PharmaSUG China 2018 – Paper AD-39

Practices of Interactive Visual Analysis for Clinical trials

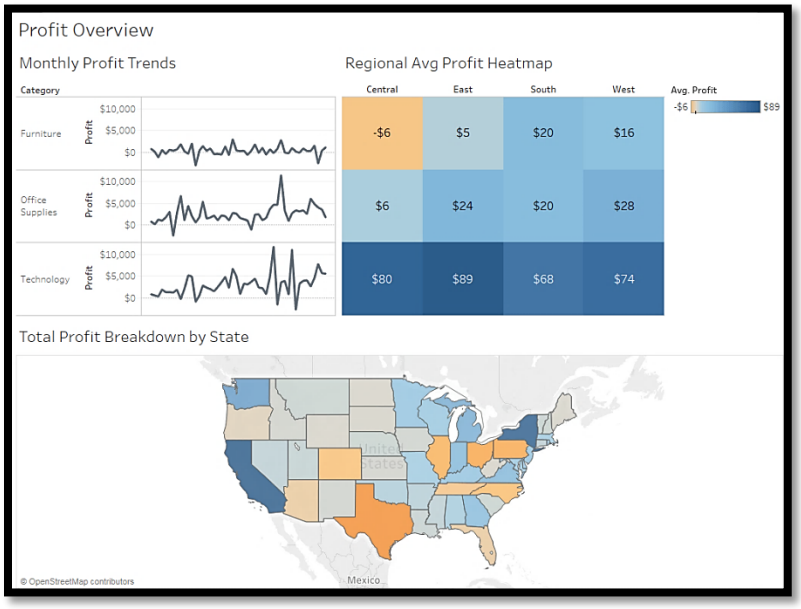
Gaoyang Li, Bayer Healthcare Co. Ltd., Beijing

ABSTRACT

This paper is to promote interactive visual analysis (IVA) for clinical trials with ideas from different roles in the process of a clinical trial. Definition and scope of IVA are explicated and key features are listed. Regarding to IVA, ideas were collected from CRA, data manager, biostatistician, statistical analyst and medical expert. Topline results interactive analysis platform, as an example, was established and illustrated to show the great benefits of this advance analytical method.

I. INTERACTIVE VISUAL ANALYSIS INTRODUCTION

Interactive Visual Analysis is more and more popular and has been applied to business analysis for many years and with business intelligence methodology. Interactive Visual Analysis is a set of techniques for combining the computational power of computers with the perceptive and cognitive capabilities of humans, in order to extract knowledge from large and complex datasets. The techniques rely heavily on user interaction and the human visual system, and exist in the intersection between visual analytics and big data. It is a branch of data visualization [1]. Following dashboard is the basic output of IVA [2].



Display 1. Process of developing topline results interactive analysis

Applying IVA method and tools, all the clinical trial data consumers could benefit from the straight forward, integrated, self- explored and real-time analysis platform. IVA is not only a dashboard, but also with other features. Following table shows some important features/advantages of IVA platform for clinical trials.

Features	Description
Visualizing	Plots, maps, tables, dashboard, story
Analytical	Individuals, summary, frequency, distribution, calculation fields, trend, cluster, etc.
Integrated	Cross domains, global parameters
Interactive	Filter, highlight, tooltip, drill-down, data view
Dynamical	Pages, animations, trace, alias
Web-based	URL, permissions, real-time
Extendable	SAS, R, Python
CDISC compatible	SDTM for data review, ADS for reports
Non-programming	Drag and Drop

Table 2. Key features of IVA platform for clinical trials

II. REQUESTS COLLECTED FROM DIFFERENT ROLES

Study Monitor

CRA is the site manager and responsible for data monitoring and site level process. 1) Besides EDC, a tool/platform to provide real-time site information summary would be much helpful for them to conduct Source data verification (SDV), such as CFDI listings or FDA Bioresearch Monitoring (BIMO). 2) Due to their academy background, CRA needs more statistical/calculation support to check complicated /linked data, such as interpret protocol wording into calculation of duration(+90 days or 3 months on calendar).

Data Manager

Data manager is ensuring data quality and much interested in data quality related analysis, especially real-time analysis of queries. They are hoping to 1) summary variables by the data distribution and find out data errors/issues. 2) Dynamically creating query reports from/for different roles would help them to identify the owner of data issues/errors and close related queries quickly than before. Except dealing with queries, data manager is responsible for data related process/meetings and creating study data related reports for study team to review or take action.3) they hope to use tools to create study events report and study status report to reduce manual work.

