



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



Ajay Gupta, PPD

HELPING DELIVER LIFE-CHANGING THERAPIES



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.

TECHNIQUE & MECHANISM

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)
```

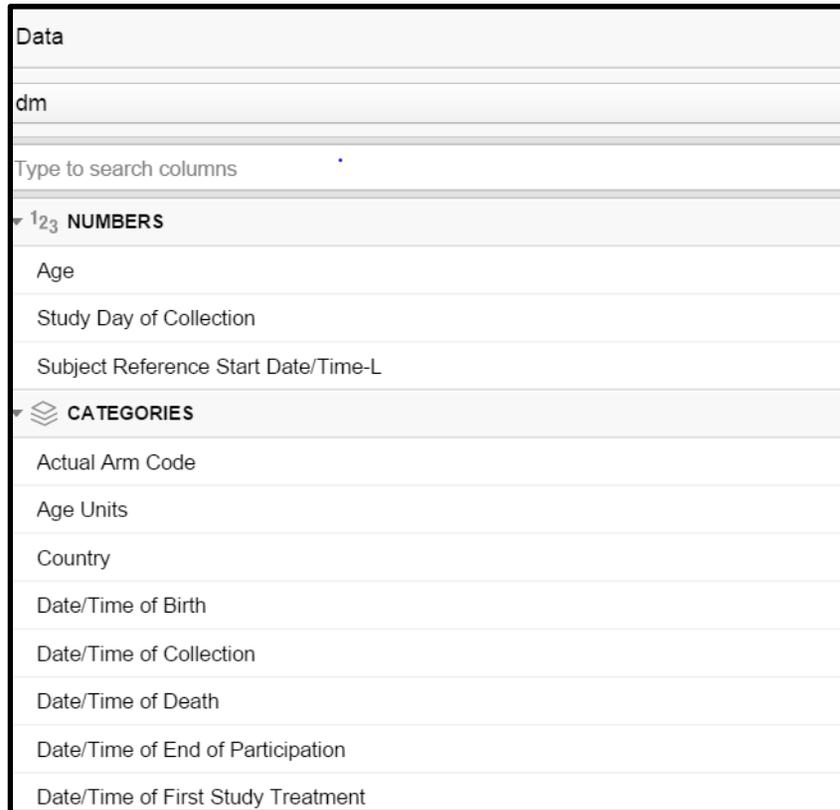
- + Import SAS datasets in TIBCO Spotfire and create data visualizations as per user specifications.

TIBCO SPOTFIRE OVERALL VIEW

The screenshot displays the TIBCO Spotfire interface with the following components:

- Data Panel:** Located on the top left, it shows a search for 'dm' and a list of columns under 'NUMBERS' and 'CATEGORIES'. A yellow oval labeled 'Data' points to this panel.
- Filters Panel:** Located on the bottom left, it shows filters for 'Study Identifier', 'Domain Abbreviation', and 'Unique Subject Identifier'. A yellow oval labeled 'Filters' points to the 'Study Identifier' filter.
- Unique Subject Identifier per Sex Chart:** A bar chart showing the count of unique subject identifiers for 'F' (Female) and 'M' (Male). The 'M' bar is significantly higher. A yellow oval labeled 'Visualization Area' points to the chart.
- Age Distribution Chart:** A box plot showing the distribution of age for 'F' and 'M' groups. A yellow oval labeled 'Visualization Area' points to this chart.
- Details-on-Demand Table:** Located on the right side, it displays a table with columns: Study Identif..., Domain Abbr..., Unique Subject Ide..., Subject Id..., and Subject R... A yellow oval labeled 'Details-on-Demand' points to the table.

TIBCO SPOTFIRE DEVELOPMENT AREA

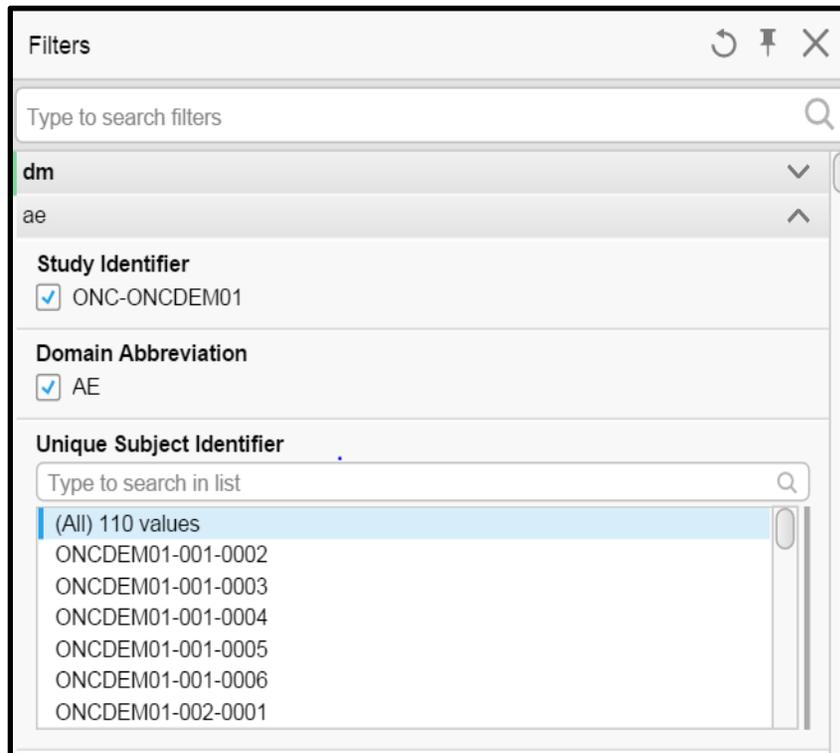


Data	
dm	
Type to search columns	
123	NUMBERS
Age	
Study Day of Collection	
Subject Reference Start Date/Time-L	
+	CATEGORIES
Actual Arm Code	
Age Units	
Country	
Date/Time of Birth	
Date/Time of Collection	
Date/Time of Death	
Date/Time of End of Participation	
Date/Time of First Study Treatment	

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.

