

Clinical Data Visualization using TIBCO Spotfire® and SAS®

Ajay Gupta, PPD

ABSTRACT

In Pharmaceuticals/CRO industries, you may receive requests from stakeholders for real-time access to clinical data to explore the data interactively and to gain a deeper understanding. TIBCO Spotfire 7.6 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 basic visualizations created using TIBCO Spotfire and SAS using raw and SDTM datasets. This paper will also discuss the possibility of creating quick visualizations to review third party vendor (TPV) data in formats like EXCEL® and Comma Separated File (CSV).

INTRODUCTION

TIBCO Spotfire 7.6.0 is, an analytics and business intelligence platform, which enables data visualization in an interactive mode, has been introduced into the pharmaceutical industry over the last few years. With its increasing implementation in the field of safety monitoring and dose escalation, TIBCO Spotfire presents its superiority for exploratory analysis. On the other hand, SAS software gives us efficient, strong statistical analysis and data processing capability. This article is to demonstrate data visualization within TIBCO Spotfire developed using SAS datasets.

This paper will demonstrate some basic visualizations created using TIBCO Spotfire and SAS using raw and SDTM datasets. This paper will also discuss the possibility of creating quick visualizations to review the third party vendor (TPV) data in formats like EXCEL and CSV.

TECHNIQUE AND MECHANISM

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

1. If the data is in EXCEL and CSV format then convert the data into SAS data set using IMPORT procedure (this step is not discussed in this paper) in SAS v9.3 or above.
2. 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.
3. Import SAS datasets 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 or above. The user can easily extend the macro to fit other SAS versions. 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 datasets 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 datasets. 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.

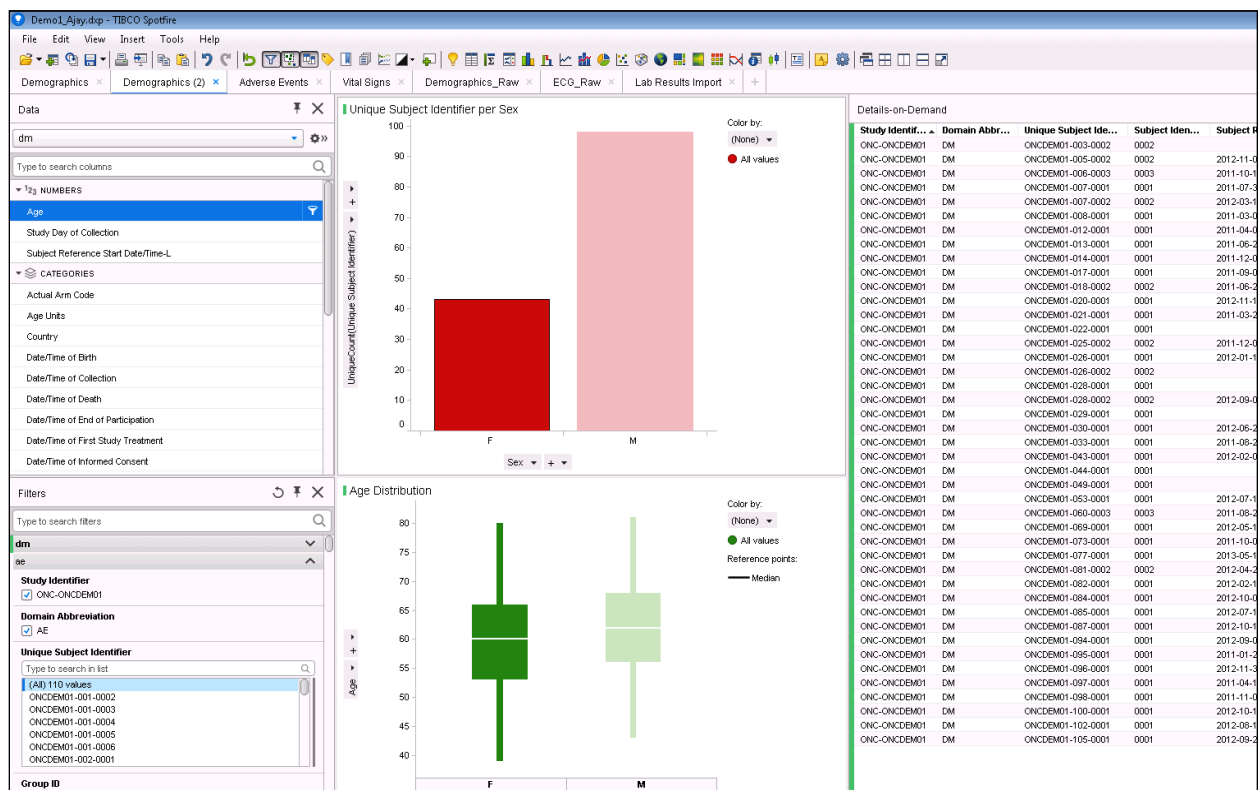
DeleteDS: Delete temporary dataset. Default value is 'Y'.