Statistical Analyst

Statistical analyst is usually SAS programmer and responsible for creating analysis datasets and producing results out such as Tables, Listings and Figures. They are very interested in data linage, data visualization and automatic validation. Even they have great programming skills, 1) finding a reliable non-programming way to validate datasets and output is preferred. And avoid of creating huge amount number of repeated tables with only one or a few parameters changed by a SAS macro or other tools would make their life easier, such as 2) automatically create subgroups analysis. Sometimes medical experts/statisticians/clinical leads need some ad hoc tables/listings/figures which are much similar to the version provided. If the 3) data consumer could investigate the data/output with a few setting, this would save much time from back and forth communication and tedious repeated job.

Biostatistician

Statisticians are hoping to 1) display clinical trial results in real-time manner, especially for efficacy. They also want to 2) dynamically compare primary endpoints between treatment groups, subgroups, and overall vs. specific subgroups. Statisticians are much interested in data reports and would like to 3) have a tool to create statistical analysis report automatically.

Study Medical Expert

Study medical experts are focusing more on safety than efficacy during the life cycle of a study. They are more interested in safety signals, especially adverse event related. They regularly review individual data from EDC or patient profile in a static way and hope 1) to review data with more exploratory options such as filtering, linking, summary, annotation, etc. 2) Interactively investigate subject safety baseline with tools such as cluster, risk trend and disease characters, would help them find out safety signals much easier.3) With baseline safety signals and prediction tool such as time sequence information of adverse events, study expert would be more proactive to take action to ensure patient safety.

III. TOPLINE RESULTS INTERACTIVE ANALYSIS EXAMPLE

Based on summary of requests from different roles, topline results interactive analysis platform (TRIA), as a pilot project, was developed for statistician to disclose study topline results to study team, including Clinical Lead, Data managers and Medical expert.

Process

The roadmap is creating dashboards based on standard data structure data, and sharing the output to study team in the web-based platform. Following chart is illustrating the process.

