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

Outline

- 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.

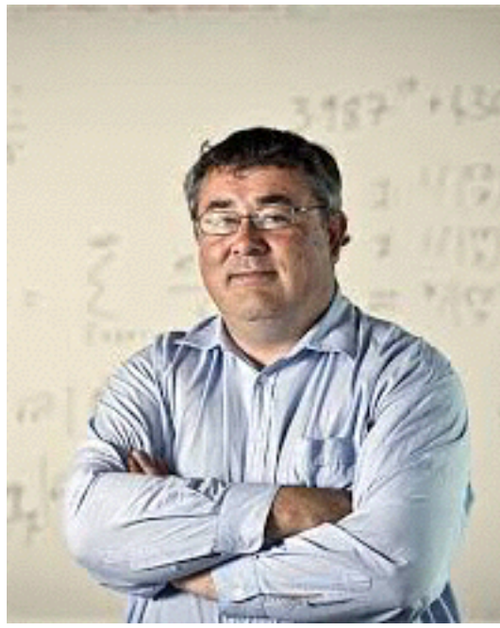
Gold standard

"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."

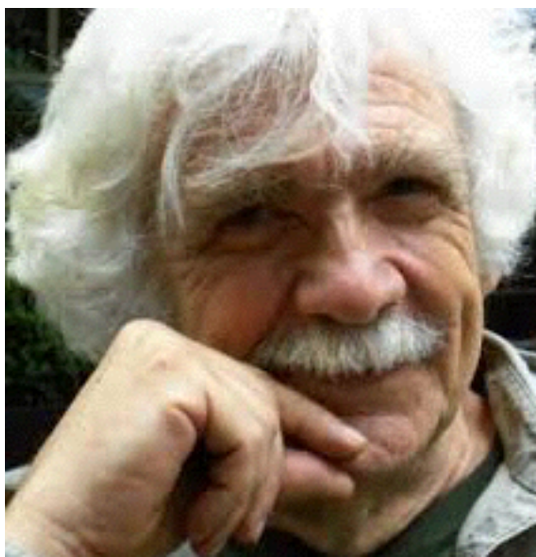


Thomas Lumley, early 90s

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
(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

```
> (setf model-menu (send menu-proto :new "Model"))
#<Object: 4055334, prototype = MENU-PROTO>
> (setf summary (send menu-item-proto :new "Summary"
                      #'(lambda () (send *current-model* :display))))
#<Object: 4034406, prototype = MENU-ITEM-PROTO>
> (setf plot (send menu-item-proto :new "Plot Residuals" :action
                      #'(lambda () (send *current-model* :plot-
residuals))))
#<Object: 3868686, prototype = MENU-ITEM-PROTO>
> (send summary :do-action)

Least Squares Estimates:

Constant                  -16.41924    (7.848271)
.....
NIL
> (send model-menu :append-items summary plot)
NIL
```

Add some items

Actions associated with menu actions

video_15.wmv

ViSta - The Visual Statistics System

File Edit Data Transform Analyze Model Options Help Window

ANOVA Regres UniTest Freqs Boot PrnCmp MulReg

Help In Out Cancel None

Alabama
Alaska
Arizona
Murder (Numer
Rape (Numeric
Robbery (Nume

Molina, Ledesma, Valero, Young

http://www.jstatsoft.org/v13/i08/paper

Idaho
Illinois
Indiana
Iowa
Kansas

Help Save Lock Maximize

Type:MulVar	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto-Thef
Size:50 X 7	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
Alabama	14.20	25.20	96.80	278.30	1135.50	1881.90	280.70
Alaska	10.80	51.60	96.80	284.00	1331.70	3369.80	753.30
Arizona	9.50	34.20	138.20	312.30	2346.10	4467.40	439.50

; Loading DataFile: c:\vistauser\mydata\crime.lsp
; File length is 3321 characters.
; DataStep created data object CrimeRates#2
; OpenData processing time: 1.15 seconds.

