iPlots eXtreme: next-generation interactive graphics
design and implementation of modern interactive
graphics

Simon Urbanek

Received: 28 October 2009 / Accepted: 3 February 2011 / Published online: 1 May 2011
© Springer-Verlag 2011

Abstract  Interactive graphics provide a very important tool that facilitates the process of exploratory data and model analysis which is a crucial step in real-world applied statistics. Only a very limited set of software exists that provides truly interactive graphics for data analysis, partially because it is not easy to implement. Very often specialized software is created to offer graphics for a particular problem, but many fundamental plots are omitted since it is not considered new research. In this paper we discuss a general framework that allows to create interactive graphics software on a sound foundation that offers consistent user interface, fast prototyping of new plots and extensibility to support interactive models. In addition, we also discuss one implementation of the general framework: iPlots eXtreme—next-generation interactive graphics for analysis of large data in R. It provides most fundamental plot types and allows new interactive plots to be created. The implementation raises interactive graphics performance to an entirely new level. We will discuss briefly several methods that allowed us to achieve this goal and illustrate the use of advanced programmability features in conjunction with R.

Keywords  Interactive graphics · Software · Data analysis

1 Introduction

Interactive graphics play an important role in applied statistics, because the first step in any analysis is to get acquainted with the data in order to be able to devise adequate subsequent process. Interactive graphics are not limited to exploratory data analysis but provide also a powerful tool during the course of choosing and validating models.
However, there are not many modern interactive graphics software packages available that could keep up with the increasing data sizes. Many of the most widely used packages do not offer fully interactive graphics. One of the reasons is that creating a consistent, data-oriented interactive graphics is very difficult and most of the existing software is specialized only on a very specific task.

Our goal was to design a general framework for interactive graphics that is flexible enough to allow design of any useful interactive graphics yet enforces the necessary consistency that is important from the user interface point of view—this aspect is often overlooked. The framework implementation should then allow fast prototyping of new ideas for interactive plots.

Although general frameworks for static graphics have been proposed by Wilkinson (1999), they cannot be easily extended to cover interactive graphics or help with the design thereof. The main reason is that interactivity has to be introduced at a very fundamental level in order to be consistent across all graphics. Other interactive graphics packages such as GGobi Swayne et al. (2003) can be extended programmatically at data and drawing level using plug-ins, but do not provide an interactivity framework for creating new plots. In this paper we describe a general framework for creating interactive graphics that focuses on the interactivity but does not impede with the flexibility of graphics. We also describe both design and implementation of modern, extensible interactive graphics on the basis of the iPlots project.

In the next section we will focus on what constitutes interactive graphics and the challenges ahead. In Sect. 3 we will discuss the design in all three main areas—plots, interactions and linking. The subsequent Sect. 4 introduces the main goals of iPlots eXtreme and highlights implementation details with respect to those goals. It is followed by a Sect. 5 illustrating the use of iPlots eXtreme on examples in R R Development Core Team (2009). Finally, we summarize the main points in Sect. 6.

2 Interactive graphics

Interactive graphics and statistical plots that allow direct interaction of the user with the plot. The most basic functionality that must be supported by interactive statistical graphics are selection (highlighting), queries, zooming, direct change of parameters and multiple views (Unwin 1999). The basic concept underlying selection and highlighting is linking, the most basic requirement for all interactions and ensures that any changes of state (e.g., selection of points) are propagated to all components (the same points are highlighted in all plots). Queries allow us to enrich plots with information without the limitations and clutter associated with static labels. Zooming (especially with logical and censored zooming) allows the user to look at very complex plots without the loss of context. Change of parameters enables the user to interactively influence properties of the plot (such as bin width in histograms) that allow viewing the data in different light. Multiple views have a similar purpose but the different view can be different yet related plots (such as boxplots and dotplots). Interactive graphics are a much larger superset of dynamic graphics (such as continuous projection plots or 3D scatterplots), the latter offer only very limited interactions (in the previous example zooming and change of projection).