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.

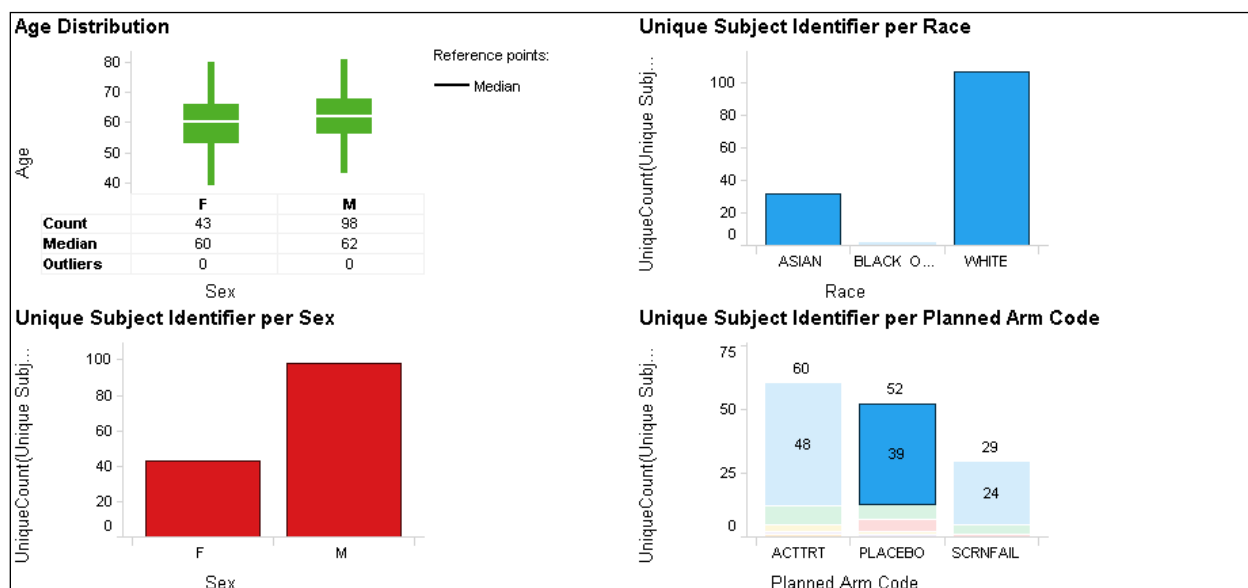
CLINICAL DATA VISUALIZATIONS

Now we will go through multiple visualizations developed using SDTM or Raw data sets. Later the paper will explore the possibility of developing a visualization using Third Party Vendor (TPV) data. Normally 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: DEMOGRAPHICS

Functionality:

Age distribution Box Plot, Subject Count by Race Bar Chart, Subject count by Gender Bar Chart, Distribution by Arm code and site.