video_15.wmv

Principal Components SpreadPlot for CrimeRates

Help Window

Help Mouse Pop Zoom

0.194 PC5 -0.21
0.169 PC4 -0.259
0.391 PC3 -0.25
0.376 PC2 -0.364
0.752 PC1 -0.566

? M Pop Zoom X Y Z Box

Spin
Home
Rock
Clip
Zoom

PC2
PC3
PC1

Up/Dn C/CC L/R Speed

Help Mouse Pop Zoom X Y

PrnCmp
BiPlot

PC2
0.769 0.385 0 -0.385 -0.769
-0.769 -0.385 0 0.385 0.769
PC1

Help

None Cancel Out In

Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois

Help Mouse Pop Zoom

PrnCmp
Box Plot

Data
1 0.5 0 -0.5 -1
PC1 PC2 PC3 PC4 PC5

Bx Dmd Mdn Means Connect

Help Pop Zoom

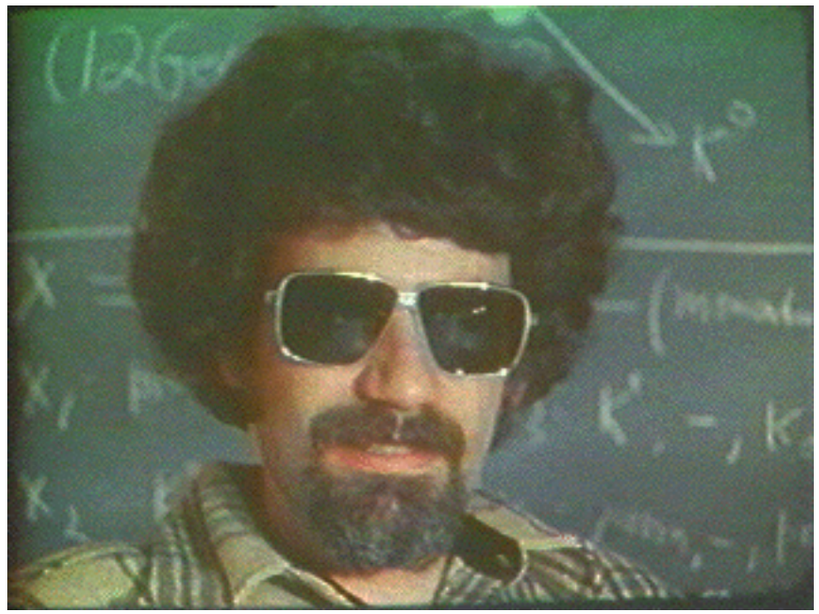
PrnCmp
Scree Plot

.5879

Proportion
0.6 0.4 0.2 0
0 2 4 6 8
Principal Components

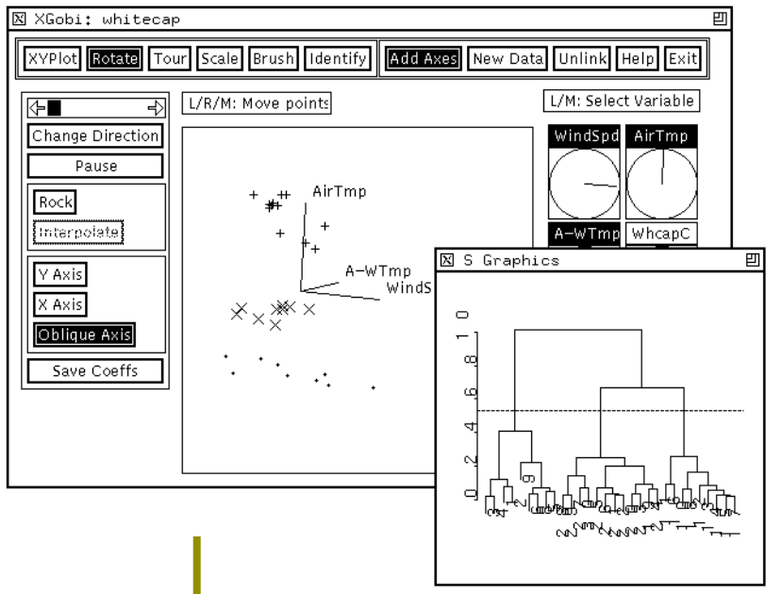
Did you notice??

