Stories of Two Decades of Efforts to Build Interactive Graphics Capacity into R

Di Cook
Econometrics and Business Statistics
Monash University, Melbourne, Australia
“Your name came up in discussions with my colleagues as you are able to offer the conference a number of things we would like: You are not an old male university professor, you have a strong insight to exploring and communicating with data and you are a user and developer of software.”

Torben Tvedebrink’s email invitation
“If we don’t actively and intentionally set out to include women, we will unintentionally exclude them”

Elizabeth Broderick, Australia’s Sex Discrimination Commissioner.

bit.ly/WhatCanIDoToo
“Your name came up in discussions with my colleagues as you are able to offer the conference a number of things we would like: You are not an old male university professor, you have a strong insight to exploring and communicating with data and you are a user and developer of software.”

Torben Tvedebrink’s email invitation
The gold standard, which we have not seen again yet
Some examples from 1992 through to today
Really new developments
Challenges to the young R developers
R has been 87 steps forward for data analysis but 12 steps backwards for interactive graphics, from where XLispStat (and perhaps Data Desk) had put the field in the 1990s.
Gold standard

“An integrated environment for statistical calculations and graphics is essential for developing an understanding of the uses of dynamic graphics in statistics and for developing new graphical techniques.”

XLispStat

http://homepage.stat.uiowa.edu/~luke/xls
“XLISP-Stat is a statistical programming system along the lines of S. It doesn't (yet) have all the features of S but it's faster, free, has better support for dynamic graphics and is being developed quite quickly.”

Thomas Lumley, early 90s
http://faculty.washington.edu/tlumley/xlispad.html
“We also had to give up some XLISP-STAT components which still have not been replicated in R. Dynamic graphics, for instance, and the byte compiler. R has to do its dynamic graphics by making calls to the standalone xgobi or ggobi programs, and it does not have tools to do dynamic graphics programming yet. I am sure this will come at some point in time, ...”

Jan DeLeeuw, 2005
http://www.jstatsoft.org/v13/i07/paper
Programmability

(load "Examples/tour.lsp")

(def normdat (quote((0 1 0 1 0 1 0 1 0 1)
(tour-plot normdat)

http://www.public.iastate.edu/~dicook/JSS/paper/paper.html
To the point of building systems like Vista...

Create a new menu for messages to a regression model

Add some items

Actions associated with menu actions
Molina, Ledesma, Valero, Young

http://www.jstatsoft.org/v13/i08/paper
Did you notice??

Activity on the plot itself, and a variety of actions and linking
If you think the interface style looked funny, take a look at Jerry Friedman’s hair from two decades before that.
XGobi meets S (Swayne, Buja, Hubbell)

- rggobi, RGtk
- iplots
- RGtk2
- cranvas
- shiny

Orca

G3Gobi

GUI development: tcl/tk, RGtk2, gwidgets
1992

XGobi meets S

- Interprocess communication - new instances of xgobi were started using the unix function in S
- Limited set of functions available

Swayne, Buja, Hubbell
XGobi meets S

- Two examples
- XSmooth
- XClust
XSmooth

- Control panel and a canvas
- Connection to S
- Controller forms the smoother command, sends it to S, and captures the response, and displays
XClust

- Control panel
- S graphics window
- XGobi instance or two
- Controller sends S commands, captures results, passes pieces to XGobi
Controller is really an interface to the hclust function in R
Number of clusters selected, labels passed to xgobi to color points by membership

Second xgobi instance would be called to display discriminant projection
XGobi: Data Pipeline

- Process the data from original form into graphical elements on the screen

(Provide mechanisms for interaction)
XGobi interaction

- Main loop controlled updates based on user action
- E.g. brushing, the screen was re-drawn
- E.g. tour, the pipeline from world to screen was re-computed
- Separate data structures held appearance attributes

```
RAW --> WORLD --> PLANAR --> SCREEN
p-D --> p-D unit-free --> d-D --> 2-D
```
XGobi features

- DRAWBACK: Only one plot window
- DRAWBACK: Tour only 2D projections
- COOL: Fast brushing, using pre-processed gridded scatterplots
- COOL: Linked list for the tour, so user could play it like a movie
- COOL: Inference using permutations
- COOL: Multidimensional scaling add-on
- COOL: High-dimensional drawing
3 GGobi's graphical user interface

GGobi's graphical user interface uses a toolkit called GTK+ (see www.gtk.org). It is used extensively by Linux programmers, which ensures a high level of quality and state of the art GUI technology. The toolkit has been ported to other platforms, though, making GGobi more portable than XGobi. It also provides GGobi with a cleaner and look and feel than XGobi, since XGobi was written using a toolkit whose development was frozen in the early 1990s.

In the following sections we discuss a few GUI-related topics, such as the supported display types and their operations, and the linking model for multiple displays.