(Cont'd.)



+ **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.

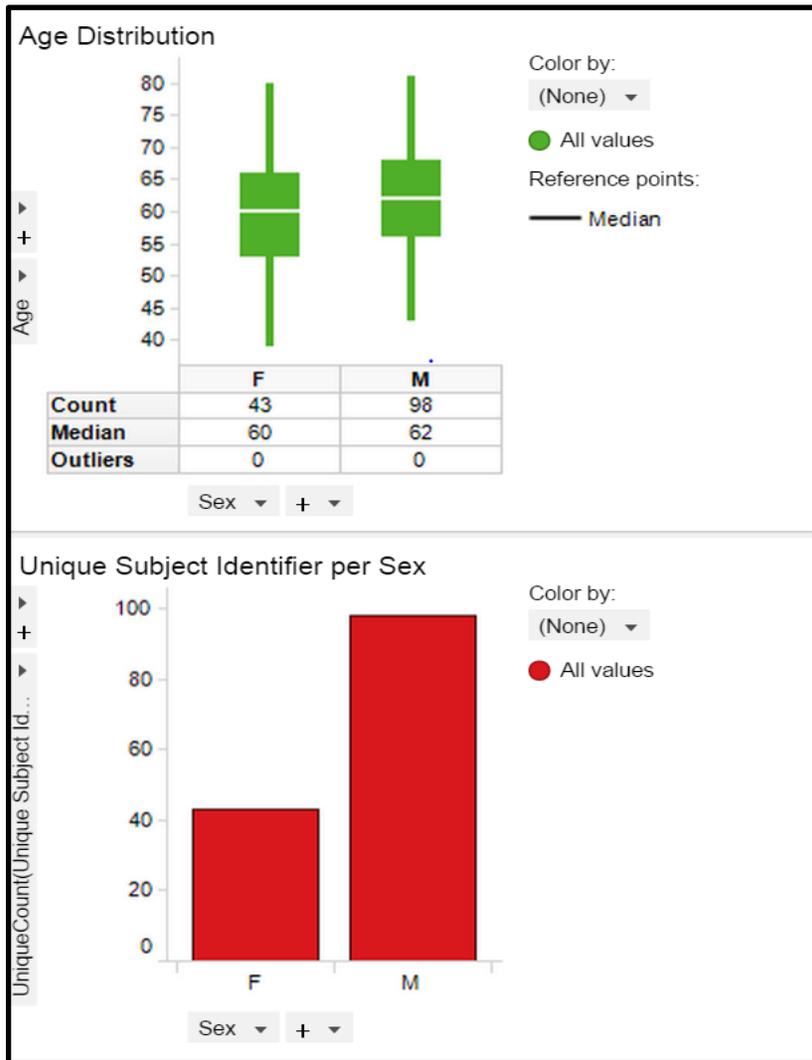
(Cont'd.)

Study Identifier ▲	Domain Abbr...	Unique Subject Iden...	Subject Ident...