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

GGobi

Orca

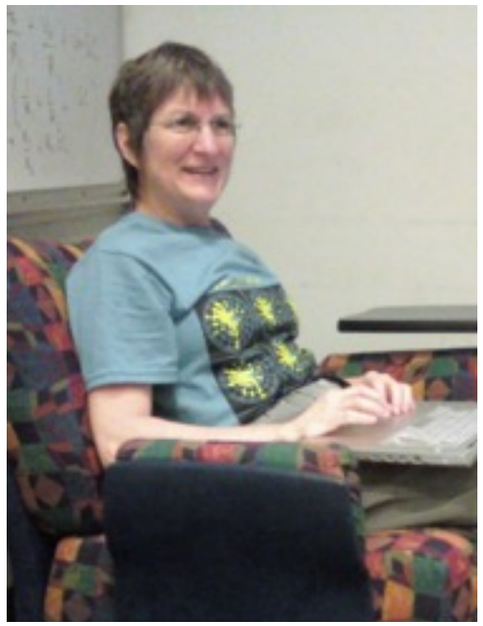
1990 1995 2000 2005 2010 2015

S -|

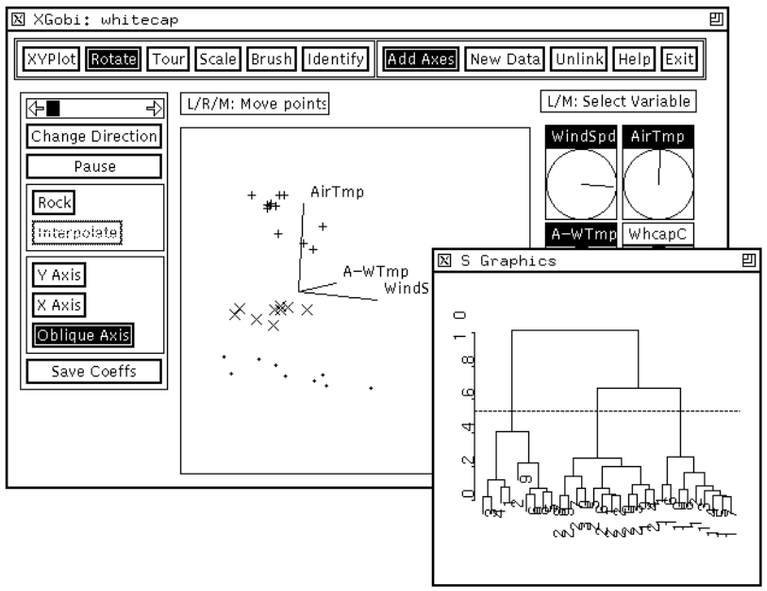
R - - - - - - - - - - - - - - - - - - -

1992

XGobi meets S

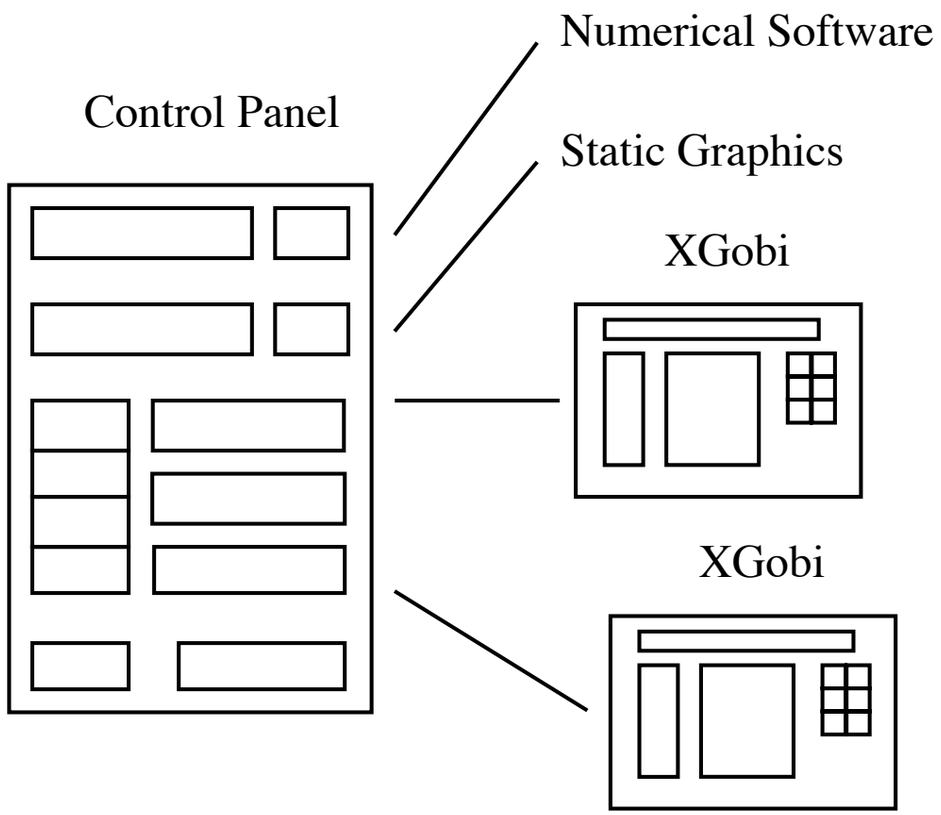


- Interprocess communication - new instances of xgobi were started using the unix function in S
- Limited set of functions available