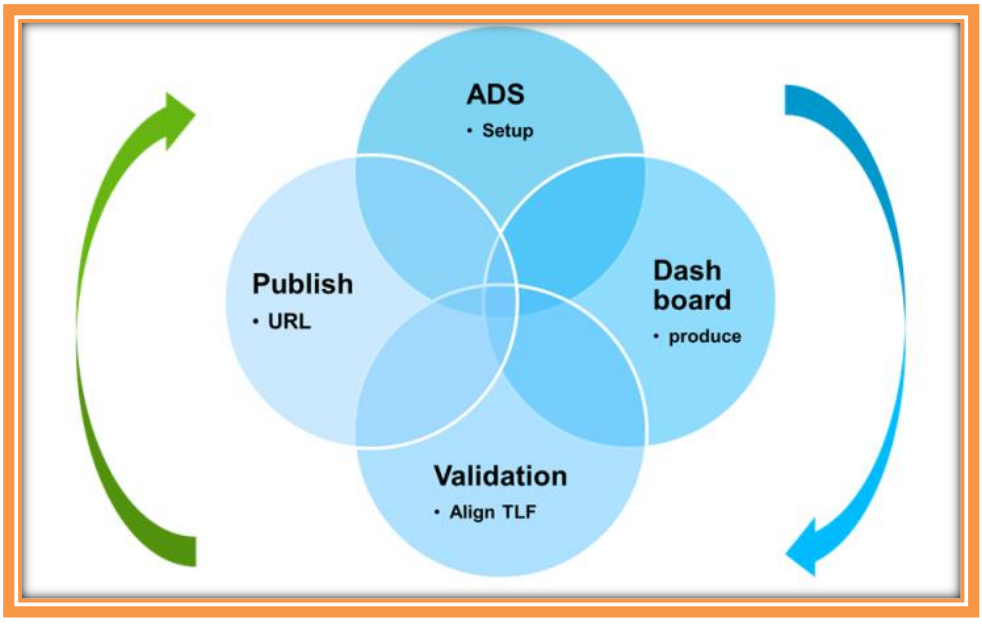
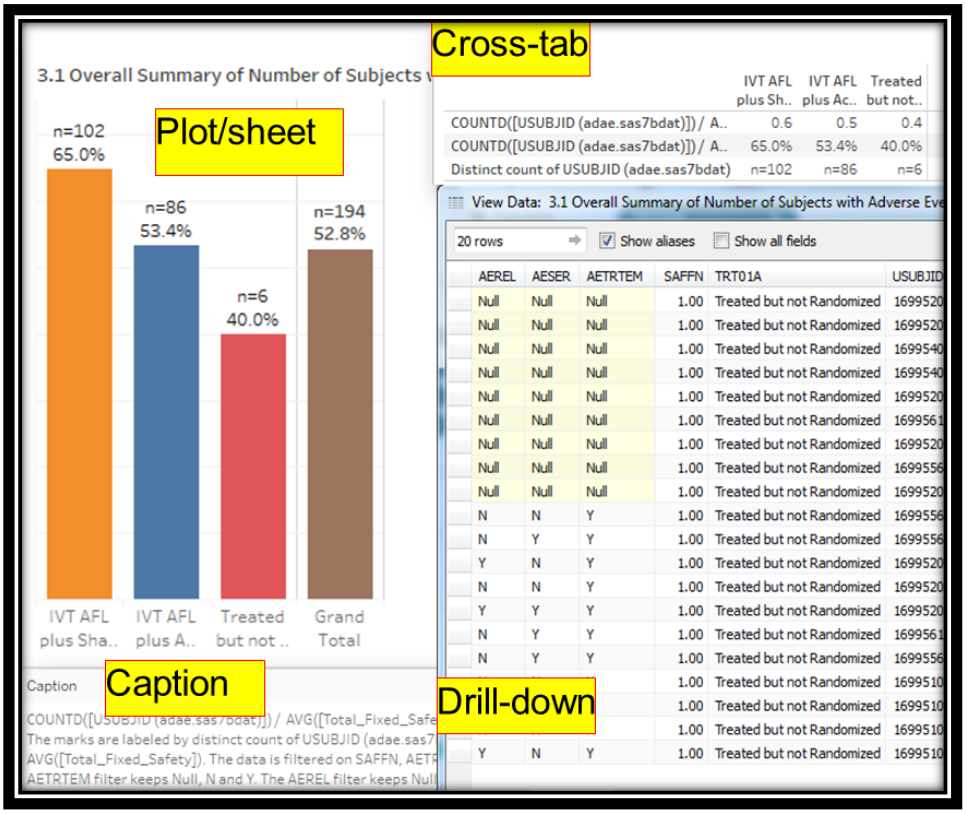


Figure 3. Process of developing topline results interactive analysis

Since standard data already created based on company SOP, they could be imported directly and no more additional data preparation needed. SDTM would be a good start point for data review, and in this case, ADaM is the start point for topline analysis according to our purpose.

According to topline template (or CSR synopsis), the 3-level architecture structure was established: Topline results of a study are Story level, and composed by 3 sections which are of Dashboard level. At the bottom key tables/figures are represented in Sheets, which are the major elements for dashboards. Advanced brushing and linkage would be applied to facilitate self-exploration of the data.

Currently traditional Tables/Listings/Figures package is the main input source for CSR and there is no guidance for using IVA for submission, cross-tab check instead of double programing or eyeball-check is in the platform. Additionally, key numbers from the platform would be compared to numbers from Tables/Listings/Figures package. In the most IVA sofeware, there are many ways to conduct validation, such as checking captions, cross-tabs and drill down the data.



Display 2. Validation options for interactive analysis

The unique URL of a specific study on the topline results platform is shared to study results review team instead of sending emails with attached files or other shared drive location. The reviewer could log in automatically by clicking the shared URL when they work on company laptop.

Topline results interactive analysis User Interface

Topline results was composed by 3 sections, study subjects and baseline, efficacy, and safety, topline results interactive analysis was composed by 3 dashboards correspondingly.