XGobi meets S
Orca
GGobi
rggobi, RGtk
iplots
RGtk2
cranvas
shiny


S ---------------
R -----------------
grid ------------------------

GUI development: tcl/tk, RGtk2, gwidgets ---
GGobi

- Re-design of XGobi
- New widget set: gtk, extensive array of GUI elements, drawing capabilities and portability
- One control, multiple displays
- Actions on the plot window, needed to actively tell ggobi which plot was the focus

Swayne, Temple Lang, Buja, Cook, re-worked by Wickham, Lawrence, Hofmann
GGobi features

- XML data description
- Handling of missing values
- Plugins!
- Arbitrary dimension tour projections
RGtk/2

- Wrappers to gtk functionality to R
- Enabled the connection between ggobi and R

“RGtk2: The nature of the R event loop prevents the continuous execution of the GTK main loop, thus preventing things like timers and idle tasks from executing reliably. This manifests itself when using functionality such as GtkExpander and GtkEntryCompletion”

http://www.ggobi.org/rgtk2/
rggobi

- Exposure of a limited set of internal ggobi data structures from R
- C functions
- R wrappers

g <- ggobi(iris)
clustering <- hclust(dist(iris[,1:4]),
    method="average")
glyph_colour(g[1]) <- cuttree(clustering, 3)
Multiple imputations

```r
> library(norm)
...
> d.tao.impute.93 <- imp.norm(d.tao.nm.93, theta.93, d.tao.93)
> d.tao.impute.97 <- imp.norm(d.tao.nm.97, theta.97, d.tao.97)
> gd[, "sea.surface.temp"] <- c(
    d.tao.impute.97[, "sea.surface.temp"],
    d.tao.impute.93[, "sea.surface.temp"])
...
```
> d.music.som <- f.ggobi.som(subset(d.music.std,
    select=lvar:lfreq), music.som)

> gd <- ggobi(d.music.som)[1]
> d.music.som.net <- f.ggobi.som.net(music.som)
> edges(gd) <- d.music.som.net + 62
> gcolor <- rep(8,98)
> gcolor[d.music.som$Type=="Rock"] <- 6

...
XGobi meets S

1990

1995

2000

2005

2010

2015

S

-----

R

-----

grid

-----

GUI development: tcl/tk, RGtk2, gwidgets ---

GGobi

1995

2005

2015

Orca

1990

rggobi, RGtk

iplots

RGtk2

cranvas

shiny
cranvas

- qt widget set - speed, but trouble with portability
- R wrappers to qt functionality
- Fixed set of plot types, but new could be programmed directly in R
- Have to program plots from first principles
- Brushes and linking defined
- Linking by mutable objects (plumbr), and reference classes (ObjectSignals)
DATA

Mutaframe

Features:
- e.g. brushed color, size

Scatterplot

Meta -- scatterplot

Histogram

Meta -- histogram

Density Plot

Meta -- density plot

Parallel Coordinates Plot

Meta -- para coord plot

qscatter()

qhist()

qdensity()

qparallel()
## events

brush_mouse_press = function(layer, event) {
    common_mouse_press(layer, event, data, meta)
}

tree = createTree(meta$xy)  # build a search tree
brush_mouse_move = function(layer, event) {
    ...
}
brush_mouse_release = function(layer, event) {
    ...
}
key_press = function(layer, event)
    run_handler(meta$handlers$keypress, layer, event)
key_release = function(layer, event)
    run_handler(meta$handlers$keyrelease, layer, event)

mouse_wheel = function(layer, event) {
    ...
}

identify_hover = function(layer, event) {
    ...
}
library(cranvas)
data(nrcstat)
nrcstat$Inst.Prg <- paste(nrcstat$Institution, nrcstat$ProgramName)
nrcdist <- dist(scale(nrcstat[,c(20,21,26,30,32,33,34,36,41,43,46)]))
nrc.hc <- hclust(nrcdist, method="ward")
plot(nrc.hc)
nrc.clust <- data.frame(nrc, cl2=cutree(nrc.hc, 2),
                       cl3=cutree(nrc.hc, 3),
                       cl4=cutree(nrc.hc, 4),
                       cl5=cutree(nrc.hc, 5),
                       cl6=cutree(nrc.hc, 6),
                       cl7=cutree(nrc.hc, 7),
                       cl8=cutree(nrc.hc, 8))
qnrc.clust <- qdata(nrc.clust)
qparallel(c(20,21,26,30,32,33,34,36,41,43,46), data=qnrc.clust,
          center = median, horizontal=T, glyph = "tick")
qparallel(80:74, data=qnrc.clust, horizontal=T,
          jitter=c("cl2","cl3","cl4","cl5","cl6","cl7","cl8"))
qscatter(RRankings5th, SRankings5th, data=qnrc.clust)
record_selector(Inst.Prg, qnrc.clust)
Parallel coordinate plot of ranking criteria