ONC-ONCDE...	DM	ONCDEM01-001-0001	0001
ONC-ONCDE...	DM	ONCDEM01-002-0001	0001
ONC-ONCDE...	DM	ONCDEM01-004-0001	0001
ONC-ONCDE...	DM	ONCDEM01-011-0001	0001
ONC-ONCDE...	DM	ONCDEM01-012-0001	0001
ONC-ONCDE...	DM	ONCDEM01-013-0001	0001
ONC-ONCDE...	DM	ONCDEM01-014-0001	0001
ONC-ONCDE...	DM	ONCDEM01-015-0001	0001
ONC-ONCDE...	DM	ONCDEM01-017-0001	0001
ONC-ONCDE...	DM	ONCDEM01-018-0001	0001
ONC-ONCDE...	DM	ONCDEM01-019-0001	0001
ONC-ONCDE...	DM	ONCDEM01-020-0001	0001

+ **Details-on-Demand:**

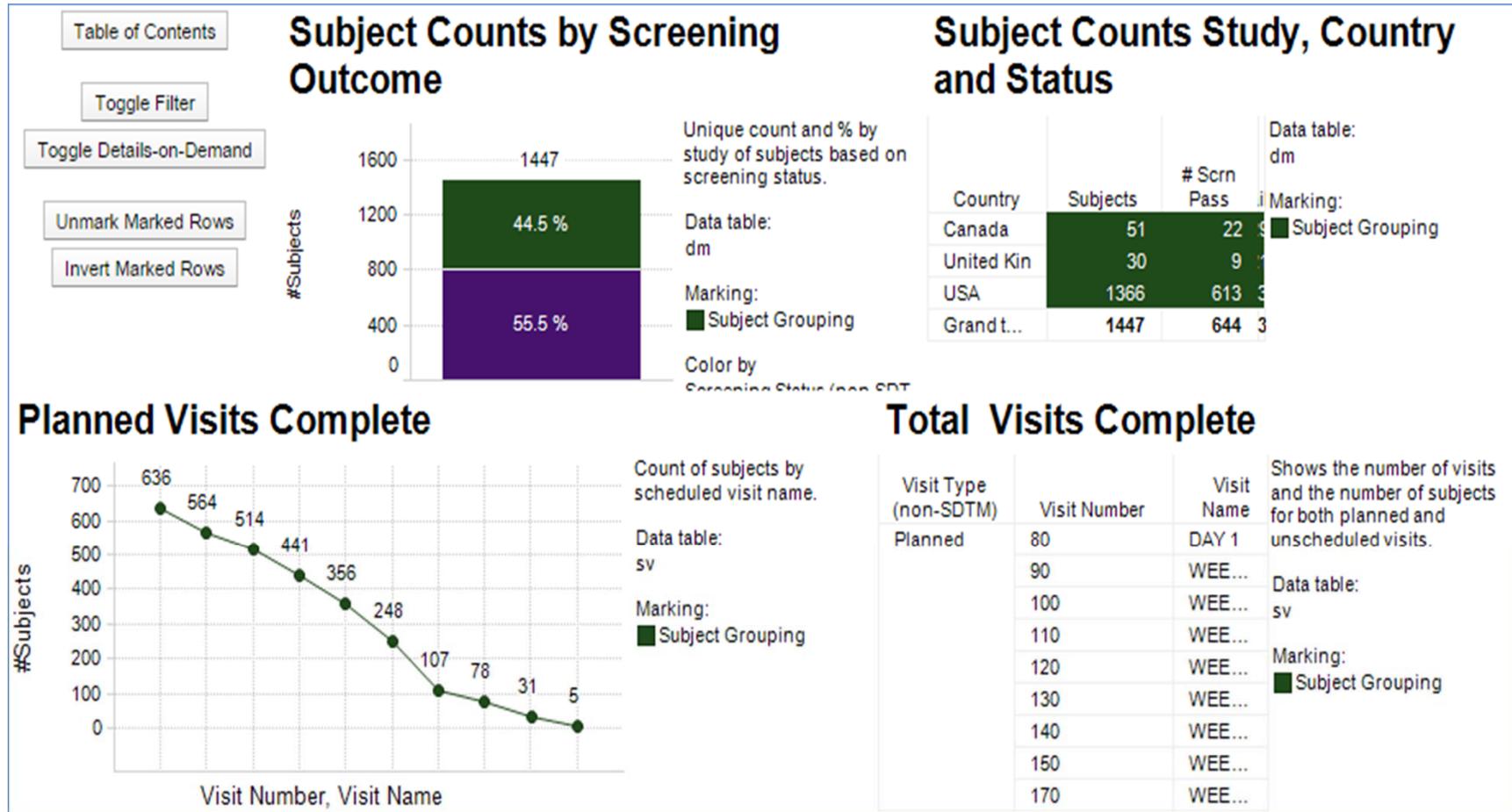
This window will provide a data set view of selected data in the visualization. It will provide information about the data set used for the particular visualization and the list of variables available in the data set. This data can be exported into Excel or .CSV files for further evaluation

(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.