Swayne, Buja, Hubbell

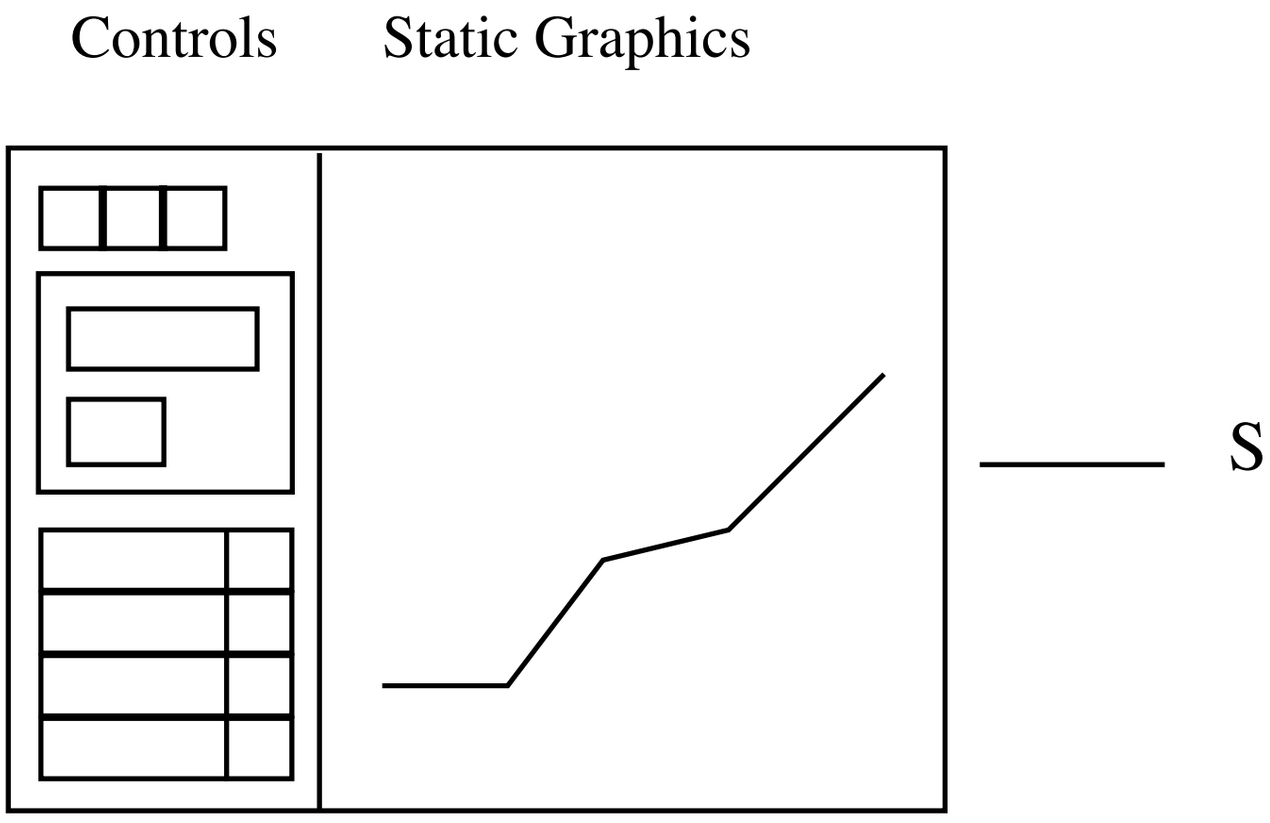
XGobi meets S



General Model

- Two examples
- XSmooth
- XClust

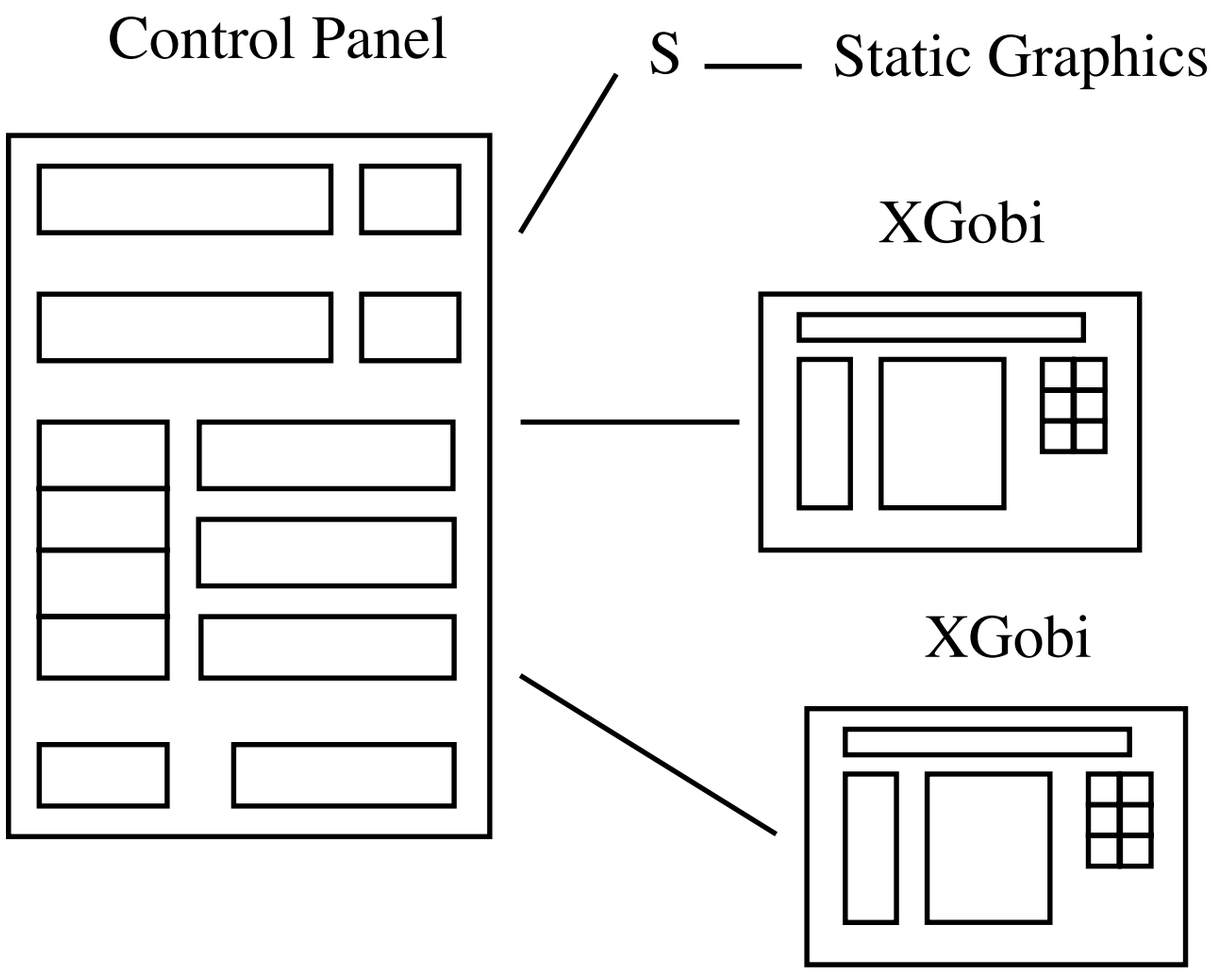
XSmooth



XSmooth Model

