



Paper DV01

Data Visualization: Where to start?

Emma Moore, Roche, Welwyn Garden City, UK

ABSTRACT

There has been a significant rise in interest around Interactive Data Displays (IDDs) from those reviewing and interpreting clinical trial analyses. As clinical data scientists we need to develop new skills to support our stakeholders, especially in this area, as new and effective ways of reviewing and interpreting data are introduced. Two tools, TIBCO Spotfire® and R Shiny®, have been popularized in this field and are now widely used at Roche. This paper will look at some tips and tricks on how to get started with these tools, sharing my experiences of how I learnt to use them. I will also look at different use cases that will highlight situations where it is advantageous to use one of these tools over the other, especially when you are developing knowledge in the use of these tools.

INTRODUCTION

The increase in demand for Interactive Data Displays has many, myself included, finding themselves looking to upskill in languages and software that they have previously not encountered or used as part of their day to day roles. Approaching a new language or software can be slightly daunting at first, especially when there are so many resources available and it can be difficult to know just where to start. Unfortunately, there is no 'magic' answer or a 'one size fits all' approach to the question "Where do I start?", but by the end of this paper, the reader should feel less intimidated by the number of resources available and have a better strategy to approaching these new tools.

SELECTING THE RIGHT TOOL

Ultimately, the first thing that needs to be done is to ensure that you are selecting the tool that will fit your needs. Spotfire and R Shiny are both incredibly powerful tools, and from experience I have found that one tends to be more fit-for-purpose than the other, depending on the task at hand. Below are a couple of examples where I have found the one tool has an advantage over the other.

SPOTFIRE

Where Spotfire shines is the ability to delve into the details of the data straight from a plot, and the ability to incorporate this functionality into interactive displays with relatively basic knowledge of the tool itself. This is especially useful for safety data displays. For example, if the IDD contains a bar chart of Adverse Events (AEs) coloured by severity type, should a stakeholder spot something within the data that they wanted to investigate on a patient or event level, they will be able to highlight that particular section of the chart and have the individual patient or AE details appear, which will definitely aid them in their review of the data.

Another benefit to Spotfire is that, thanks to the point and click nature of the software, it takes a relatively short amount of time to get an IDD ready for use. In more recent versions of the software, suggestions of which visualization to create are shown after the upload of a dataset. If you are looking to learn a new tool and get an IDD out to stakeholders in a relatively short amount of time, this is a big plus. This isn't to say that mastering this software is simple, but you can still produce powerful analysis without having to delve too much into the custom expressions & other features available to you.

R SHINY

As R Shiny is a package within R itself, the ability to use R code to do analyses within the R Shiny app itself is very powerful. There are statistical packages that can be called and displayed within the IDD that may not necessarily be available for use in Spotfire, and by calling these programmatically, the developer will have more control over the parameters used and ultimately what is shown to the stakeholder. As data scientists, it is part of our role to empower the stakeholders to make informed decisions based on the data, and so this ability to have more control over what can be displayed is a great advantage to R and R Shiny.

With R Shiny, and R, it is also possible to create more complex visualizations, again with more control over what is being shown, as whilst a point and click software is useful, it can often be difficult to go back and reproduce the same



results without the code. It can even be difficult to utilize the same type of plot with a different dataset without having to go through and recreate all the steps, whereas with programming you can copy and paste the code and adapt it in less time than it would have taken to create again from the beginning. If you're looking to create a large app with several domains and some complex analysis, which will most likely apply to efficacy trial data, then R and R Shiny would be my recommended route.

GENERAL TIPS

As someone who is essentially self-taught in using both R and Spotfire, the overwhelming feeling when starting with both these tools is one that is all too familiar. The following tips are ones which have been gathered over the months of learning and can be applied to learning anything, not just R and Spotfire.