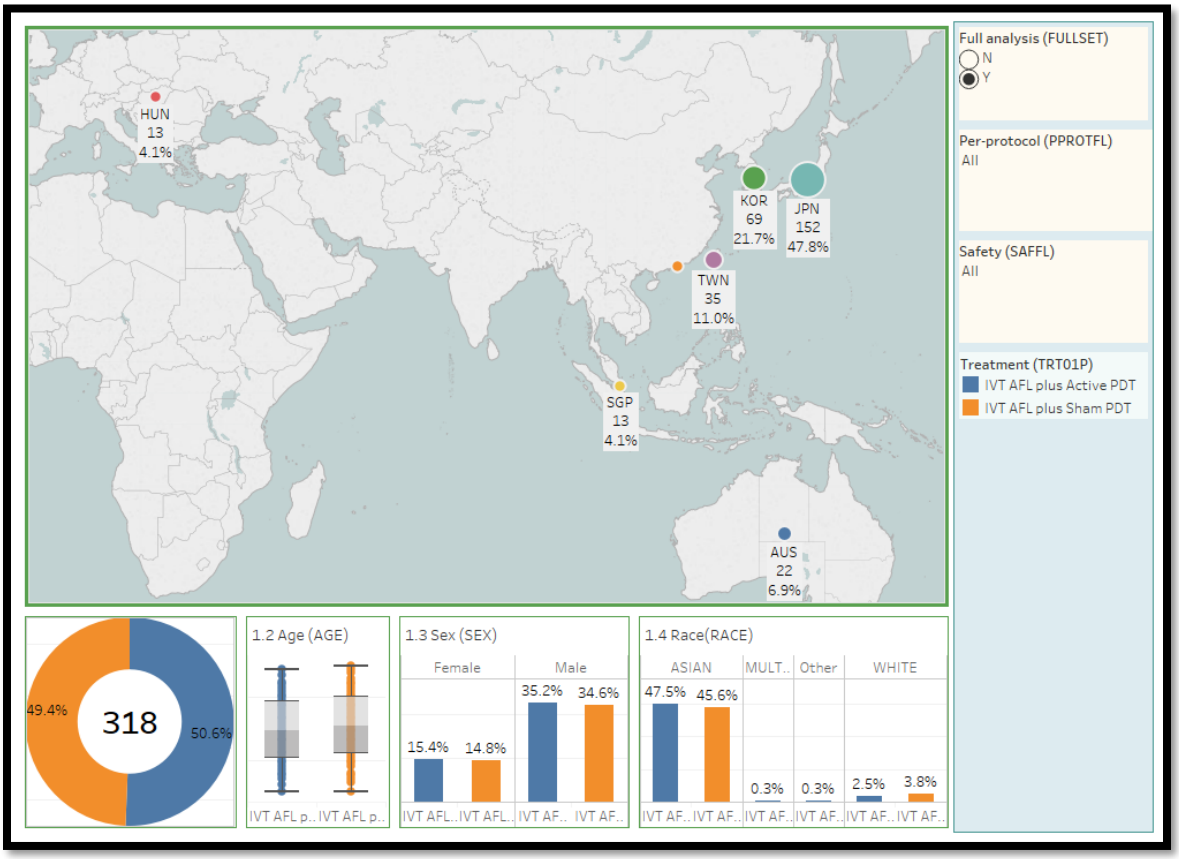
Study Subjects dashboard

Study Subjects information dashboard contains common fields such as Country/Region, Site, Number of Subjects, Age, Sex, and Race, etc. Baseline and other more information could be added based on specific studies.

Study Subjects UI is composed by 2 areas, the right control panel and the other plot area. The control panel includes 3 population filters and the treatment group legend, which is highlight selector as well.

On the left top of the dashboard is the number and percent of each Country/Region on the map. The size of the circle on each country indicates the amount of subjects, and the circle could act as country subgroup selector. When one country was selected, the numbers of the plots on the bottom would change to the country level immediately since the map and the plots were linked together.

Age was presented with Box and whisker plot to show mean, min, max, 25% quartile and 75%, and some individual values as well. Different from measurement field Age, other fields are categorical fields.