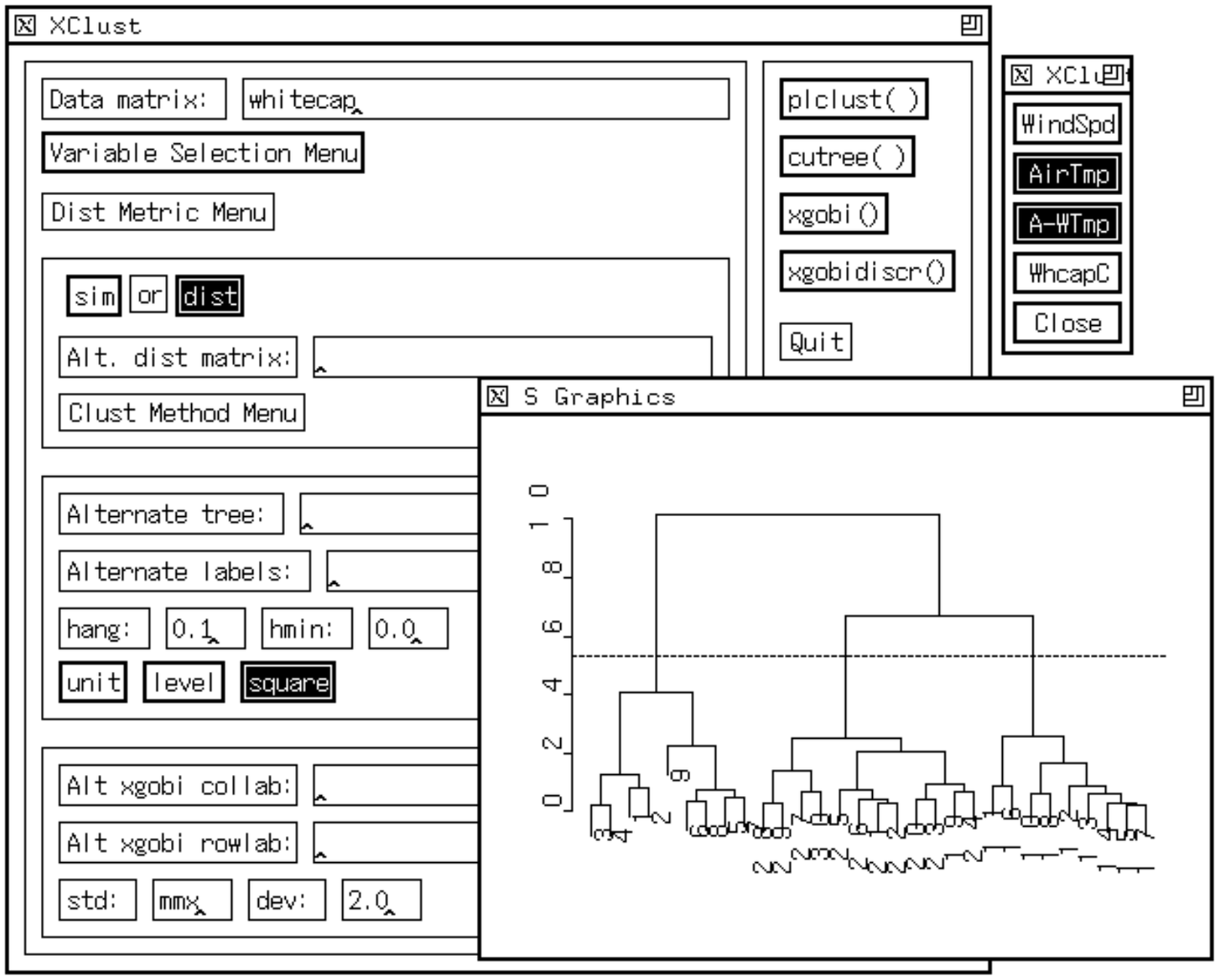
- Control panel and a canvas
- Connection to S
- Controller forms the smoother command, sends it to S, and captures the response, and displays