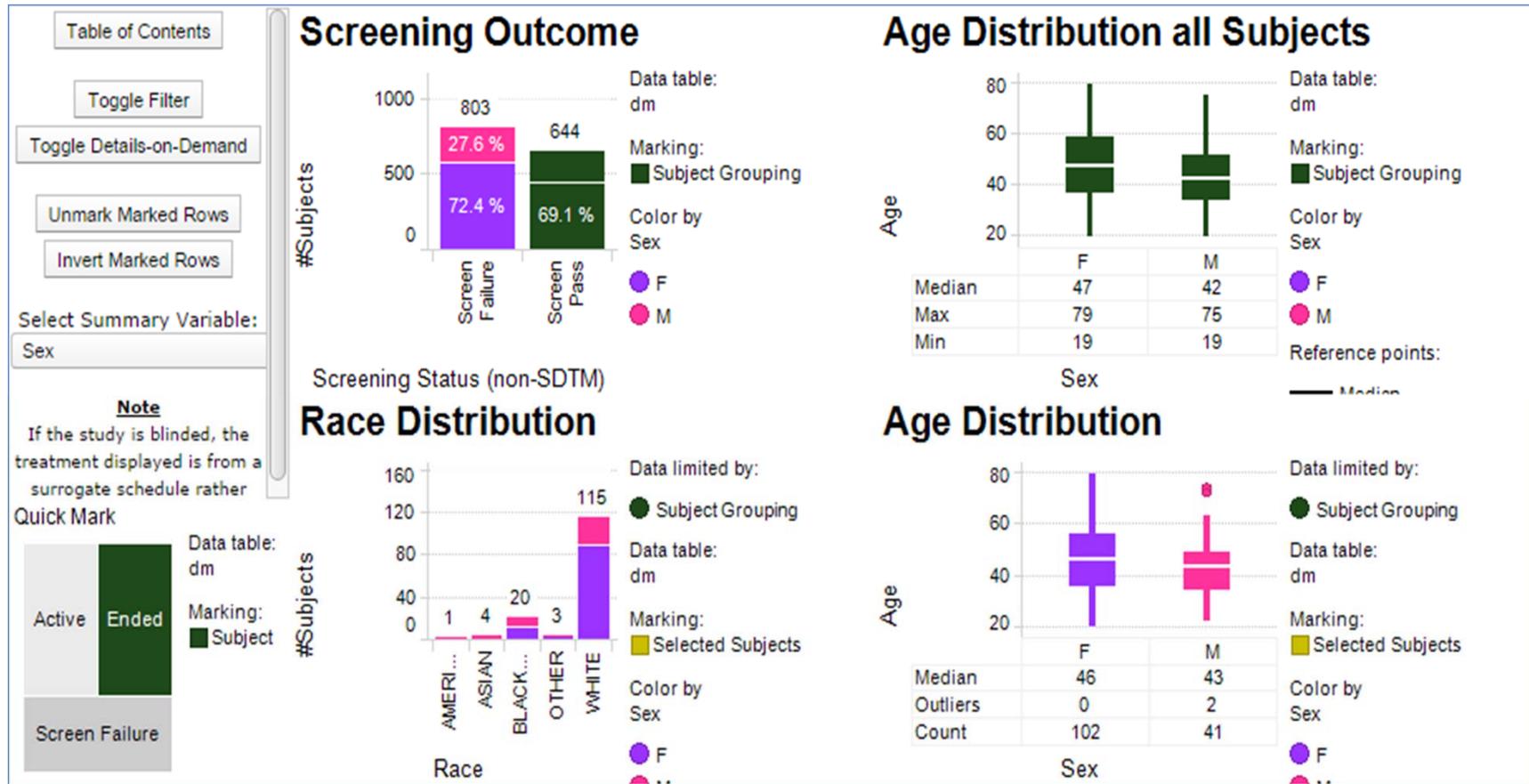
PRECLARUS® EXAMPLE 1: STUDY SUMMARY



(Cont'd.)

- + 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



(Cont'd.)

- + 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

Table of Contents

Toggle Filter

Toggle Size by Subject Count

Toggle Details-on-Demand

Data not restricted to

Subject Grouping

Unmark Marked Rows

Invert Marked Rows

Reset Axis

Count Type: Subject

Select Hierarchy for Frequency of AEs:

- Action Taken with Study Treatment
- Actual Arm Code
- Age
- Body System or Organ Class
- Body System or Organ Class Code
- Category for Adverse Event

Frequency of AEs

	BLOOD AND LYMPH...	CARDIAC DISOR...
EAR AND LABYRINTH DIS...	ENDOCRINE DISOR...	EYE DISORDERS
GASTROINTESTINAL DI...	GENERAL DISOR...	IMMUNE SYSTEM DISOR...
INFECTIOUS AND IN...	INJURY, POISONING AN...	INVESTIGATIONS
METABOLISM AND N...	MUSCULOSKELETAL AN...	NEOPLASMS BENIG...
NERVOUS SYSTEM DIS...	RENAL AND URINAR...	REPRODUCTIVE SYSTE...
PSYCHIATRIC DISORDERS	RESPIRATORY, THORACI...	SKIN AND SUBCUTANEOUS
		VASCULAR DISORDERS

Data table: ae

Marking: Events/Interventions

Color by #Subjects

Max (118)

Min (1)

Hierarchy: Body System or Organ Class

Drill-Down: AE Subject Count by Term

Dictionary-Derived Term	#Events
ASTHENIA	1
CHEST DISCOMFORT	1
CHEST PAIN	1
CHILLS	2
DRUG INTOLERANCE	1
FATIGUE	8
HUNGER	1
INFLUENZA LIKE ILLNESS	3

Counts by AE term. %Subjects: subject count by term divided by total subjects randomized.

