLE 6: SUBJECT REPORTS

Description:

An important display in any dashboard is a patient profile type view where all of a subject's medical detail can be viewed together on a timeline. In this display, different areas of interest, e.g. AEs, labs, study drug, concomitant medication, visit schedule etc. can be selected at the left as required and plotted against the day on which it occurred within the subject's progress through the study. The tooltip displays some summary information about the respective area. This is a good opportunity to consider some challenges. While SDTM, and its design of the RELREC domain, make it difficult to display explicit links between study data held across different domains, the user can view potential links when different data are plotted together using universal variables, such as study day. In some cases, RELREC can be used, for example to show explicit relationships between Adverse Events and Concomitant Medications. This can be seen in this example via the marker label. The use of TIBCO Spotfire marker labels (extra information shown for elements that are clicked) and tooltips (shown when the mouse is hovered over an item) can greatly enhance the user experience.



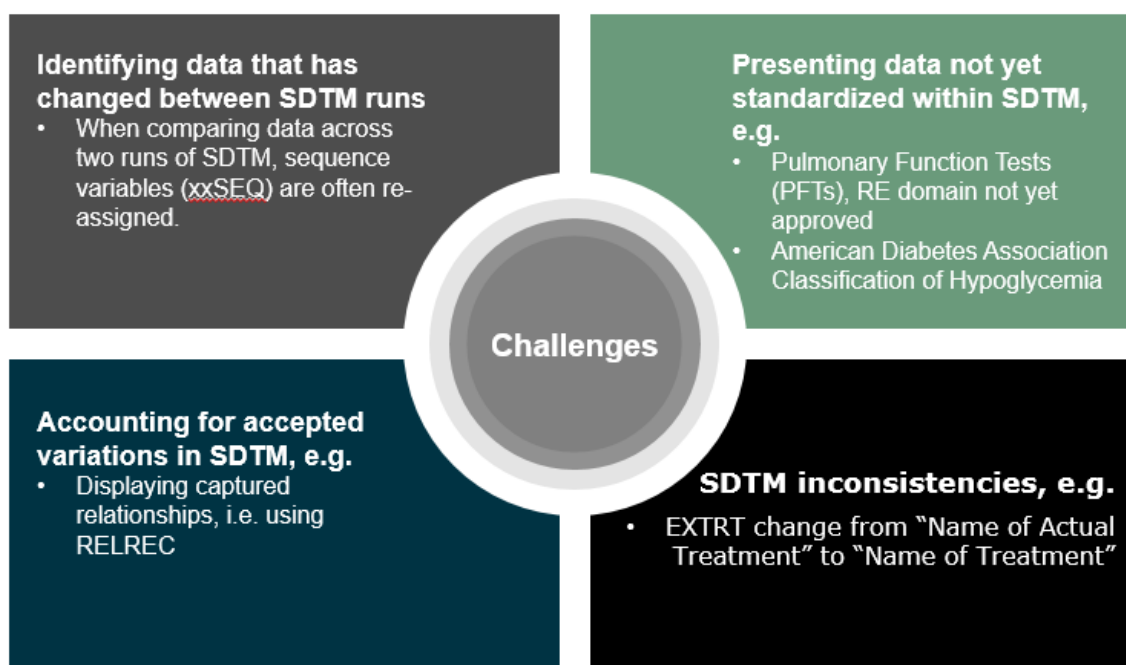
Display 10. Visualizations for subject reports



Display 11. Visualizations for study reports

CDISC SDTM CHALLENGES ENCOUNTERED

In this section, I explore some of the challenges encountered when using CDISC SDTM specifically for dashboard visualizations.



IDENTIFYING DATA THAT HAS CHANGED BETWEEN SDTM RUNS

The ability to isolate data that was previously reviewed from data that is changed or new is a common request for dashboards that are regularly updated. SDTM data sets are not designed to be compared against previous runs of SDTM for the same trial. SDTM sequence variables (xxSEQ) are often reassigned between one SDTM run and the next, so cannot be used. We have established programs to manage this, but this is a complex business. Consideration should also be given to what data the user needs to review under these circumstances, for example only the data that has changed or all data for subjects, where any of their data has changed. General solution to overcome this challenge will be to drop the xxSEQ or group variables during comparison as shown in macro %Spotfire_Dataprep.

MANAGING PERMITTED VARIATIONS IN SDTM

The algorithms that drive the Preclarus PDD need to account for any permitted variations in the CDISC SDTM standard that may be encountered in the data sets. RELREC is a good example of where the standards allow multiple variations. Even in the relatively simple relationship between Adverse Events and Concomitant Medications, this can become vastly complicated, for example where xxGRPID variables are used. One answer is to set the dashboard to exclude features, where the data does not conform. This is a much better approach than the dashboard failing or even worse, surfacing erroneous data.

PRESENTING DATA NOT YET STANDARDISED WITHIN SDTM

There are various additions that we would like to add that are not yet defined within CDISC SDTM. Examples include domains that are not yet approved, such as Respiratory System Findings (RE). This is where studies would capture Pulmonary Function Test (PFT) data. Another example are standards required for capturing the American Diabetes Association Classification of Hypoglycemia. Gaps such as these, mean that data is recorded anywhere within SDTM, so we cannot locate the relevant data, to display, via our algorithms. To overcome this challenge, we request user to map the data into existing domains where possible. Also, we add new domains in our dashboard on regular basis.

