

Medical Data Review and Exploratory Data Analysis using Data Visualization

Joel Allen, Roche Innovation Center Basel, UKIntroduction

Background and Context

Since the advent of the computer, there has been an explosion of innovation in both industry and our personal lives. The ability to share our lives more easily with loved ones, to manage our daily lives (both work and at home), to connect with people in an instant, or to find out news from the other side of the world that we would normally not hear about. All of these innovations have a common thread, they deal with data. And as more and more of these innovations become common, the amount of data they deal with becomes larger and larger. Does this sound familiar?

As we collect more and more data, in our work and in our private lives, it becomes an important question of what are we going to do with all of this data? Do we want to look for data that lies at the extremes of a scale? Do we want to filter out all of the data we consider to be extraneous? Do we want to find subgroups of the data and cluster them together? Are we looking to characterize the entire distribution of data or are we looking for a correlation between multiple data points? These are all questions that can be answered through data mining and data visualizations.

Data Visualization Tools

Data mining is just searching through large amounts of data to find points of interest that might hold answers to questions we have about a population. Sometimes the answers that it holds are unexpected or different than we thought. Data visualizations are just a way of expressing that data through a graphical means. Data visualizations have the advantage of being able to aggregate large amounts of data into a graphic with only a few axes. This allows us to focus only on the data we would like to see. Typical graphs might only have 2 or 3 axes but as our tools become more developed it is becoming increasingly common to add in the axis of time so that we can see how are data develops as time goes on. Also our tools are allowing interactions between the tool and the users to dive deeper into the data to look at only the data that interests us, or change scales on the axis to allow us to look at the data in different ways. Doing this in real time interactions with the user, simply and efficiently, gives the user better control of what they are looking at and has the advantage of decreasing repetitive work to look more closely at a subset of data. Hence, using smart visualizations can allow us to reuse graphs for different analysis of the data without having to continuously reprogram our visualizations . Programs like Spotfire, Tableau Public, Google Charts, Gephi, Java script and R are easy to use and set up so that visual representations of the data are user friendly, have axis's, scale, labels can be changed at runtime by the user and reduce the need to do repetitive work.

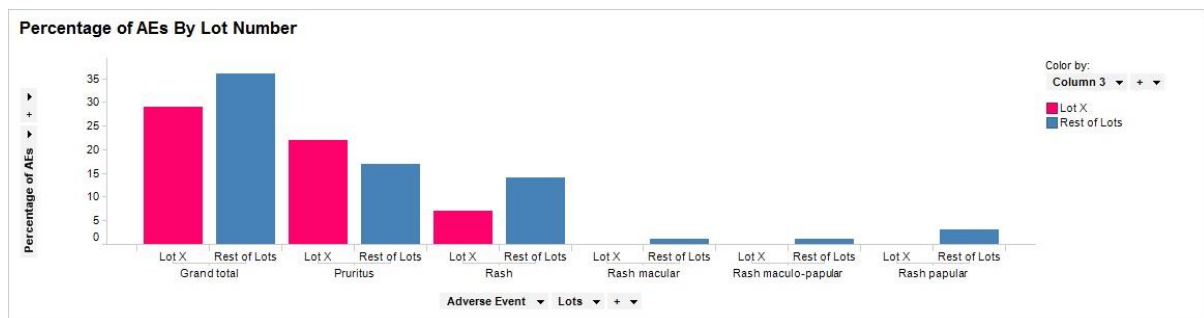
Data Visualization in Clinical Research

Data visualizations in the medical field have been used historically in and will continue to be used in the future. It is a clear and concise method to convey information in regards to the medical data collected. Refer to any clinical trial that has been completed and you will find a “tables and listing” section. Using data visualizations is nothing new, but with new innovations in computing technology and the ways we collect and store data, it is now easier and more accessible to present data than ever before. With programs mentioned above, we can access data almost up to the second that it has been collected - giving us the most up to date view of the data we are looking at. Using technology services like cloud applications allows you to pass this information to many users instantaneously even if they are not connected directly to your network. Data connectivity allows you to access your database from anywhere in the world and get instant access to your data.

With all these tools at our disposal, why are we still only looking at the tables and listings once we have collected all of the data? In the past, we had to program every table and listing individually because they were meant to be looked at through a paper copy whether for submission to a journal or to a regulatory body. But more and more of our work and interactions are now done through a computer. It makes sense to update the way we work and think to have an increased interactive approach, making our data interactive as well. No more should we be thinking about a paper solution and by removing the use of paper solutions we can start to look at the data as it is collected in an interactive fashion. By doing so, especially in the medical research field, we can start to answer questions about the data as we collect it. In clinical trials the questions have always been in relation to patient safety data. The usual process was via data dumps into table and giving it to a scientist to sort through. Now we can work interactively with the data to get answers to these questions fast and accurately and with more ease and user-friendliness to the scientists.

Compare the data below.

	Patients exposed to Lot X (n=27) Number of patients (%)	Control patients not exposed to Lot X (n=71) Number of patients (%)
Pruritus	6 (22%)	12 (17%)
Rash	2 (7%)	10 (14%)
Rash macular		1 (1%)
Rash maculo- papular		1 (1%)
Rash papular		2 (3%)
Grand total	8 (29%)	26 (36%)



Both of the above answer the question “Is there a difference in AE’s between Lot x of a drug and the rest of the population.” By reading the table we can see there is not a big difference however with the visualization we can see immediately that there is very little difference. The difference is with the visualization we can interactively change which lot we are interested in to compare, for example Lot y with the rest of the population, without having to redo all of our calculations. (Unfortunately, this can only be done in the programs listed above and not in a text editor. We are only ½ way there.) But you can see how answering questions through interactive visualizations can work with higher volumes of data and increasing number of questions quickly and efficiently without having to recalculate all of our variables.

Data Visualization Experts

These tools are great and you can see the immediate advantages of these over paper solutions. Although it is not enough just to hand these tools off to scientists on a clinical study and expect them to see immediate advantages. As with most things in life if you expect to use a tool well, you need experience using that tool to get the results you wanted. You cannot hand a paintbrush to anyone and expect them to paint a Picasso on their first try. As experience is gained, they will become better and better at using that tool to get the expected results that they want. But can we wait for results or tolerate errors when we are conducting human trials with drugs? A better solution would be to have experts in the programs work directly with scientists to answer questions about the data for them. This brings the Picassos to the canvas quicker rather than waiting for a masterpiece.

By embedding programmers into the scientific community, we can ensure that scientists’ questions are answered quickly, effectively and interactively so that we can see immediate results. Having programmers who are solely responsible to work directly with scientists, we can reduce the amount of time taken to answer questions that may arise during a clinical trial. We can ensure integrity of the study design, save time and money in developing drugs for market and most importantly keep patients safe at the same time. We can develop answers to real world questions, in real time, and distribute those answers to the people who need them most wherever they may be in the world. With the smart use of our technological tools, and the pairing of computer specialists with scientific specialists, we can make the most of tools that we have available to us.

Contact Information:

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

Joel Allen
Hoffman La Roche
Grenzacherstrasse 124
CH-4070, Basel, Switzerland
Tel: +41 (0)61 688 2796
email: joel.allen@roche.com