XClust

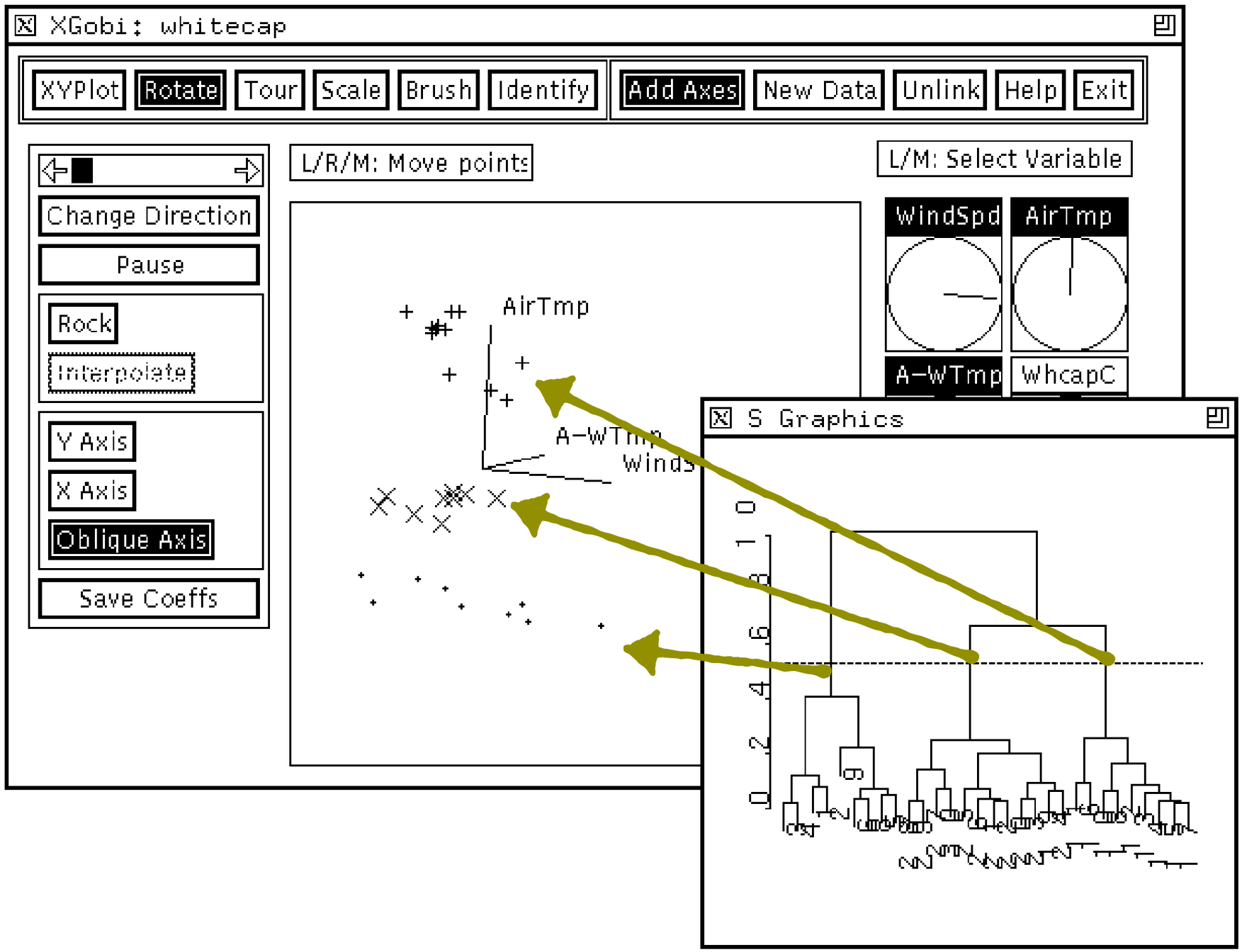


- Control panel
- S graphics window
- XGobi instance or two
- Controller sends S commands, captures results, passes pieces to XGobi

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

src — bash — 156×35

```
dyn-59-191-217-164:src dicook$ ls -asF
total 4904
  0 ./                          24 help.c
  0 ../                         24 hermite.c
  8 .purify                     [illegible]
 16 Corba.nw                    48 identify.c
 16 DrawingA.c                  40 inference.c
  8 DrawingA.h                  16 initialize.c
  8 DrawingAP.h                 16 install-sh*
 16 Imakefile                   24 jitter.c
 40 Makefile                    24 kernel.c
  8 Makefile.corba              24 legendre.c
  8 README.corba                72 line_editor.c
 24 XGobiCorbaServer.idl       152 make.out
 40 XGobiServer.h               48 make_axes.c
  8 aclocal.m4                  72 missing.c
 16 ash1d.c                     32 move_points.c
 48 brush.c                     16 mt19937-1.c
 24 brush_cbacks.c              24 natrl_hermite.c
 56 brush_init.c                24 new_data.c
 48 brush_send.c                24 paint.c
 32 callbacks.c                112 parcoords.c
  8 central_mass.c              48 pipeline.c
160 corr.c                     32 plot1d.c
 16 corr_index.c               48 plot_once.c
 56 corr_pursuit.c            104 prt_plotwin.c
  8 corr_util.c               120 pspline.c
 16 de_indices.c               8 qnorm.c
 16 diffs                      64 read_array.c
  8 dummy.c                   120 read_data.c
 56 exclusion.c                24 redesign
 16 f2c.h                       8 rpc_aiiac.h
 16 getfname.c                  8 rpc_aiiac_err.h
 40 gt_ctls.c                  24 rpc_client.c
dyn-59-191-217-164:src dicook$
```