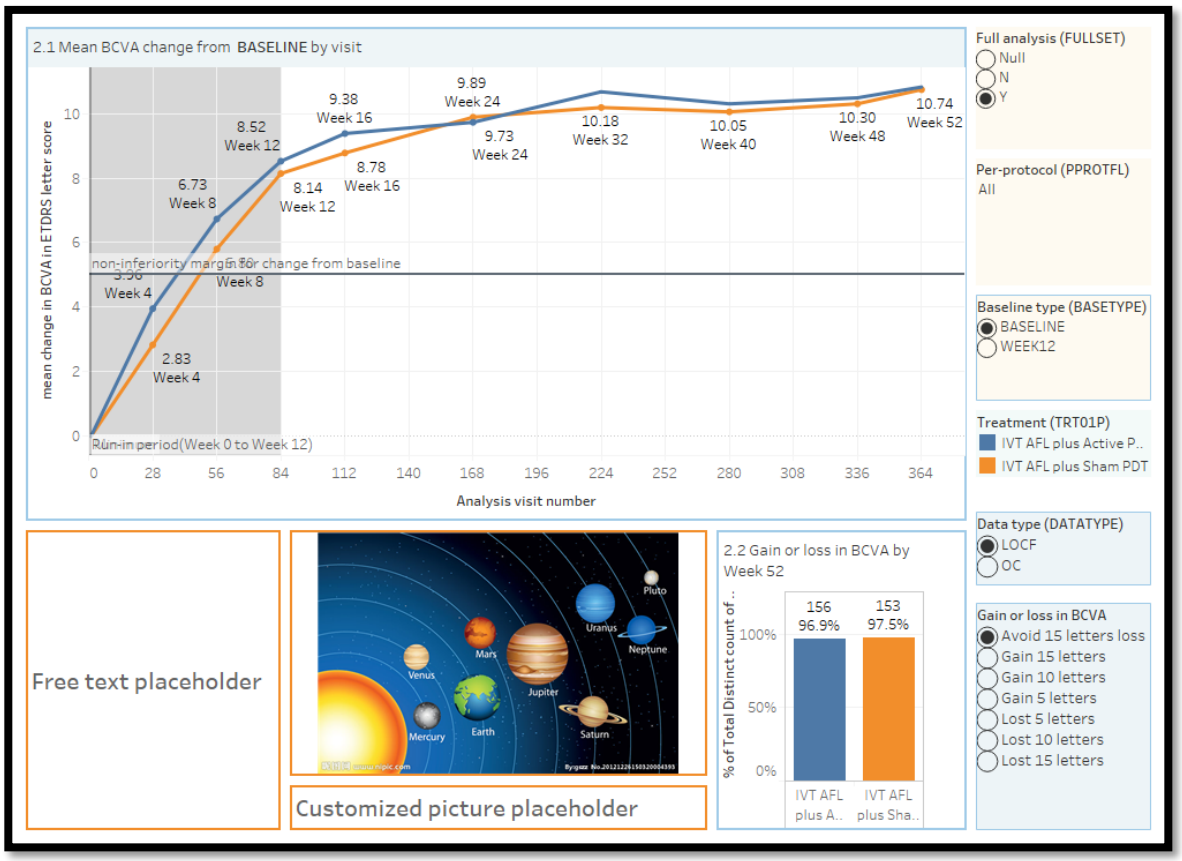


Display 3. Study subjects dashboards of topline interactive analysis

Efficacy dashboard

Efficacy information dashboard contains Primary endpoints and Secondary endpoints. In this case, The primary endpoint is BCVA change from baseline/week 12 by week 52, and second endpoint is number of subjects avoids 15 letters loss by Week 52.

There are 2 different setting of Efficacy dashboard from Study subjects dashboards.1) in the control panel, control parameters were added to give more options for data consumers, the parameter Data Type for the sheet 2.1 mean BCVA from BASELINE/WEEk12 by visit, and Parameter Gain or Loss for BCVA for the sheet Gain or loss in BCVA by week 52. 2) There is placeholder area for text and pictures added in the dashboard to display the conclusions or explanations or other specific results, or whatever related to efficacy.