Display 2. Visualizations for demographics data

Descriptions:

The above interactive visualization is created using the demographics (DM) data set from the SDTM database. It consists of a box plot and multiple bar charts. 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 male subject within the box plot then the bar containing the number of male subjects will be highlighted in the Unique Subject Identifier

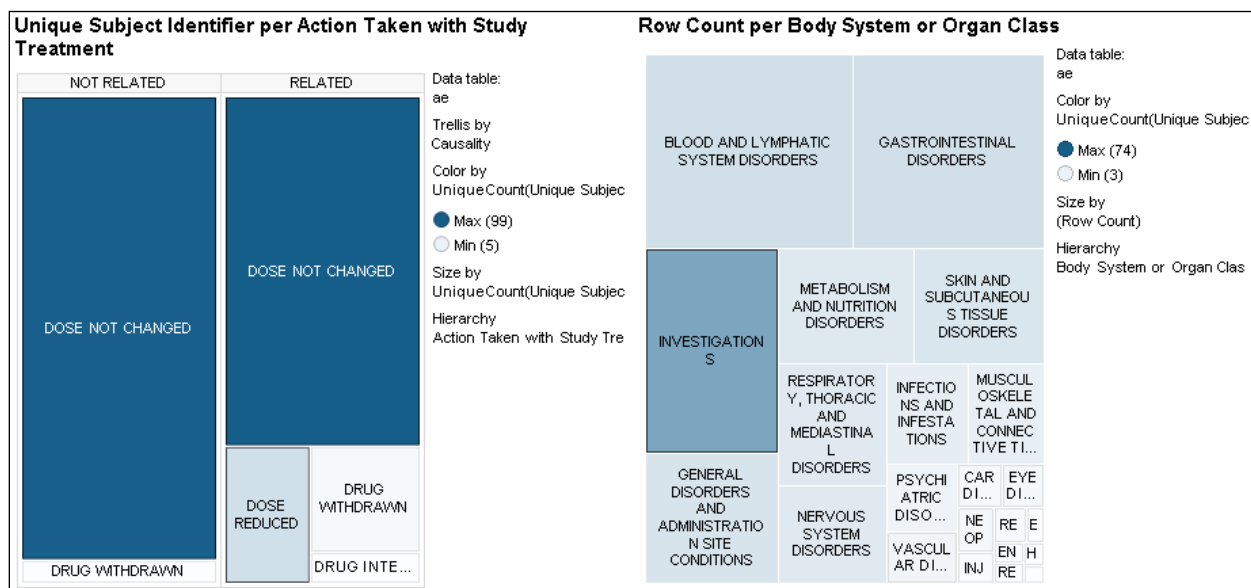
per Sex bar chart. Also, the filter option within Spotfire can be used to select a specific subgroup.

1. Age Distribution Box Plot: This interactive box plot will give user information about the age distribution across each gender and provide unique subjects count in a particular gender.
2. Unique Subject Identifier per Sex Bar Chart: This interactive bar chart will give the unique subjects count by gender.
3. Unique Subject Identifier per Race Bar Chart: This interactive bar chart will give the unique subjects count by race.
4. Unique Subject Identifier per Planned Arm Code per Site: This interactive bar chart will give the unique subjects count by planned arm code. Each bar is further grouped by site identifier.

EXAMPLE 2: ADVERSE EVENTS

Functionality:

Tree Map for Action Taken with Study Treatment, Tree Map for Body System or Organ Class.



Display 3. Visualizations for Adverse Events data

Descriptions:

Another graphic that is sometimes used to show adverse events is the Treemap. A treemap for adverse events from the SDTM database is shown in the screenshot above. This exploratory graphic allows a user to start at high-level overview of adverse events and then drill-down to a patient level view.

