PharmaSUG China 2019 – Paper DV-054

Using R to Query Dataset from SAS Server to Generate Clinical Trial Graphs

Jei Li, Janssen Pharmaceuticals, New Jersey, USA
Xiang Li, Janssen Pharmaceuticals, New Jersey, USA

ABSTRACT

The strength of SAS® is based on its powerful data manipulation capability, but it can be a challenge when using SAS PROC TEMPLATE to generate customized plots, especially for beginners. Alternatively, R has exceptional flexibility in creating customized plots compared to SAS. However, the data management function of R tends to be time consuming and not as straightforward as SAS. To leverage the strength of each language, we will display how to integrate SAS dataset in R environment to generate clinical trial graphs.

INTRODUCTION

As we know, R is mainly designed for statistical analysis and data visualization, it is very powerful in creating plots with extreme flexibility. However, R still has its limitation, for example: data manipulation. It could be very difficult to do data manipulation by using R, especially for beginners. Using SAS to do data mining is easier than using R. In this paper, we will execute a prepared SAS program in R environment to get an input dataset and generate different kinds of clinical trial graphs using R graphic code based on the dataset.

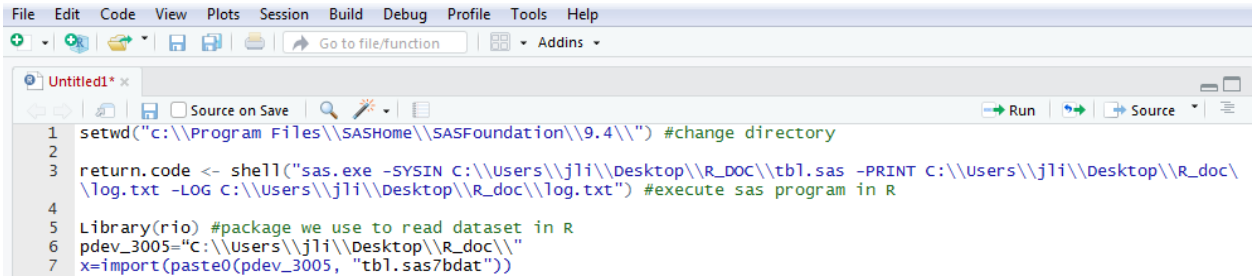
IMPLEMENTATION

EXECUTE SAS PROGRAM IN R

Before actual plotting, prepare a SAS program to get an input dataset. Run this program in R environment to avoid submitting SAS program in other servers.

1. Change directory to the directory containing the SAS executable file
2. Run prepared SAS program by using shell() function in R to get our input1987wl dataset
3. Read the dataset in R

By submitting the following code, we will have our input dataset ready and can start to generate different graphs based on the dataset.



GENERATE CUSTOMIZED PLOTS IN R

When talking about the advantage of R plotting, the first thing that comes to mind is flexibility. By using R, we can move each element in the plot to anywhere we like it to be.

To get a customized plot, we need to set up a frame (plotting area) first.
Calling par() function to control the size of margin area and using png() function to save the plot as a png file.
Code example:

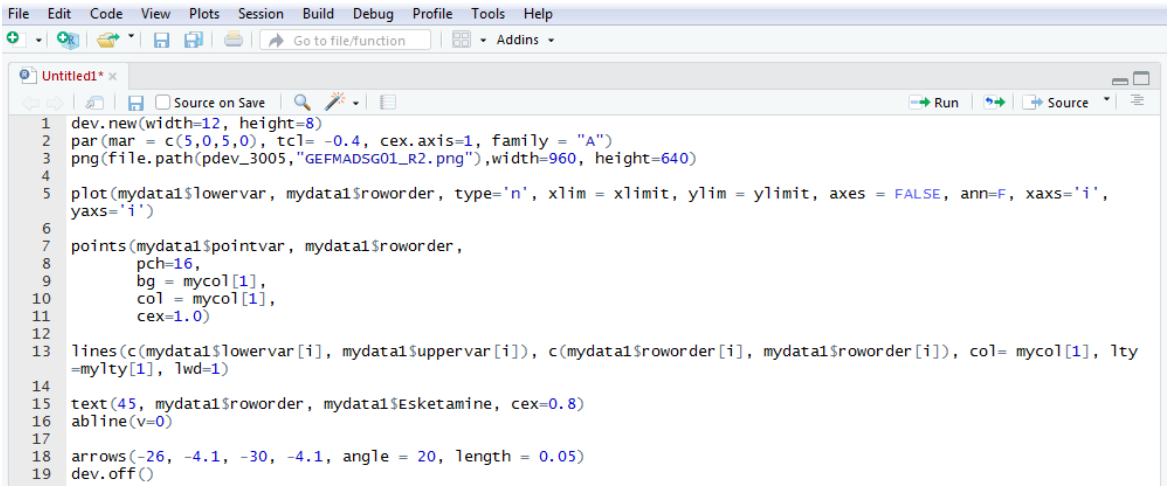
```
par(mar=c(bottom, left, top, right))
png(file.path(), width=, height=)
```