SET ASIDE TIME

The most common challenge when starting to upskill is setting aside the time to work on personal development. It's easy to put this to the bottom of a to do list when requests come in and the day to day work piles up, but the importance of dedicating time to upskilling cannot be stressed enough. Try and block out some time in a calendar, an afternoon a week or a day every fortnight, and be strict about keeping that time for development.

HAVE A CLEAR GOAL

Whilst the notion of learning a new skill for the sake of learning a new skill is a great one, it's a good idea to have a clear goal in mind. This would usually take the form of an IDD for a stakeholder. Having this end product in mind will help keep the motivation to learn this skill at the forefront of your mind. It will also mean that once these skills have been learned, it's less likely that you will have to go back and relearn some topics, as you're putting your skills into practice on an almost daily basis and increasing the retention of them.

R SHINY

As someone who is essentially a self-taught R programmer, the overwhelming feeling when starting to learn R is one that I am all too familiar with. Having such a plethora of resources at one's disposal is a problem that many would envy, but when starting out it can be difficult to know where to begin.

START SIMPLE

Before developing an R Shiny app, it would be good to learn some of the basics in R if you are not already familiar with the language. There are plenty of online courses available from websites such as DataCamp® and Udemy®, which provide a very good starting point and a comprehensive guide to learning base R. From there you can build on your knowledge by looking into other packages available. For data visualization, the two packages I would recommend looking into would be plotly and ggplot2, not only are there plenty of resources available online to give guidance, but between the two you can create almost any plot.

WORK STEP BY STEP

Starting small and building upon it is a good way to ensure that you don't take on too much of the language too fast. Whilst it's good to have a complex app as the ultimate goal, taking on the challenges at a beginner stage can demotivate some people when learning as they don't feel they understand the reasons why some example code might add in additional lines and what their purpose is.

If you're learning R as an additional language, it may help to reproduce a plot that you have created in another language to gain a better understanding of where the differences and similarities between the two languages lie. Selecting a plot produced by someone else or in a software that you're unfamiliar with would also help as then there is a finished product to compare the R visual to and improve upon. Once you have this static plot, you can then look to incorporate this into an app and look to add in some additional features.

LEARN THE BASICS OF R SHINY

The past two sections have focused on R in general, so now would be a good time to learn the basics of R Shiny and the first step would be looking into how the different components of the R Shiny package work together. Apps are built on two main components, the user interface (ui) and the server. The ui is where the frontend of the app will be



defined, what the stakeholder will see and how they see it, and the server is the backend of the app, where the data is processed and the visuals are created. There are two main ways to create an R Shiny app using these components, they can be called as functions within one file named app.R, or they can be created as two separate files, ui.R and server.R, and the run from within the R session. As a beginner, I would recommend creating the two separate files until you have got to grips with the language and the technique of creating the app, as these files can be a lot easier to manage and find pieces of code within than having the app stored in one file. There is of course no wrong way to create the app, as long as the developer is comfortable with the setup and the user is happy with the end result.

FIND EXAMPLES

As mentioned in the introduction, there is no 'one size fits all' approach, what works for one programmer may not necessarily work for another. However, given the amount of examples of R Shiny apps there are available online, I would be remiss not to encourage you to take advantage and learn from other creations. The R Shiny gallery is filled with examples and features that you may want to incorporate into your app, and whilst looking at these, it's often common to spot a solution or a different technique for a different part of the app that you may not have considered before.

Examples of apps may not be limited to just those on the Shiny server or online, there may be apps that have been produced within your company that you can look to for inspiration and guidance. If these are working on the same data types and have been built to incorporate company standards, then they will be an invaluable resource.

ONLINE RESOURCES

Online resources can take the form of official guidance from the R Shiny site, or searching through forums such as Stack Overflow®. It's very likely that if you have a question or an idea of what you would like to implement, somebody else has asked the same question and had it answered.