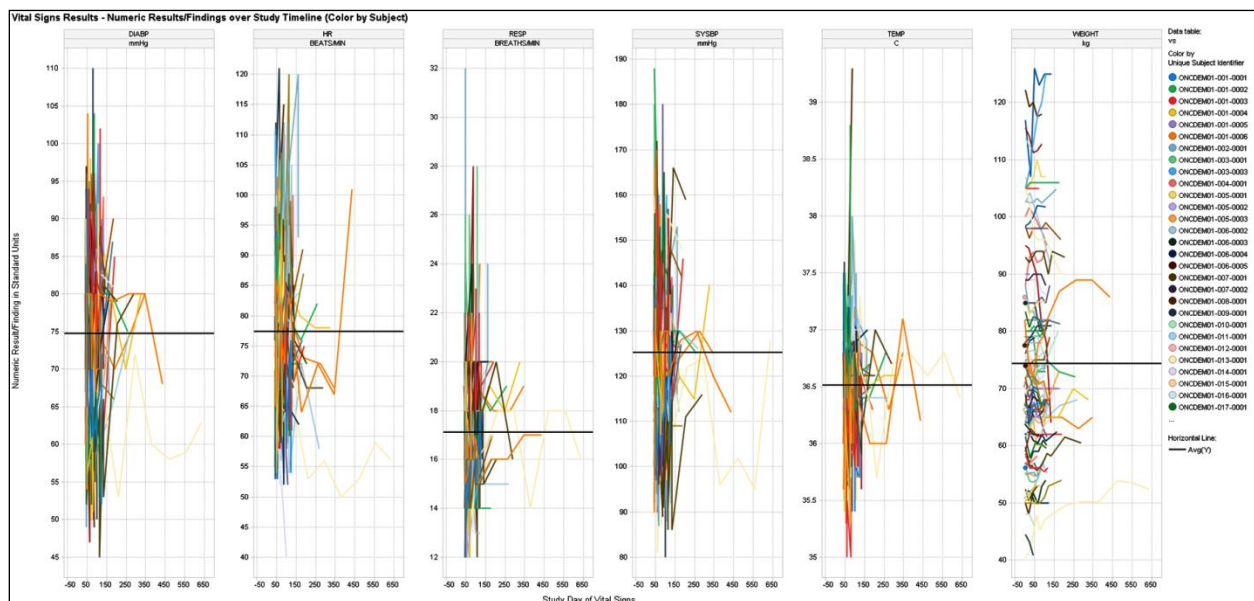
1. Tree Map for Action Taken with Study Treatment: The treemap graph shown above is an interactive graph which displays a hierarchy ordered by causality and action taken with study treatment. The size of the rectangles corresponds to the number of patient where this action taken appeared. The density of population is shown by the color of the areas. The tooltip displays some summary information about the respective area. Users can mark a particular area on the graph and access the data using the Details-on-Demand tab.
2. Tree Map for Body System or Organ Class: The treemap graph shown above 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. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab.

EXAMPLE 3: VITAL SIGNS

Functionality:

Line graph of Vital signs numeric results/findings over study timeline (color by subject) and panel by Vital Signs test and units.