Dendrogram displaying results from 2-8 clusters

Scatterplot of ranks by two different schemes

List of programs

<table>
<thead>
<tr>
<th>List of programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSTON UNIVERSITY Biostatistics</td>
</tr>
<tr>
<td>CARNEGIE MELLON UNIVERSITY Statistics</td>
</tr>
<tr>
<td>CASE WESTERN RESERVE University Statistics</td>
</tr>
<tr>
<td>COLORADO STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK Biostatistics</td>
</tr>
<tr>
<td>CORNELL UNIVERSITY Biostatistics</td>
</tr>
<tr>
<td>DUKE UNIVERSITY Biostatistics</td>
</tr>
<tr>
<td>FLORIDA STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>GEORGE WASHINGTON UNIVERSITY Biostatistics</td>
</tr>
<tr>
<td>IOWA STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>JOHN HOPKINS UNIVERSITY Applied Mathematics</td>
</tr>
<tr>
<td>KANSAS STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>MICHIGAN STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>NORTH CAROLINA STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>NORTH DAKOTA STATE UNIVERSITY Main Campus</td>
</tr>
<tr>
<td>NORTHWESTERN UNIVERSITY Statistics</td>
</tr>
<tr>
<td>OHIO STATE UNIVERSITY Main Campus</td>
</tr>
<tr>
<td>OHIO STATE UNIVERSITY Main Campus Scott College of Engineering</td>
</tr>
<tr>
<td>OKLAHOMA STATE UNIVERSITY Main Campus</td>
</tr>
<tr>
<td>OREGON STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>PENN STATE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>PURDUE UNIVERSITY Main Campus Statistics</td>
</tr>
<tr>
<td>RICE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>RUTGERS THE STATE UNIVERSITY OF NEW JERSEY Statistics</td>
</tr>
<tr>
<td>SOUTHERN METHODIST UNIVERSITY Statistics</td>
</tr>
<tr>
<td>STANFORD UNIVERSITY Statistics</td>
</tr>
<tr>
<td>TEMPLE UNIVERSITY Statistics</td>
</tr>
<tr>
<td>TEXAS A &amp; M UNIVERSITY Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA–BERKELEY Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA–DAVIS Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA–LOS ANGELES Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA–RIVERSIDE A Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA–SANTA BARBARA Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CHICAGO Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF CONNECTICUT Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF FLORIDA Statistics</td>
</tr>
<tr>
<td>UNIVERSITY OF GEORGIA Statistics</td>
</tr>
</tbody>
</table>
qd <- qdata(data, alpha = alpha, size = size)

add_listener(qd, function(i, j){
  if(j == "\.brushed")
    ...
    plotInter(.data, i[1], line = line, error = FALSE) + ylim(ylim) + s
  ...
}
cranvastime

See examples of working with time series and maps at https://vimeo.com/chxy/videos

See also Loon (Waddell and Oldford)
XGobi meets S

GGobi

Orca

rggobi, RGtk

iplots

RGtk2

cranvas

shiny


S

R

ggrid

GUI development: tcl/tk, RGtk2, gwidgets ---
RStudio

- Great for getting up and running
- Organizing your work

- BUT, its back to ONE SINGLE WINDOW for GRAPHICS....
library(scales)
library(tourr)
library(ggvis)
library(shiny)

aps <- 2
fps <- 30

mat <- rescale(as.matrix(flea[1:6]))
tour <- new_tour(mat, grand_tour(), NULL)
start <- tour(0)

proj_data <- reactive({
  invalidateLater(1000 / fps, NULL);
  step <- tour(aps / fps)
  data.frame(center(mat %*% step$proj), species = flea$species)
})

proj_data %>% ggvis(~X1, ~X2, fill = ~species) %>%
  layer_points() %>%
  scale_numeric("x", domain = c(-1, 1)) %>%
  scale_numeric("y", domain = c(-1, 1)) %>%
  set_options(duration = 0)
```r
library(tourr)
library(animint)
mat <- rescale(as.matrix(flea[1:6]))
tour <- new_tour(mat, grand_tour(), NULL)
tour_dat <- function(step_size) {
  step <- tour(step_size)
  proj <- center(mat %*% step$proj)
  data.frame(x = proj[,1], y = proj[,2], species = flea$species)
}

steps <- c(0, rep(1/15, 200))
stepz <- cumsum(steps)
dats <- lapply(steps, tour_dat)
datz <- Map(function(x, y) cbind(x, step = y), dats, stepz)
dat <- do.call("rbind", datz)

p <- ggplot() + geom_point(data = dat,
  aes(x = x, y = y, colour = species, showSelected = step))
plist <- list(
  plot = p,
  time = list(variable = "step", ms = 100),
  duration = list(step = 200)
)
animint2dir(plist, "tour", open.browser = FALSE)
servr::httd("tour")
```

see also gridSVG (Murrell)
plotly

- python
- ggplot2 re-scripted with interaction

Also see rbokeh (Hafen)
shiny

- Reactive objects
- Building GUIs is easy
- Support for interactive graphics on plot is improving

https://gallery.shinyapps.io/095-plot-interaction-advanced/
Vanderplas, Graham, Cook

Work funded for the USDA

examining soybean

milestone cultivars
Click on data points in the plot to see field trial data.

Yield, Protein, and Oil by Year

More Field Trial Data by Year

Yield, Protein, Oil Pairwise Plots

Yield

Protein

Oil

Maturity

Lodging
Challenges to the young developers

- Interactivity on the plot
- Different types of brushes
- Different kinds of linking between plots
- Programmability
- Strong connection with model fitting
- Portability, easy install, web compatible
- Large quantities of data
- Incorporating inference
- Conceptual framework
This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.