Advanced Data Visualization using TIBCO Spotfire® and SAS® using SDTM

Ajay Gupta, PPD
INTRODUCTION

+ 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 dynamic 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 presentation will demonstrate this, including examples of advanced visualizations from Preclarus® Patient Data Dashboard within PPD® created using TIBCO Spotfire and SAS and share our experiences and challenges while creating this dashboard.
The general process of creating data visualizations in TIBCO Spotfire is as follows:

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

+ Import SAS datasets in TIBCO Spotfire and create data visualizations as per user specifications.
TIBCO SPOTFIRE OVERALL VIEW

Data

Details-on-Demand

Filters

Visualization Area
TIBCO Spotfire Development area consists of the following four main windows which can be resized as needed:

+ **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.

```
<table>
<thead>
<tr>
<th>Data</th>
<th>dim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type to search columns</td>
<td></td>
</tr>
<tr>
<td>NUMBERS</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Study Day of Collection</td>
<td></td>
</tr>
<tr>
<td>Subject Reference Start Date/Time-L</td>
<td></td>
</tr>
<tr>
<td>CATEGORIES</td>
<td></td>
</tr>
<tr>
<td>Actual Arm Code</td>
<td></td>
</tr>
<tr>
<td>Age Units</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td></td>
</tr>
<tr>
<td>Date/Time of Birth</td>
<td></td>
</tr>
<tr>
<td>Date/Time of Collection</td>
<td></td>
</tr>
<tr>
<td>Date/Time of Death</td>
<td></td>
</tr>
<tr>
<td>Date/Time of End of Participation</td>
<td></td>
</tr>
<tr>
<td>Date/Time of First Study Treatment</td>
<td></td>
</tr>
</tbody>
</table>
```
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.
(Cont’d.)

+ 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.
(Cont’d.)

+ **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.
PRECLARUS® EXAMPLE 1: STUDY SUMMARY

Subject Counts by Screening Outcome

Subject Counts Study, Country and Status

Planned Visits Complete

Total Visits Complete

HELPING DELIVER LIFE-CHANGING THERAPIES
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 (select a section of a visualization) a particular area on the graph and access the data using the Details-on-Demand tab.

All plots using similar datasets are interlinked together. 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.
PRECLARUS® EXAMPLE 2: DEMOGRAPHICS

Screening Outcome
- Screen Failure: 803
- Screen Pass: 644
- Data table: dm
- Marking: Subject Grouping
- Color by Sex

Age Distribution all Subjects
- Median: F 47, M 42
- Max: F 79, M 75
- Min: F 19, M 19
- Data table: dm
- Marking: Subject Grouping
- Color by Sex

Screening Status (non-SDTM)
- Active: 1
- Ended: 4
- Data table: dm
- Marking: Subject Grouping

Race Distribution
- Data limited by: Subject Grouping
- Color by Sex
- American: 115
- Asian: 3
- Black: 4
- Other: 20
- White: 3

Age Distribution
- Data limited by: Subject Grouping
- Color by Sex
- Median: F 46, M 43
- Outliers: F 0, M 2
- Count: F 102, M 41
- Reference points: Median

Note: If the study is blinded, the treatment displayed is from a surrogate schedule rather than the actual treatment.

Quick Mark
- Screen Failure
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. 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.
PRECLARUS® EXAMPLE 3: ADVERSE EVENTS
(Cont’d.)

+ Above 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 study adverse events and then drill-down to a patient level view.
PRECLARUS® EXAMPLE 4: Vital Signs

Vital Signs Results Box Plot - Change from Baseline

Vital Signs Tests

Visit No... Visit Name Vital Signs Vital Signs Test Short... Subject Count
0 UNSCH V5 SY/SBP 3
160 CYCLE 3 SY/SBP vs DIABP vs 3
17001 CYCLE 4 SY/SBP vs DIABP vs 3
180 CYCLE 5 SY/SBP vs DIABP vs 3

Vital Signs Tests

Data table: vs
Marking: Selected Subjects

Vital Signs Results - Numeric Results/Findings over Study Timeline (Color by Subject)

Study Day of Vital Signs
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.

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.
PRECLARUS® EXAMPLE 5: LABS

Lab Results Box Plot - Change from Baseline

Lab Results - Numeric Results/Findings over Study Timeline (Color by Subject)
(Cont’d.)
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.

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 and/or query.
PRECLARUS® EXAMPLE 6: SUBJECT REPORTS
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. 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.
CDISC SDTM CHALLENGES ENCOUNTERED

Identifying data that has changed between SDTM runs
- When comparing data across two runs of SDTM, sequence variables (xxSEQ) are often re-assigned.

Presenting data not yet standardized within SDTM, e.g.
- Pulmonary Function Tests (PFTs), RE domain not yet approved
- American Diabetes Association Classification of Hypoglycemia

Accounting for accepted variations in SDTM, e.g.
- Displaying captured relationships, i.e. using RELREC

SDTM inconsistencies, e.g.
- EXTRT change from “Name of Actual Treatment” to “Name of Treatment”
IDENTIFYING DATA THAT HAS CHANGED BETWEEN SDTM RUNS

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. 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.
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 STANDARDIZED 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). To overcome this challenge, we request user to map the data into existing domains were possible. Also, we add new domains in our dashboard on regular basis.
+ **SDTM INCONSISTENCIES**

+ 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”. 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. as data, in Spotfire, are located by label.

+ In general, to overcome this challenge user can update the labels in macro %Spotfire_Dataprep and make it dashboard compliant.
SUMMARY

+ TIBCO Spotfire provides an interactive platform for exploratory analysis.
+ With its simplicity to adjust axes symbols and text, able to co-relate sources of data to subset quicker/on the fly, and its ability to export data for further user analysis/query, TIBCO Spotfire enables faster data review, quality assessment and process improvement.
+ Also, TIBCO Spotfire saves time consumed through a traditional ad-hoc 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.
QUESTIONS?

Contact Author:

Ajay Gupta
PPD
Ajay.Gupta@ppdi.com

DISCLAIMER

The contents of this paper are the work of the author and do not necessarily represent the opinions, recommendations, or practices of PPD.