Display 4. Visualizations for Vital Signs data

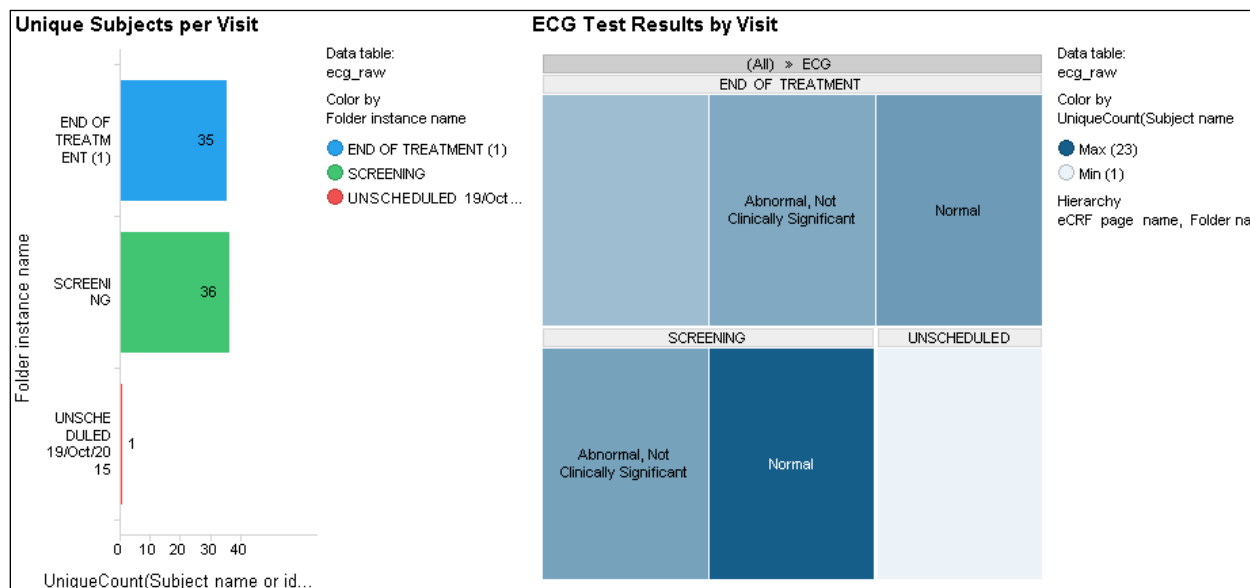
Descriptions:

The above interactive visualization is created using the Vital Signs (VS) data set from SDTM database. It consists of a line graph of Vital signs numeric results/findings over study timeline (color by subject) and a panel by Vital Signs test and units. 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 line graph 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 4: ELECTROCARDIOGRAM (ECG) USING RAW DATA

Functionality:

Horizontal Bar Chart for unique subject count per visit, Tree map for ECG results by visit.



Display 5. Visualizations for ECG Raw Data

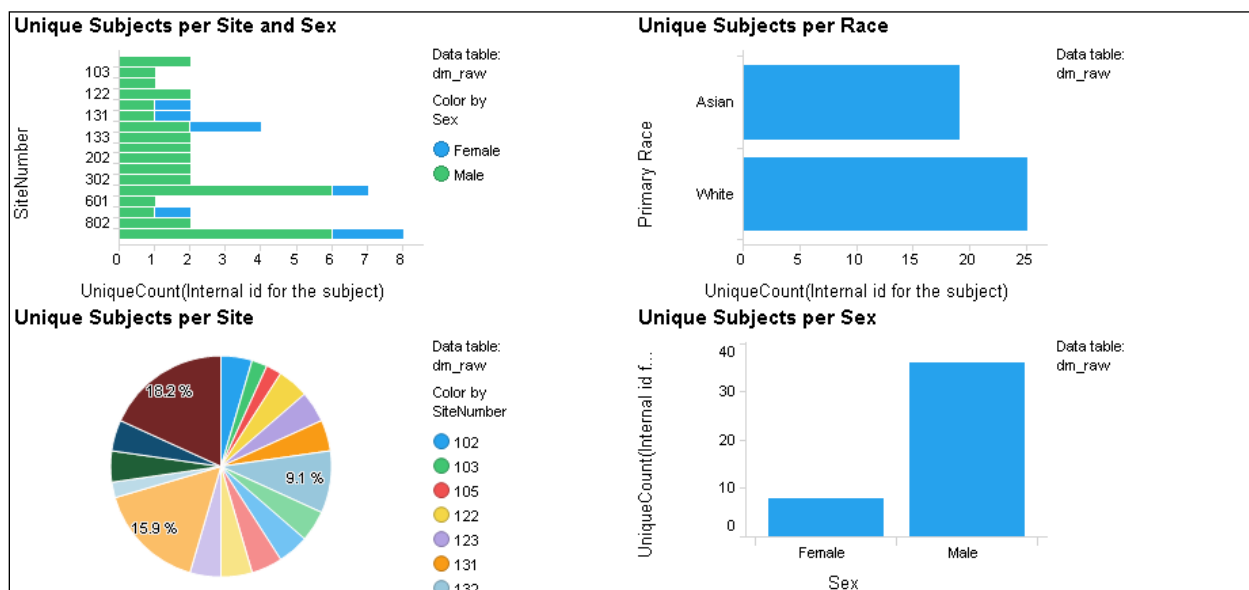
Descriptions: The above interactive visualization is created using Electrocardiogram (ECG) data set from the RAW database in Rave. It consists of a bar chart and a tree map. The important note here is that variables from the RAW database can be directly used in a visualization similar to the SDTM database. For example, in the RAW database the visit information is stored in the variable labeled 'Folder Instance Name' and ECG test results are mapped to the variable labeled 'Overall Interpretation of ECG' which are used in the Tree Map. Users can mark a particular area on the graph and access the data using the Details-on-Demand tab. All plots using similar data are interlinked together. So, if you mark/select a particular area on a particular plot then area containing the selected data on other plot will be highlighted. e.g. if you select a particular visit in horizontal bar chart then rectangle containing similar data will be highlighted in tree map. Also, filters in Spotfire can be used to select a specific subgroup.