All R plotting are generated based on XY coordinate, we can put label, text and symbol anywhere in the graph based on the values of x and y. Here we can always find a position of a point and tell R where we want to put an

element on this point because each point has its own coordinate inside the plotting area. SAS, however, doesn't have this kind of flexibility. There are only three position options in SAS GRAPH: LEFT, CENTER and RIGHT. In this case, it will be a challenge if we would like to put a label or a legend on a specific place in the graph.

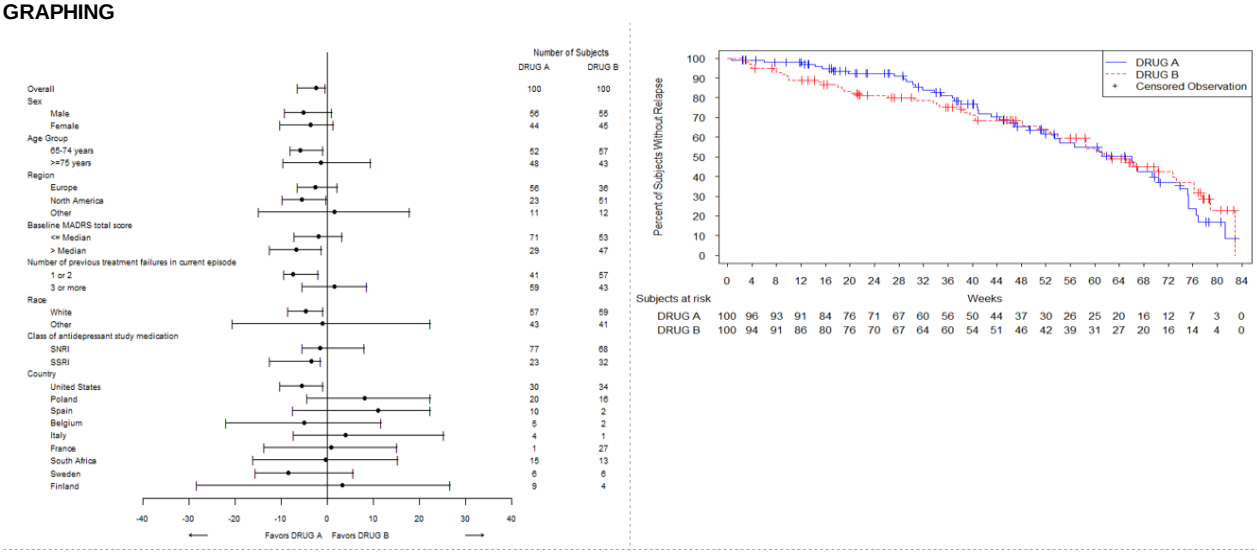
Code example:

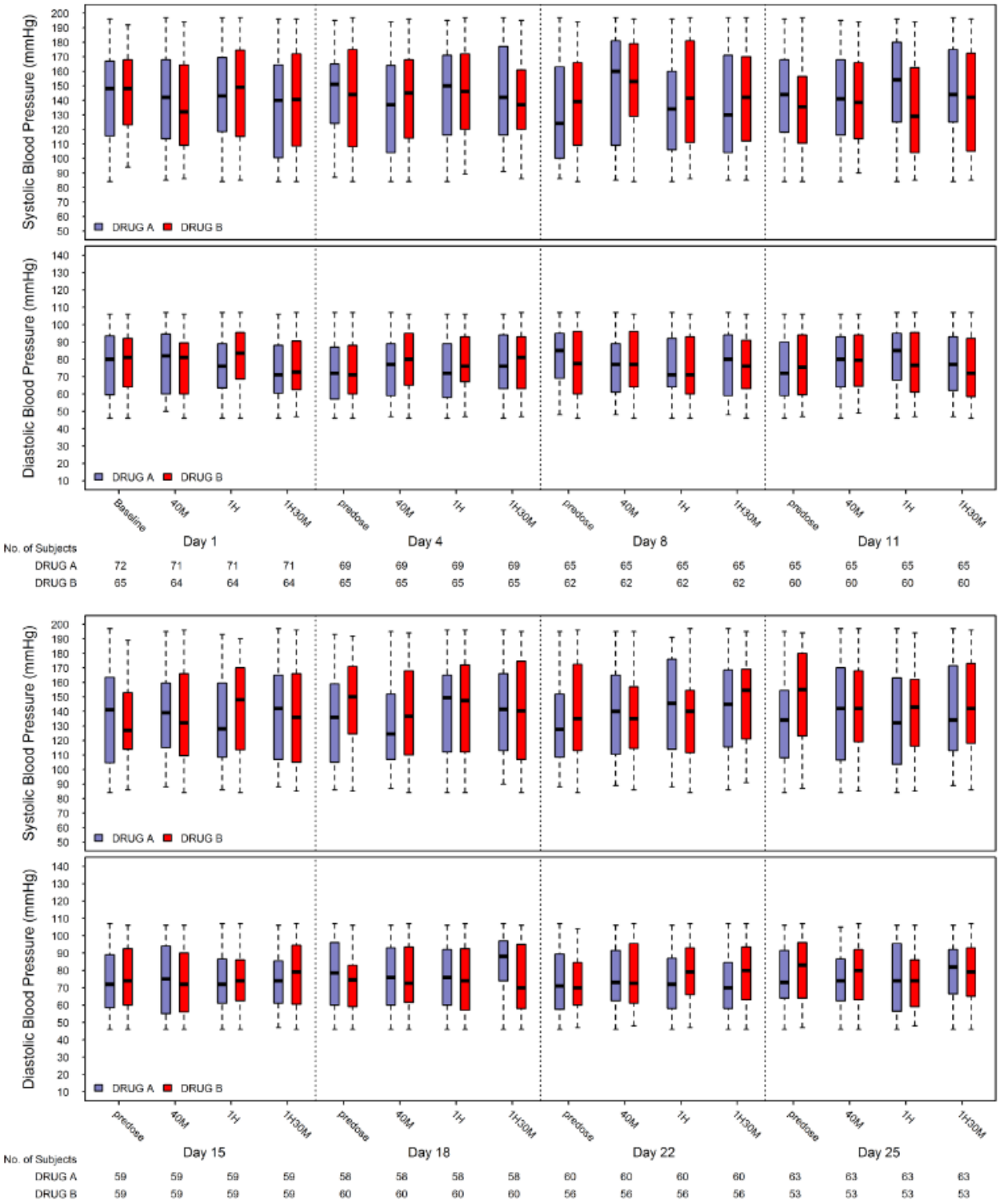
```
points(x coordinate, y coordinate, pch=, …)      ,
lines(x coordinate, y coordinate, lty=, …)       ,
text(x coordinate, y coordinate, labels=" ", …) ,
legend(x coordinate, y coordinate, …)            , etc.
```



Different visually appealing clinical trial graphs can be generated by using the functions mentioned earlier. We only need around 50 lines of code to create a forest plots as follows. By using SAS PROC TEMPLATE, more than 100 lines of code is required to get the same forest plot outputs. R graphing code is exceedingly simple, straightforward and time saving compare to SAS PROC TEMPLATE.

CUSTOMIZED PLOTS BY USING R GRAPHING





CONCLUSION

There are many advantages of using R to create figures. Not only can R create visually appealing plots, the main advantage is it is extremely flexible and could adjust almost all the graph parameter in the plots. R graph related functions are straightforward, easy to pick up and beginner-friendly. Compared to SAS PROC TEMPLATE, R code is straightforward and simple when creating customized figures. However, using R to do data manipulation is cumbersome. As a result, we still need to rely on other coding languages such as SAS and PROC SQL to create input datasets. Neither SAS or R is perfect, but in combination, generating graphs can be as efficient as possible.

CONTACT INFORMATION

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

Jei Li
Janssen Pharmaceutical Company of Johnson and Johnson
1125 Trenton Harbourton Rd,
Titusville,
NJ 08560
jli319@its.jnj.com

Xiang Li
Janssen Pharmaceutical Company of Johnson and Johnson
1000 US-202,
Raritan,
NJ 08869
xli256@its.jnj.com