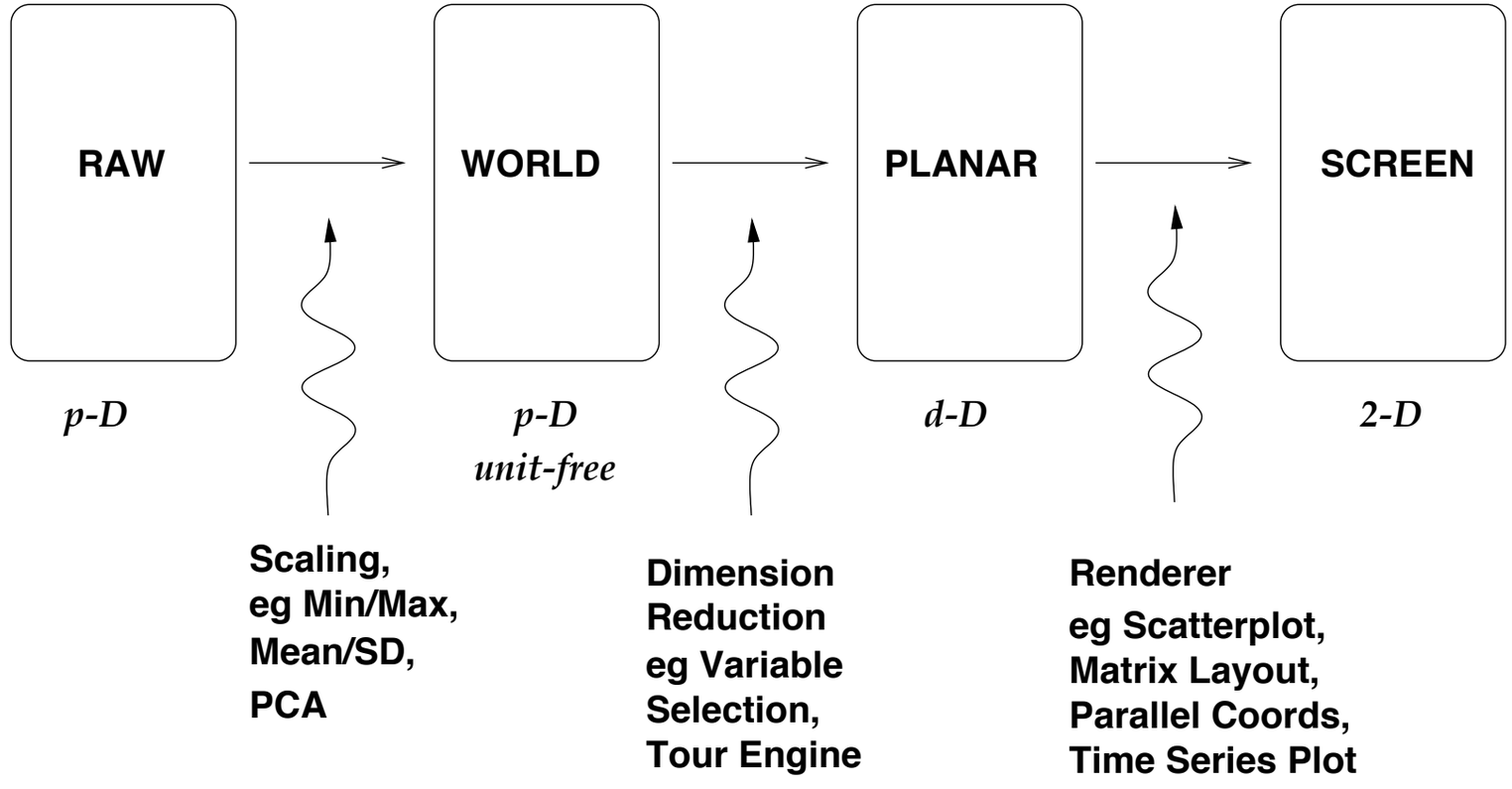
communication code

```
  8 rpc_client.h
  8 rpc_dce.acf
 24 rpc_dce.h
 16 rpc_dce.idl
  8 rpc_dceclient.acf
  8 rpc_dceclient.h
  8 rpc_dceclient.idl
  8 rpc_dceerror.h
 32 rpc_functions.c
 16 rpc_server.h
 40 rpc_server_proc.c
 48 rpc_server_svc.c
 40 rpc_spatial.c
  8 rpc_vars.h
208 rpc_xgobi.c
  8 rpc_xgobi.h
 40 rpc_xpl.c
136 save_data.c
 32 scale_cbacks.c
 24 scaling.c
 16 show_message.c
  8 skewness.c
 48 smooth.c
 24 smooth_fns.c
 16 smoothsk.c
 32 sphere.c
 64 spin.c
 32 spin_cbacks.c
  8 stdize.c
 48 subset.c
 16 svd.c
 24 texture.c
```

```
120 tour.c
 64 tour_cbacks.c
 56 tour_init.c
240 tour_pp.c
 24 tour_section.c
 16 tour_send.c
 16 tour_util.c
 96 transform.c
 16 utils.c
 96 var_panel.c
 48 vc_lists.c
 80 widgets.c
 40 xgobi.c
 64 xgobi_init.c
 72 xgobiexterns.h
 40 xgobitop.h
 40 xgobitypes.h
  8 xgobivars.h
 48 xgv_anchor.c
 40 xgv_cbacks.c
 40 xgv_help.c
 32 xgv_histogram.c
 48 xgv_mds.c
 16 xgv_qsort.c
 48 xgv_read_data.c
 16 xgv_stressplot.c
 72 xgv_widgets.c
 32 xgvis.c
  8 xgvis.h
  8 xincludes.h
 32 xyplot.c
```