Limit #Subjects to only randomized subjects via Dispositions tab.

Data limited by: Events/Interventions

Data table: ae

Drill-Down: Adverse Events by Summary Variable

NOTE: For subject level, non-serious events are excluded for subjects that also have serious events

Data limited by: Expression

Data table: ae

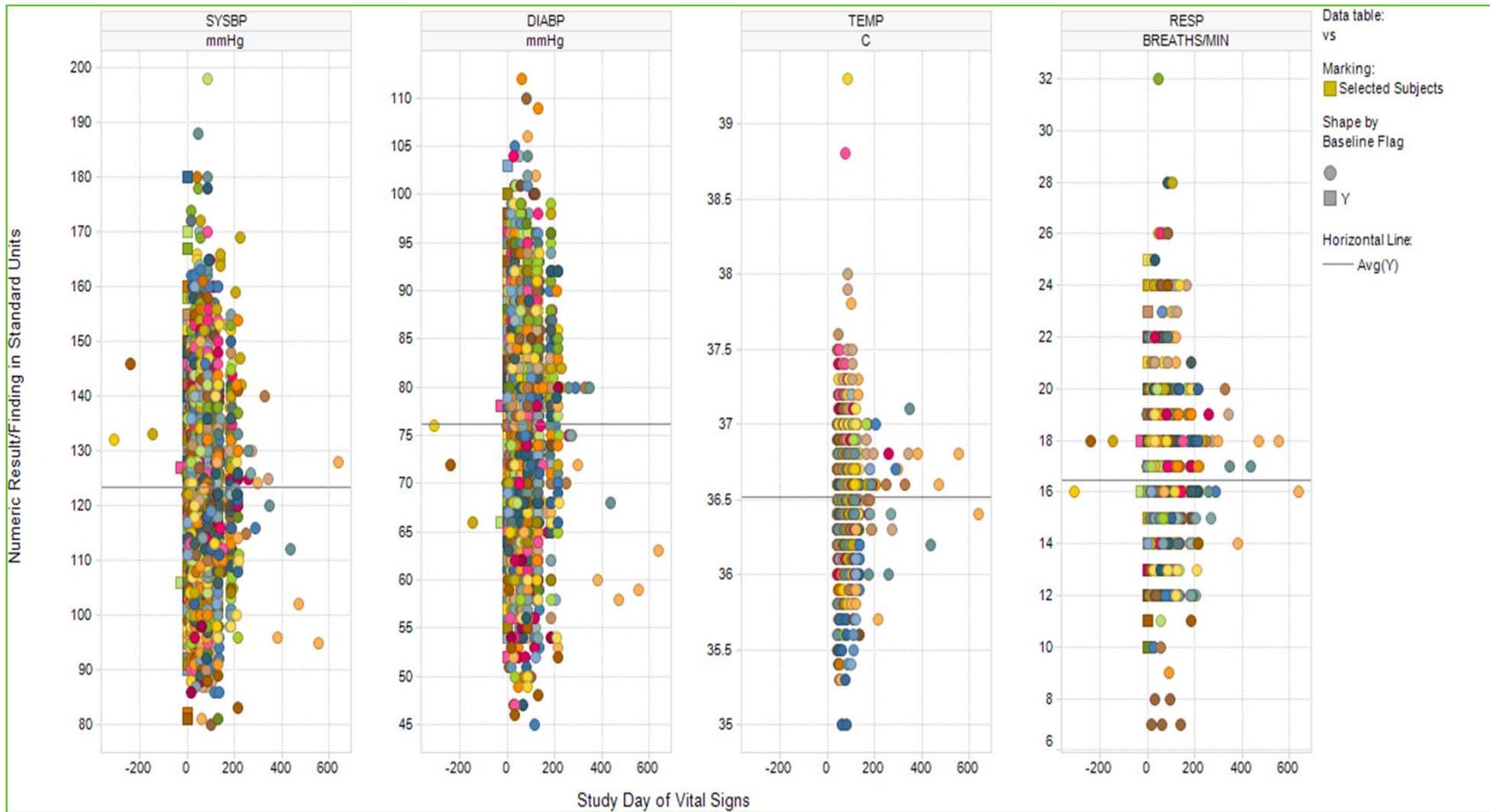
(Cont'd.)