PEERS & COLLEAGUES

Although touched on in the 'Find Examples' subsection, it's not only the work of your colleagues that can be helpful when starting to create an app, reaching out and asking those who have a greater knowledge of the language that you will often get you the answers that you need quicker and with more context than looking through example code and forums. If there is an internal course being produced, even if it covers the topics you have already learned, sign up and attend, there will be chances to ask questions and learn the 'softer' skills within programming such as good coding practices and more efficient ways to process data.

SPOTFIRE

As mentioned previously in this paper, one of the great benefits of this software is the point and click nature and the best tip that can be given when learning Spotfire is to load in a dataset and explore all the different options available for each of the visuals.

BE CURIOUS

One of the best tips that can be given when learning Spotfire on initiative alone is to be curious. Load in a dataset and explore all of the options available for each of the visuals. By altering the different properties in the software and examining the results, you will possibly have a much deeper understanding of how each one affects the data and the IDD than sitting through and reading some training slides or instructions.

The filtering schemes and marking within Spotfire are the functionalities which when combined, produce the result mentioned in the 'Selecting the right tool' section, so it would be recommended to focus on these parts of the plot properties if you were hoping to add that functionality to the IDD. Other properties such as the 'custom expression' can be used to not only limit the data shown, but add in other analyses within tables and plots that may not be available by using only the point and click method.

WORK STEP BY STEP

As with creating an R Shiny app, break down the task at hand into steps. For Spotfire it may be more manageable to create the IDD one tab or domain at a time, so that you don't overwhelm yourself by taking on too much all at once. If the IDD you are creating is for multiple domains, focus on just one and think about which visuals would best suit the



data type and the stakeholder needs. Create these on separate tabs to keep the IDD manageable at the start, as these can always be combined into one tab later, and then add in other features you want to include. Once you have the basics of the IDD in place, then you can start to build in more complex functionalities and look at creating custom expressions to limit and control more of the data, as well as adding in document properties that function similarly to macro variables in SAS and can be used to control visuals across the entire IDD, not just for one tab.

ONLINE RESOURCES

Again, there are plenty of resources online that can be utilized when learning something new. There are community boards set up by Tibco, the company who have developed Spotfire, and whilst the tool may not be as common on other online forums or courses, there is plenty to be found to help with learning. Tibco also have plenty of videos on tips and tricks that can improve Spotfire IDDs.

ATTEND A TRAINING COURSE

As mentioned before, with this tool it can be quick to gain enough skills to get an IDD ready for use, but to fully understand the power of Spotfire and the full capabilities, if the online resources haven't covered the topics you would like to learn about in enough detail, it would be recommended to try and attend a training course that covers the more intricate functionalities.

CONCLUSION

There are many different tools available for use to create Interactive Data Displays, and whilst this paper has briefly covered only two tools, the tips available for those starting to incorporate data visualization into their work should make the task of learning a new skill more manageable and less daunting. The links to the websites referenced within this paper can be found in the 'Recommended Reading' below.

RECOMMENDED READING

R Shiny gallery - <https://shiny.rstudio.com/gallery/>
The basic parts of a Shiny App - <https://Shiny.rstudio.com/articles/basics.html>
TIBCO Spofire – Learn - <https://www.tibco.com/products/tibco-spotfire/learn>
R Graphics Cookbook, Winston Chang
TIBCO Communities - <https://community.tibco.com/products/spotfire>
Udemy - <https://www.udemy.com/>
DataCamp - <https://www.datacamp.com/>
Stack Overflow - <https://stackoverflow.com/>

ACKNOWLEDGMENTS

Dave Summers, Tim Barnett, Craig Gower, Kieran Martin, Carolyn Fisher, Vinod Kerai

CONTACT INFORMATION

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

Emma Moore
Roche Products Ltd.
Hexagon Place, 6 Falcon Way,
Welwyn Garden City, AL7 1TW
+44 (0)1707 366014
+44 (0)7823 421393
emma.moore.em1@roche.com

Brand and product names are trademarks of their respective companies.