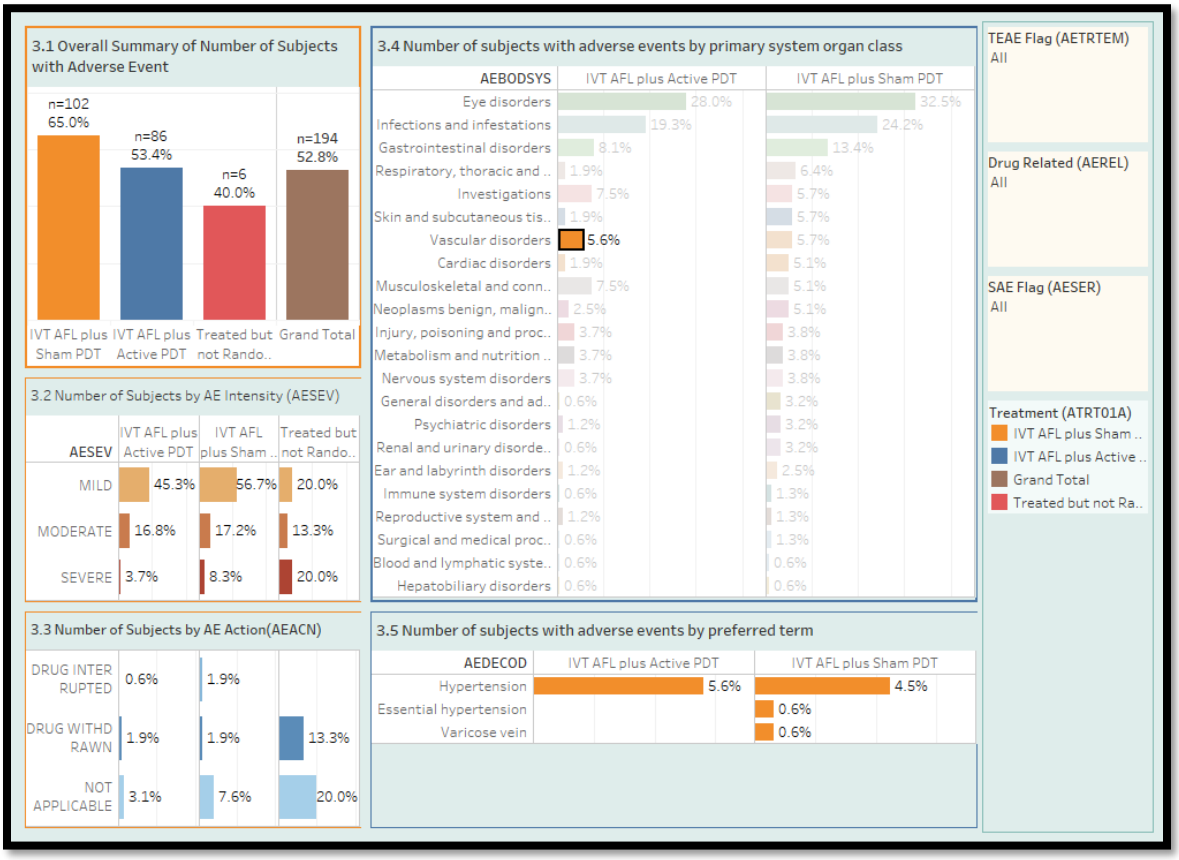
Display 3. Efficacy dashboards of topline interactive analysis

Adverse Events dashboard

Safety section usually contains summary information of adverse events and other test information such as vital signs and laboratory test. In this case, only adverse event information is of topline results.

The same as subject dashboard, this dashboard contains control panel on the left and data present area. In this dashboard, the data preset area was divided into 2 vertical sub -area: The left (orange border color) overall summary of numbers of subjects with AE and in the middle (blue border color), Number of subjects with adverse events by primary system organ class (and by preferred term). Each sub -area has it main sheet and secondary sheets. If the

main sheet was filtered or highlighted, then the secondary sheets would present only the selected data. It means sheet 3.2 and sheet 3.3 linked to sheet 3.1, and sheet 3.5 linked to sheet 3.4. These settings provide the flexibility to check overall summary and check other detailed attributes as the same time. Medical expert could check data by primary system organ class and check the numbers under this specific system organ class by preferred term.



Display 4. Adverse Events dashboards of topline interactive analysis

SUMMARY

Interactive visual analysis is easy to develop and explore the data based on related software or package. It could shorten the process from data collection to data disclosure. Interactive visual anlysis method/platform would strengthen study team since all roles of a clinical trial could access data easily rather than heavily rely on statistician or analyst. Statistician and analyst could eventually evolve as data scientist, for the interactive visual analysis could integrate standard deliverables in a real-time manner.

ACKNOWLEDGMENTS

The author would like to thank Data science and Analytics department of Bayer Healthcare Co. Ltd. for supporting innovation method and tools test for clinical trials. Specially, the author would like to thank my line manager Michael Badlani and Eric Zhang, for both budget and working time, and their insights and promotion. The author would like to thank all members of Data science and Analytics department for their precious input. Special thanks to Alice Liao for review this paper and her constructive comments.

REFERENCE

[1] Wikimedia Foundation, Interactive visual analysis, https://en.wikipedia.org/wiki/Interactive_visual_analysis

[2]Tableau, Built-in features made for advanced dashboarding, https://www.tableau.com/solutions/business-dashboards

CONTACT INFORMATION

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

Name: Gaoyang Li

Enterprise: Bayer Healthcare Co. Ltd.

Work Phone: +86 10 65360717

E-mail: gynju@163.com

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.