1 Tutorial Title

Understanding and creating interactive graphics

2 Instructors

Name	Institution	Address	Email
Toby Dylan Hocking	McGill	Montreal, Canada	toby.hocking@mail.mcgill.ca
Claus Thorn Ekstrøm	Univ. Copenhagen	Copenhagen, Denmark	ekstrom@sund.ku.dk

3 Short Instructor Biography

Toby Dylan Hocking has designed and implemented several R graphics packages. He is the designer and primary maintainer of directlabels (Best Student Poster Prize at useR 2011) and animint (presented in an invited session at Joint Statistical Meetings 2015). He is also the original designer of the ggplot conversion feature of the plotly package.

Claus Thorn Ekstrøm is the creator and contributor to a number of R packages (MESS, Meth-Comp, SuperRanker) and is the author of "The R Primer" book. He has previously given (tutorials on Dynamic graphics in R) and the role of interactive graphics in teaching, and won the C. Oswald George prize for his article "Teaching 'Instant Experience' with Graphical Model Validation Techniques" in 2014.

4 Brief Description of Tutorial

An interactive graphic invites the viewer to become an active partner in the analysis and allows for immediate feedback on how the data and results may change when inputs are modified. Interactive graphics can be extremely useful for exploratory data analysis, for teaching, and for reporting.

Because there are so many different kinds of interactive graphics, there has been an explosion in R packages that can produce them (e.g. animint, shiny, rCharts, rMaps, ggvis, htmlwidgets). A beginner with little knowledge of interactive graphics can thus be easily confused by (1) understanding what kinds of graphics are useful for what kinds of data, and (2) finding an R package that can produce the desired type of graphic. This tutorial solves these two problems by (1) introducing a vocabulary of keywords for understanding the different kinds of graphics, and (2) explaining what R packages can be used for each kind of graphic.

5 Goals

1. Explain and emphasize the role that interactive graphics have in reporting, scientific journals, and in teaching.

2. By showing several examples, explain different categories of graphics: animated, multi-panel, and multi-layer, interactive (direct vs indirect manipulation).

3. Explain how existing R packages can be used to create these different types of graphics.

4. Explain the strengths and weaknesses of the existing R packages, to highlight directions for future work.

6 Detailed Outline

6.1 A vocabulary for understanding interactive graphics, 30 minutes

In this section we will give a high-level introduction about interactive graphics, without going into details about R code for specific packages.

Motivation: interactive graphics in exploratory data analysis, reporting results and teaching.

Vocabulary for describing interactive graphics:

interactive user interaction changes what is displayed. Useful when there are many similar plots for different data subsets, but you **don't** want to see them all at the same time.

> **direct manipulation** interacting with plot elements (lines, points, etc).
>
> **indirect manipulation** interacting with keyboard, mouse clicks on widgets (buttons, menus, etc).

animated An animated graphic automatically advances over time, like a video. Animated graphics are most useful when data sets have a time dimension. The only interaction possible is moving forward and backward in time.

multi-layer A multi-layer graphic uses several geometric elements to show several data sets and/or variables. Multi-layer plots are useful for showing relationships between data sets and/or variables.

multi-panel A multi-panel graphic shows different things in different panels (sub-plots) which each have their own axes (perhaps different from each other). Useful when there are many similar plots for different data subsets, and you **do** want to see them at the same time. Also useful for showing different plots with aligned axes.

Compare and contrast:

interactive vs animated only interaction possible in an animated graphic is moving forward and backward in time (animated graphics are thus a subset of interactive graphics).

interactive vs multi-panel both useful for many similar plots with different data subsets. Do you want to see all the subsets at the same time? (yes=multi-panel, no=interactive)

6.2 Quiz questions

The previous section introduced a vocabulary for describing interactive graphics. In the following section, after showing a new graphic, we will ask the audience to take 1 minute to discuss with their neighbour about which vocabulary words can be used to describe that graphic.

6.3 Creating interactive graphics using R packages

In this section we will show specific R code examples from the various packages.

High-level interactive plotting packages, 30 minutes

- Simple approaches like rotating plots (rgl package) and simple user interaction (wallyplot from MESS package).
- Interactive bar plots (rCharts, several different JavaScript interfaces, interfacing with JavaScript libraries to change axes and legends)
- Interactive scatter plots showing happiness and tax rate (rCharts, and clickme packages, several different JavaScript interfaces, add dropdown effects and improve tooltips)
- interactive maps and choropleths (the rMaps packages)
- Discussion of frustrations that new users unfamiliar with JavaScript may encounter when interfacing with JavaScript libraries

Interactive graphics with shiny and plotly, 30 minutes

- Teaching least squares estimation (shiny)
- Teaching power calculations (shiny)
- Reproducing some of the previous graphics on happiness and tax rate in plotly (ggplot2, and ggplotly, adding tooltips/hover effects, and dropdown)
- Graphics on prediction accuracy for Danish population predictions (plotly, adding sliders)

Multi-layer graphics, ggplot2 package, 15 minutes

- A map that shows a circle for every city, and a line for borders of each country.
- A plot of a linear model that shows data as circles, a regression line, and model residuals as line segments.

Multi-panel graphics, facets in ggplot2, 15 minutes useful in two different situations:

Same plot for different data subsets a linear model fit to each of several data subsets.

Different plots with aligned axes World Bank data viz with one time series panel, and one scatterplot panel.

Animated graphics, animation package, 15 minutes

- Gradient descent (time=iterations).
- Two-panel World Bank data viz (time=years).

Interactive + animated + multi-panel + multi-layer, 45 minutes a few packages are able to produce complex graphics which can be described by several vocabulary words.

shiny + ggplot2 World Bank data viz, interacting with widgets changes selected year, countries, regions.

shiny + ggvis same kind of graphic with World Bank data.

animint World Bank data viz, direct manipulation changes selected year, countries, regions.

7 Justification

The role of graphics is expanding and is moving away from simple static representations found in scientific journals to more interactive representations where the user is directly involved in exploring different facets of the data. In that sense, the reader indirectly takes on the role of the analyst, and R is the ideal tool to produce integrated, interactive graphics and for interfacing with some of the external graphics libraries that exist.

The various implementations for interactive graphics found in R packages are often highly specialized in the same way as high-level plots, and each package typically has a completely different syntax. This tutorial will

- introduce a vocabulary for categorizing interactive graphics,

- present practical examples of how to produce interactive graphics using existing R packages,

- describe how to overcome frustrations typical of new users to interactive graphics, and

- highlight advantages and room for improvement in existing packages.

8 Background Knowledge

Since we plan to present state-of-the-art interactive graphics, people should know how to use R data structures (lists, data.frames) and the ggplot2 package.

Even though many examples will be interactive web graphics, we will assume only knowledge of R, not HTML/JavaScript.

There are two classes of potential attendees:

- UseRs who are not very familiar with interactive graphics should benefit the most, since we will give a high-level overview of many different packages.

- DevelopeRs of interactive packages are encouraged to come, to discuss the current state-of-the-art and future directions.

9 Expected Number of Attendees

We're not sure how to estimate this, but typically interactive graphics are a popular topic so it would be best to have a decent sized lecture hall (50-150 people).