

Research Profile

Jimmy Oh

December 11, 2014

My interest is in Statistical Computing, which to me means I have a background in Statistics and I write software that makes life easier.

An obvious application would be *efficiency tools*. These tools take tasks that can be done manually, such as basic calculations, and do it for you faster (and often more accurately) by using a computer. A less obvious application would be *enabling tools*. These tools enable their users to do things they would not have been able to do before. Though I create efficiency tools from time to time, my passion is in creating enabling tools so that I can share the power of the computer to more users, opening their eyes to new possibilities. The lines can blur between these two tools as there exist tasks that would take too long to do manually so as to be practically impossible. A tool that does this task quickly then is both an efficiency tool and an enabling tool.

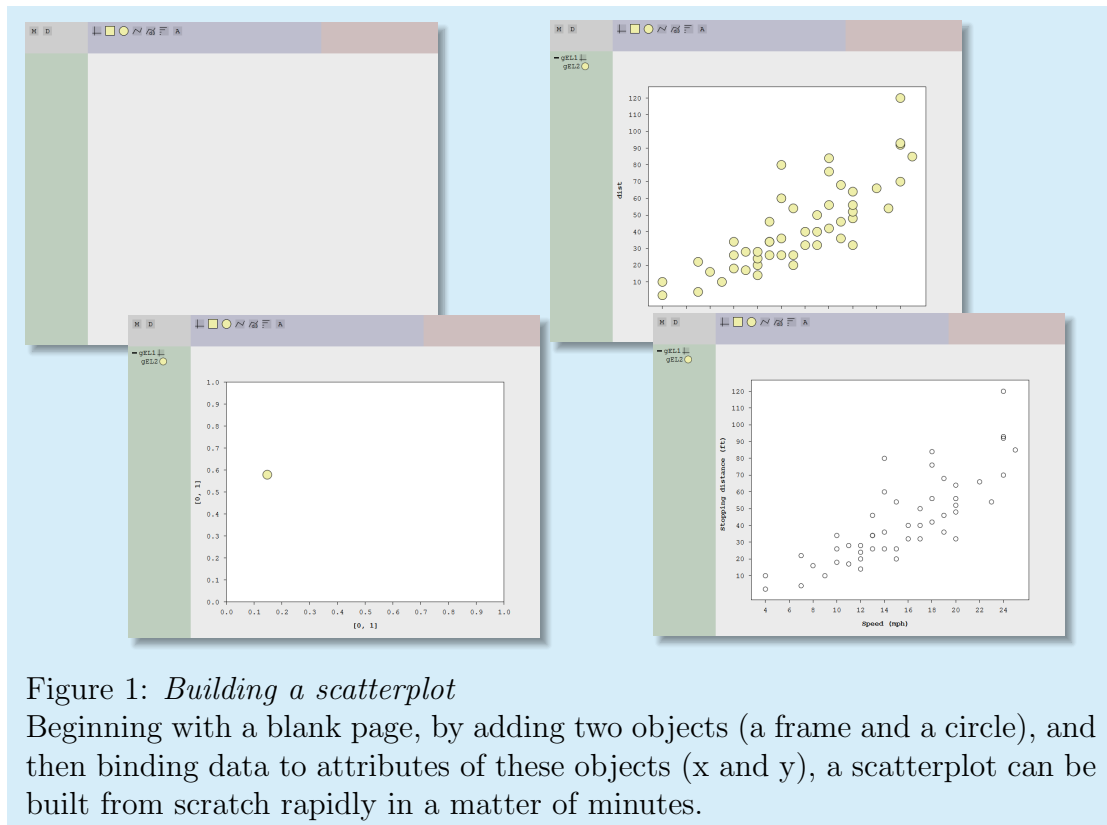
TableToLongForm is an example of both. When I investigated Open Data (data released freely to the public, often by government organisations) I discovered much of the data were released in hierarchical tables unsuitable for analysis with existing statistical software. TableToLongForm is an R package that attempts to automatically convert these tables into machine-friendly longform dataframes, thus making it possible to better utilise the abundant Open Data that is being released globally. A reviewer for the RJournal article for TableToLongForm described it as “a really important contribution to R. These file types are everywhere and they waste a huge amount of time to read for virtually every analyst. I’m unaware of any alternative approaches in R that are anywhere near this easy to use.” At its most basic, TableToLongForm is an efficiency tool, as it is generally possible to convert the tables to computer-friendly formats manually with some time. However when data must be aggregated over hundreds of tables, manual conversion becomes practically infeasible. In these situations TableToLongForm becomes an enabling tool, that now allows analysis over the entire aggregated dataset.

WeBIPP is intended to be an enabling tool, but also works as an efficiency tool. It aims to satisfy the following objectives:

- An interactive, graphical user interface for producing visualisations quickly.
- Does not require coding knowledge to use.
- Easy to utilise data in creating and modifying the visualisations.
- Easy to deploy, by being a purely web-based tool that requires nothing more than a modern browser.
- Extensible and easy for developers to create new addons that extend capability.

- Keep a record of actions taken in creating the visualisations such that steps (not just results) are fully reproducible, and thus reusable.
- Easy to further manipulate the output with other tools, such that using WeBIPP does not limit the user to only what WeBIPP can do.

While still early in development, WeBIPP already satisfies these objectives and is available online requiring nothing more than a modern web browser to use. Due to being a prototype, the degree of satisfaction with which the objectives are met can sometimes be questionable. In its current implementation it is mostly a proof of concept that a tool can exist that satisfies the objectives listed above.



As the two tools above show, I write tools that can do quite different tasks, hence future research I can do is diverse. I am happy to collaborate with other researchers to see if my computing knowledge and skills might be leveraged to contribute to their research. For instance I worked with Rachel Fewster and Emma Carroll to develop a graphical user interface front-end to a powerful algorithm they had developed for R. The target audience were ecologists who desired powerful tools, but did not have the technical expertise to jump in and use R. The GUI solves the problem of accessibility.

Another potential topic I am interested in is leveraging easy-to-deploy interactive web tools for the purposes of teaching. Already, WeBIPP shows potential for teaching statistical graphics to students, as you can quickly build plots from scratch in a matter of minutes, showing their component parts and how they fit together to create the final output. I would like to pursue other ways such tools could be used for interactive teaching, with the eventual goal of making an enabling tool that makes it easy for others to build such interactive tools on their own, even without extensive coding knowledge or expertise.