1. Horizontal Bar Chart for unique subject count per visit: This bar chart will show the number of unique subjects by visit using variables from raw database. The tooltip displays some summary information about the respective area. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab.
2. Tree Map for ECG test results by visit: The treemap graph shown below is an interactive graph which displays a hierarchy ordered by visit and ECG test results. The size of the rectangles corresponds to the number of patients with particular ECG results for a given visit. The density of population is shown by the color of the areas. The tooltip displays some summary information about the respective area. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab.

EXAMPLE 5: DEMOGRAPHICS USING RAW DATA

Functionality:

Unique Subject Identifier by Site and Sex Bar Chart, Unique Subject Identifier by Site Pie Chart, Unique Subject Identifier by Race Bar Chart, Unique Subject Identifier by Sex Bar Chart.



Display 6. Visualizations for Raw Demographics data

Descriptions:

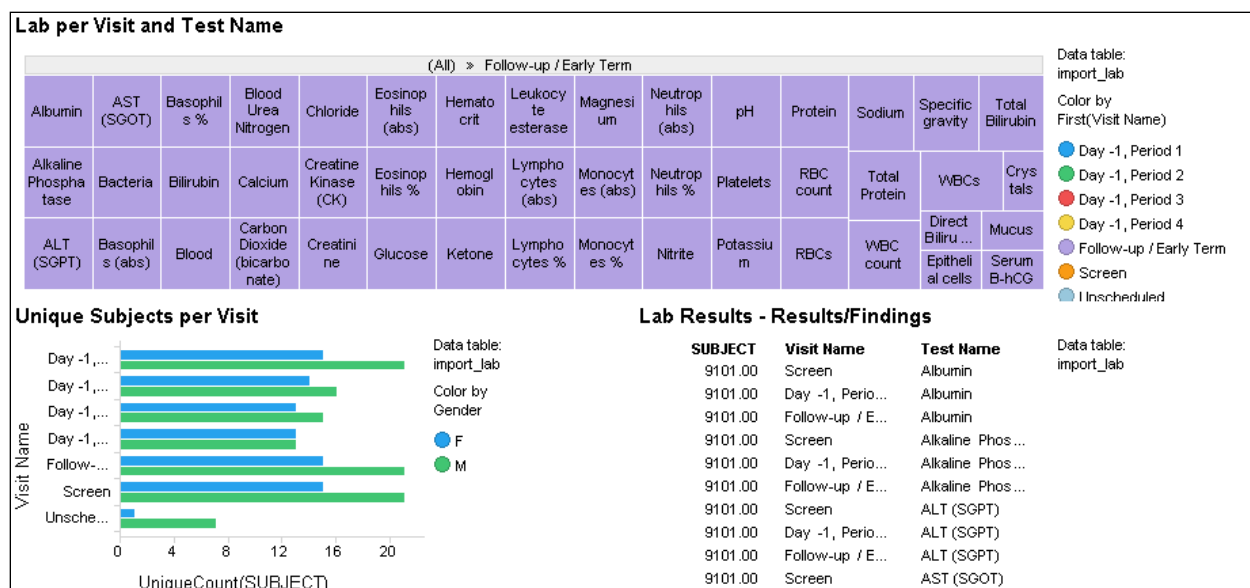
The above interactive visualization is created using the Demographics (DM) data set from RAW database in Rave. It consists of multiple bar charts and a pie chart. The important note here is that variables from the RAW database can be directly used in visualizations similar to the SDTM database. For example, variables such as site number, subject, sex, and race. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab. All plots using similar data are interlinked together. So, if you mark/select a particular area on a particular plot then the area containing the selected data on the other plot will be highlighted. e.g. if you select only male subjects in Unique Subject Identifier per Sex bar chart then the area containing the number of male subjects will be highlighted in other bar/pie charts. Also, filters in Spotfire can be used to select a specific subgroup.