SDTM INCONSISTENCIES

Our system is designed to accommodate different versions of SDTM, which is largely successful due to the excellent standard management incorporated into CDISC SDTM. Care does need to be taken however, as small adjustments between versions can cause problems for algorithms. For example, the variable EXTRT, in the Exposure (EX) domain, had a label change between SDTM v3.1.3 and v3.2 from "Name of Actual Treatment" to "Name of Treatment". In another example, the QSSTRESC label, "Character Result/Finding in Standard Format", in the Questionnaires (QS) domain, breaks the 40-character rule. This means that many studies will assign an alternative label. In such cases, it is important to ensure that the dashboard continues to locate the data. It is important to make the algorithms robust to such variations by looking for and recognizing these cases. In general, to overcome this challenge user can update the labels in macro %Spotfire_Dataprep and make it dashboard compliant.

CONCLUSION

TIBCO Spotfire provides an interactive platform for exploratory analysis using CDISC SDTM data for clinical trials. With its simplicity to adjust axes symbols and text, and its ability to export data for further user analysis/query, TIBCO Spotfire enables faster data review, quality assessment and process improvement. This paper shows how far CDISC SDTM has come and how it can be effectively utilized today to greatly enhance the monitoring and review experience on clinical trials. Also, TIBCO Spotfire saves time consumed through a traditional adhoc process of creating statistical graphics. While still following a standard process including on-demand development, quality review and final production, TIBCO Spotfire leaves the feasibility for modification as per customer's need.

REFERENCES

Gupta Ajay, 2017. Clinical Data Visualization using TIBCO Spotfire® and SAS®. Proceedings of the PharmaSUG 2017 Conference, paper BB18.

Bartle Philip, 2017. Visualising CDISC SDTM for Monitoring and Review. Proceedings of the PhUSE 2017 Conference, paper DV05.

<http://spotfire.tibco.com/>

<http://www.sascommunity.org>

ACKNOWLEDGMENTS

Thanks to Ryan Wilkins, Lindsay Dean, Philip Bartle, Tammy Jackson, Ken Borowiak, Richard DAmato, Lynn Clipstone, and PPD Management for their reviews and comments. Thanks to my family for their support.

CONTACT INFORMATION

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

Ajay Gupta, M.S.
PPD
3900 Paramount Parkway
Morrisville, NC 27560
Work Phone: (919)-456-6461
Fax: (919) 654-9990
E-mail: Ajay.Gupta@ppdi.com,
Ajaykailasgupta@aol.com



DISCLAIMER

The content of this paper are the works of the authors and do not necessarily represent the opinions, recommendations, or practices of PPD.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX I

```
%MACRO SPOTFIRE_DATAPREP (IN=, OUT=, PREVDS=, DROPVARS=, SORTVARS=, MERGEDS=,
                        MERGEVARS=, MERGESORT=, DELETEDS=Y);

/*Get Sort Variable from input dataset*/
PROC CONTENTS data=&in out=contents (KEEP=name sortedby
                        WHERE=(SORTEDBY ~=.)) NOPRINT;

RUN;

DATA _null_;
    CALL SYMPUTX ('obs', nobs);
    IF 0 THEN SET contents NOBS=nobs;
    STOP;
RUN;

%IF &nobs=0 %THEN %DO;
    /*If no sort done then set to sort provided by user*/
    %LET sort_v=&sortvars;
%END;

%ELSE %DO;
    PROC SORT DATA=contents ;
        BY sortedby;
    RUN;
    PROC SQL NOPRINT;
        SELECT name INTO: sort_v SEPARATED BY ' ' FROM contents
            ORDER BY sortedby
        ;
    QUIT;
%END;

/*Set input dataset*/
DATA &out;
    SET &in;
RUN;

/*Exclude sequence variable from datasets for compare*/
%IF &PrevDS~="" %THEN %DO;
    DATA In_1;
        SET &out;
        DROP &dropvars;
    RUN;
    DATA PrevDS_1;
```

```

        SET &PrevDS;
        DROP &dropvars;
RUN;
/*Compare Previous and New Dataset*/
PROC SQL NOPRINT;
    CREATE table In_2 AS
        SELECT * FROM in_1
        EXCEPT
        SELECT * FROM PrevDS_1
    ;
QUIT;
PROC SQL NOPRINT;
    CREATE TABLE In_3 AS
        SELECT * FROM In_1
        EXCEPT
        SELECT * FROM In_2
    ;
QUIT;
/*Add modified flag*/
DATA In_4(KEEP=&sort_v modified_flag);
    SET In_2 (IN=ina) In_3;
    IF ina THEN modified_flag="Y";
RUN;
PROC SORT DATA=In_4;
    BY &sort_v;
RUN;
PROC SORT DATA=&out.;
    BY &sort_v;
RUN;
DATA &out;
    MERGE &out(IN=a) in_4;
    BY &sort_v;
    IF a;
RUN;
%END;
%IF "&mergeds"~="" %THEN %DO;
    /*Merge Common Variables for filter*/

```

```
PROC SORT DATA=&out;
  BY &mergesort;
RUN;
PROC SORT DATA=&mergeds(KEEP= &mergesort &mergevars);
  BY &mergesort;
RUN;
DATA &out;
MERGE &out(IN=a) &mergeds;
  BY &mergesort;
  IF a;
RUN;
%END;
/*Delete Tmp datasets*/
  %IF &DeletedDS=Y %THEN %DO;
    PROC DATASETS LIB=work KILL NOLIST;
    QUIT;
    RUN;
  %END;
%MEND;
```