- + 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



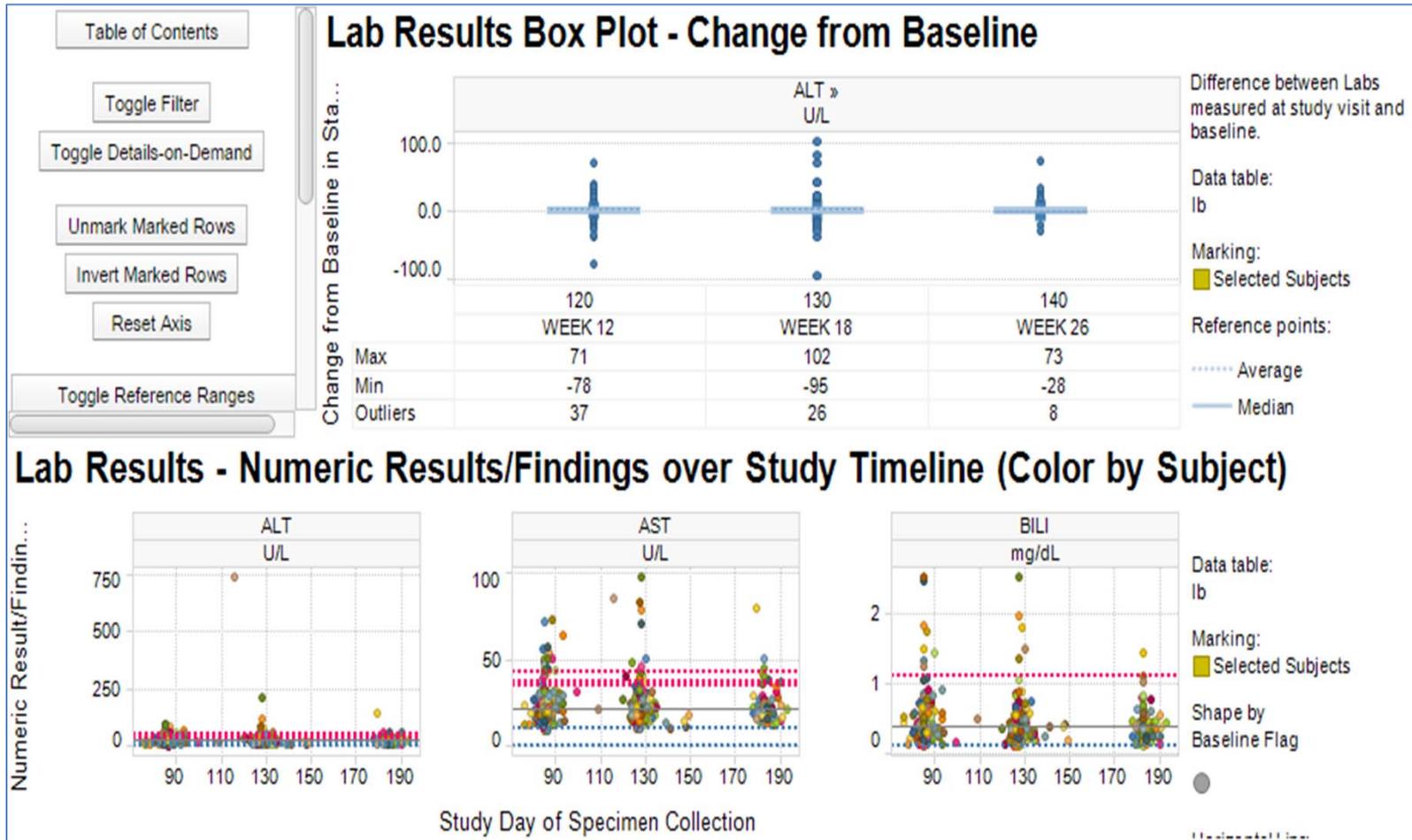
(Cont'd.)



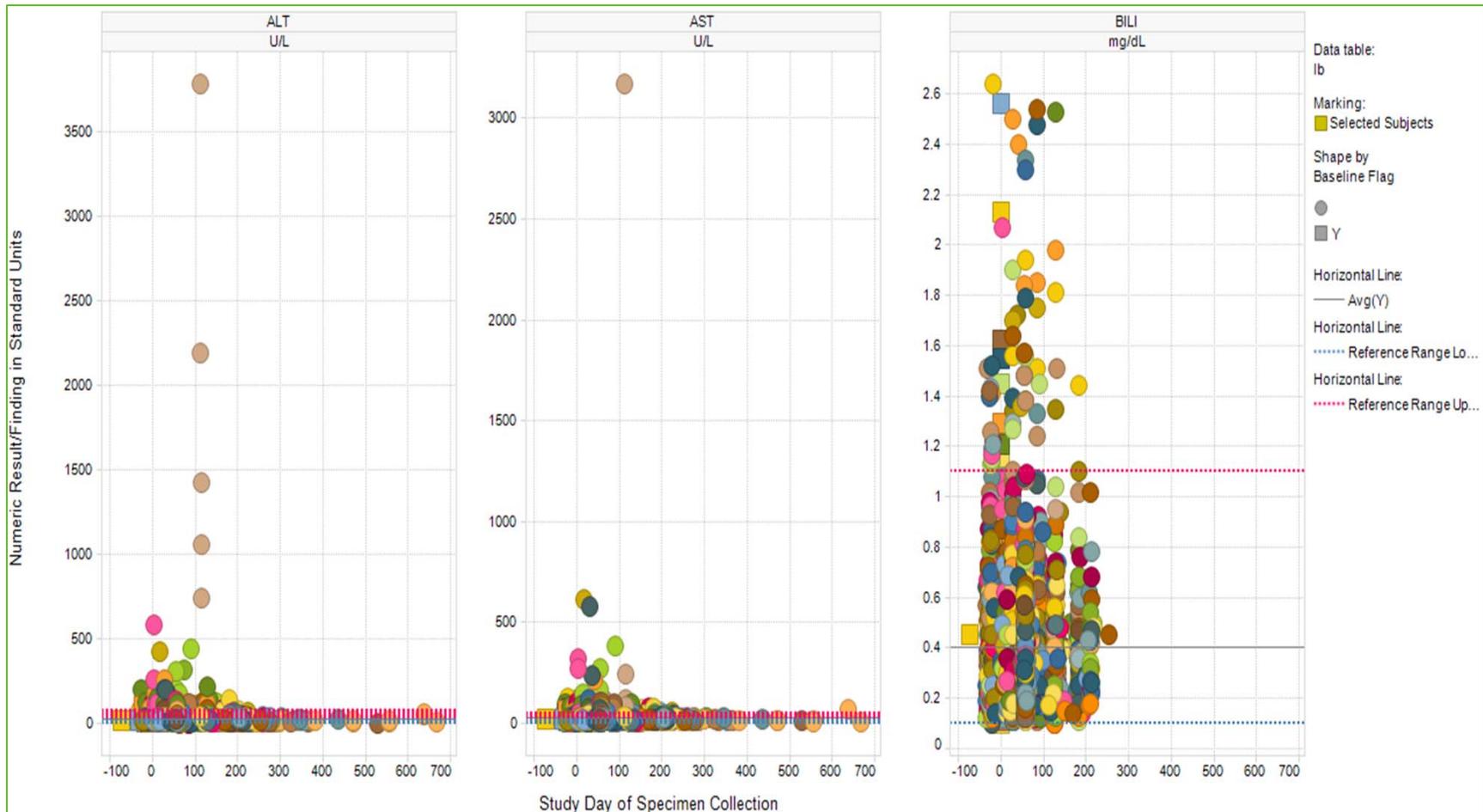
(Cont'd.)

- + 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



(Cont'd.)



(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

2 - Questionnaires x 2 - Disease Response x 2 - Swimmer Plot x 3 - Subject Report x 3 - AE/CMMH Data Tables x 3 - EX/PR/SU Data Tables x 3 - CE/IR/DD Data Tables x 3 - Subject Labs x 3 - Subject + < >

Table of Contents

Toggle Filter

Toggle Details-on-Demand

Unmark Marked Rows Invert Marked Rows

Reset Axis

Select Event Types to display:

- AE
- ConMed
- Disease Response
- Disposition
- DM Status
- Dose
- Lab
- MedHist
- Visit

Use slider to set Marker size:

Note
To drill down into the data of particular subjects in section 3, first add those patients to the Selected Subjects marking to populate the Subject List table; then mark the row from this table corresponding to the subject of interest.