1. Unique Subject Identifier per Site per Sex: This interactive horizontal bar chart will give the unique subjects counts by site identifier. Later the bar is grouped by sex.
2. Unique Subject Identifier per Site: This interactive pie chart will give the unique subjects counts by site identifier. The color in the pie chart is unique for each site.
3. Unique Subject Identifier per Race Bar Chart: This interactive bar chart will give the unique subjects counts by race.
4. Unique Subject Identifier per Sex Bar Chart: This interactive bar chart will give the unique subjects counts by gender.

EXAMPLE 6: LABORATORY USING TPV DATA

Functionality:

Lab Results/Findings table, Horizontal Bar Chart for Unique subject identifier count by Visit and Sex, Tree Map for lab data by Visit and Test name.



Display 7. Visualizations for TPV Lab data

Descriptions:

The above interactive visualization is created using the Laboratory data set from a third party vendor (TPV). It consists of a table, bar chart, and tree map. The important note here is that variables from TPV data can be directly used in a visualization similar to the SDTM database. For example, in the TPV the lab data variable 'TESTNAME' has lab test information, and lab results are stored in the variable "RESULT" which can be directly use in the treemap and graph. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab. All plots using similar data are interlinked together. So, if you mark/select a particular area on a particular plot the area containing the selected data on the other plot will be highlighted. e.g. if you select a particular visit in the horizontal bar chart the rectangle containing similar data will be highlighted in the tree map. Also, filters in Spotfire can be used to select a specific subgroup.

1. Horizontal Bar Chart for unique subject counts per visit: This bar chart will show the number of unique subjects by visit and sex. This graph gives a clear picture of all available visits. The tooltip displays some summary information about the respective area. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab.
2. Tree Map for Lab test by visit: The treemap graph shown above is an interactive graph which displays a hierarchy ordered by visit and lab test. This treemap will give a clear picture of all available lab tests in a given visit. A user can further drill down into the data on a subject level. The size of the rectangles corresponds to the number of patients with a particular lab test results for a given visit. The tooltip displays some summary information about the respective area. A user can mark a particular area on the graph and access the data using the Details-on-Demand tab.
3. Lab Results/Findings Table: This table will show some selected variables from the TPV data which can be useful for further evaluation.

CONCLUSION

TIBCO Spotfire provides an interactive platform for exploratory analysis. 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. 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.

REFERENCES

Gu Yi, 2016. Smart Statistical Graphics – A Comparison between SAS ® and TIBCO Spotfire ® In Data Visualization. Proceedings of the PharmaSUG China 2016 Conference, paper DV05.

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

<http://www.sascommunity.org>

ACKNOWLEDGMENTS

Thanks to Lindsay Dean, Tammy Jackson, Ken Borowiak, Ryan Wilkins, David Gray, Richard DAmato, Lynn Clipstone, Ed Lunk, 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_;
    IF 0 THEN SET contents NOBS=nobs;
    CALL SYMPUTX ('obs', nobs);
    STOP;
RUN;

%IF &obs=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;
```