Quick Mark

Active Screen Failure

Ended

Data table: dm

Marking: Selecte...

Online

24,637 of 194,893 rows 8 marked 31 columns SubjectReport

Subject List

Unique Subject Identifier	Age	Sex	Race	Country
DEMO-039-0018	33.00	F	WHITE	USA
DEMO-291-0006	35.00	F	WHITE	USA
DEMO-403-0002	26.00	F	WHITE	USA
DEMO-404-0004	76.00	F	WHITE	USA
DEMO-569-0007	38.00	F	WHITE	USA

Data limiting: Selected Subjects

Data table: dm

Marking: Subject Report

LABS: High High Panic Low Abnormal Normal

AE: Mild(1) Moderate(2) Severe(3) (4) (5) Serious

MH: Ongoing During and/or After Coincident Before

DS: Randomized Death/Withdrawn/Term

RS: PD SD PR VGPR nCR CR sCR NE UNK

NO DAY: No Date Date with No Day No Day COLOR: No Data

Subject Details

Unique Subject Identifier: DEMO-291-0006
 EventCategory: AE: Non-Serious (Color = Severity)
 Event: <Coded>: PRURITUS GENERALISED
 EventDetail: ITCHING ALL OVER, Day: 6, Outcome: RECOVERED/RESOLVED, Action Taken: DOSE NOT CHANGED, End: 2012-10-...

EventCategory

- AE: Non-Serious <Coded>: PRURITUS GENERALISED
- AE: Serious <Coded>: ALANINE AMINOTRANSFERASE INCREASED
- <Coded>: ASPARTATE AMINOTRANSFERASE INCREASED
- <Coded>: CITALOPRAM HYDROBROMIDE
- CM: CONCOMITANT MEDICATION <Coded>: DEXLANSOPRAZOLE
- <Coded>: DIPHENHYDRAMINE HYDROCHLORIDE
- <Coded>: IRON
- DOSE: TREATMENT
- DS: DISPOSITION ADVERSE EVENT
- DS: OTHER EVENT LAST DOSE 2012-09-04
- DS: PROTOCOL MILESTONE INFORMED CONSENT OBTAINED 2012-10-03
- RANDOMIZED
- ALP
- ALT
- AST
- BILDIR
- CMVIGGAB
- EBCIGGAB
- MPV
- RDW
- CREAT
- K
- VISIT: Planned
- VISIT: Unplanned

Study Day

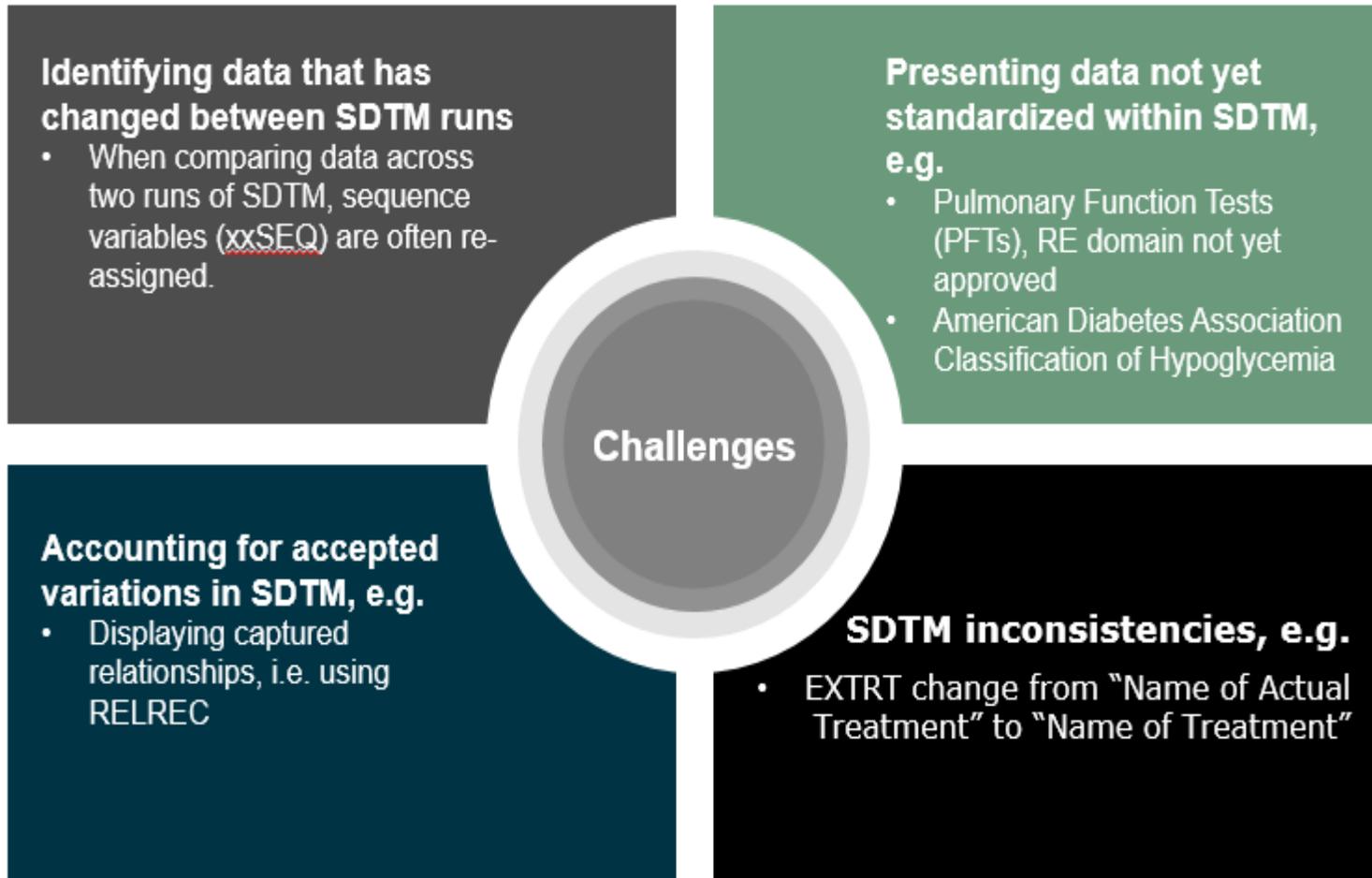
(Cont'd.)



(Cont'd.)

- + 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



(Cont'd.)

+ 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.

(Cont'd.)

- + 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 where possible. Also, we add new domains in our dashboard on regular basis.

(Cont'd.)

+ **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.