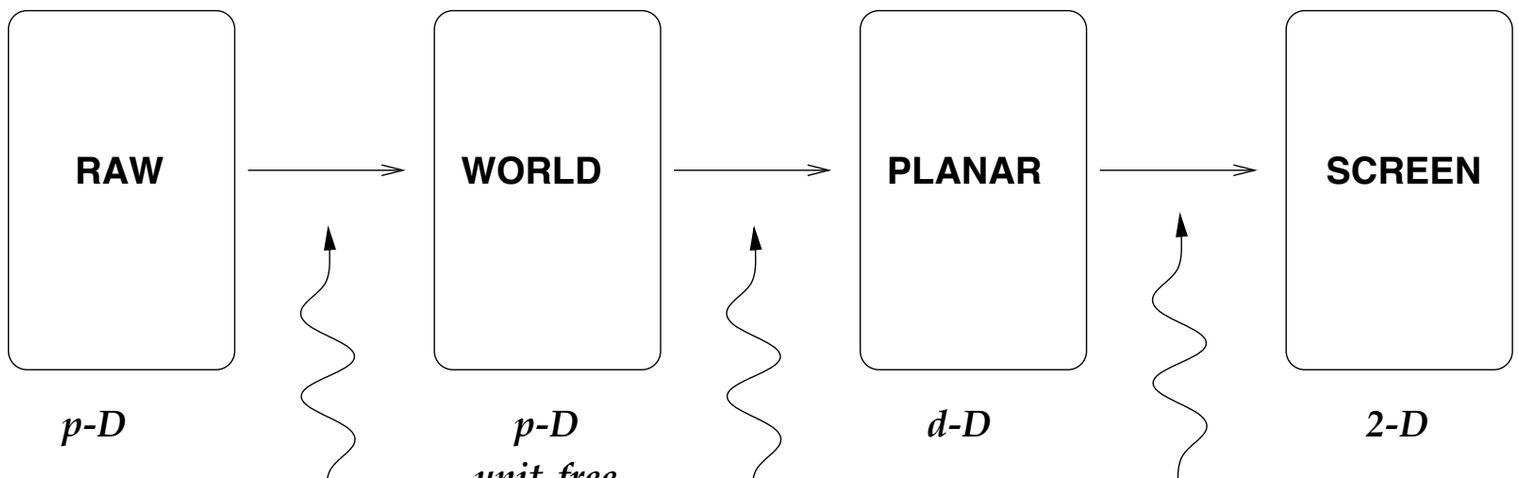
XGobi: Data Pipeline

- Process the data from original form into graphical elements on the screen
- (Provide mechanisms for interaction)



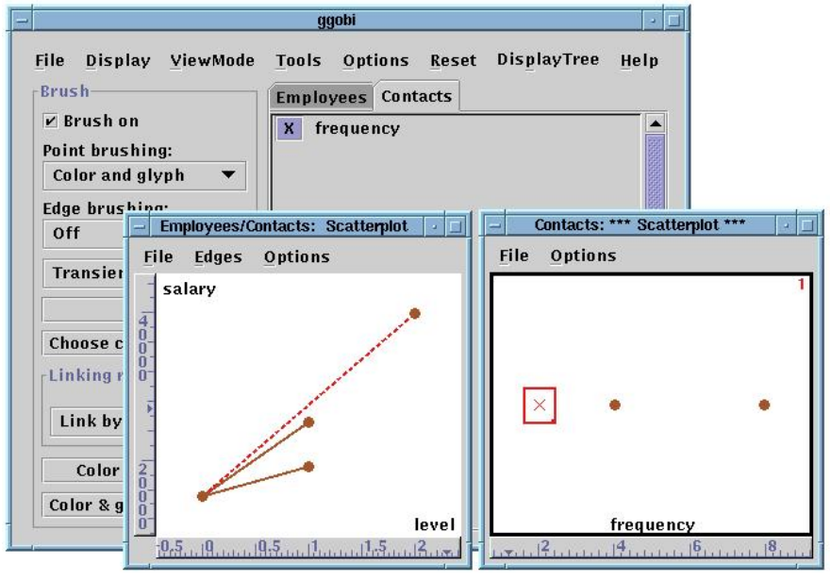
XGobi interaction

- 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



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



rggobi, RGtk

iplots

RGtk2

cranvas

shiny

XGobi meets S

Orca

GGobi

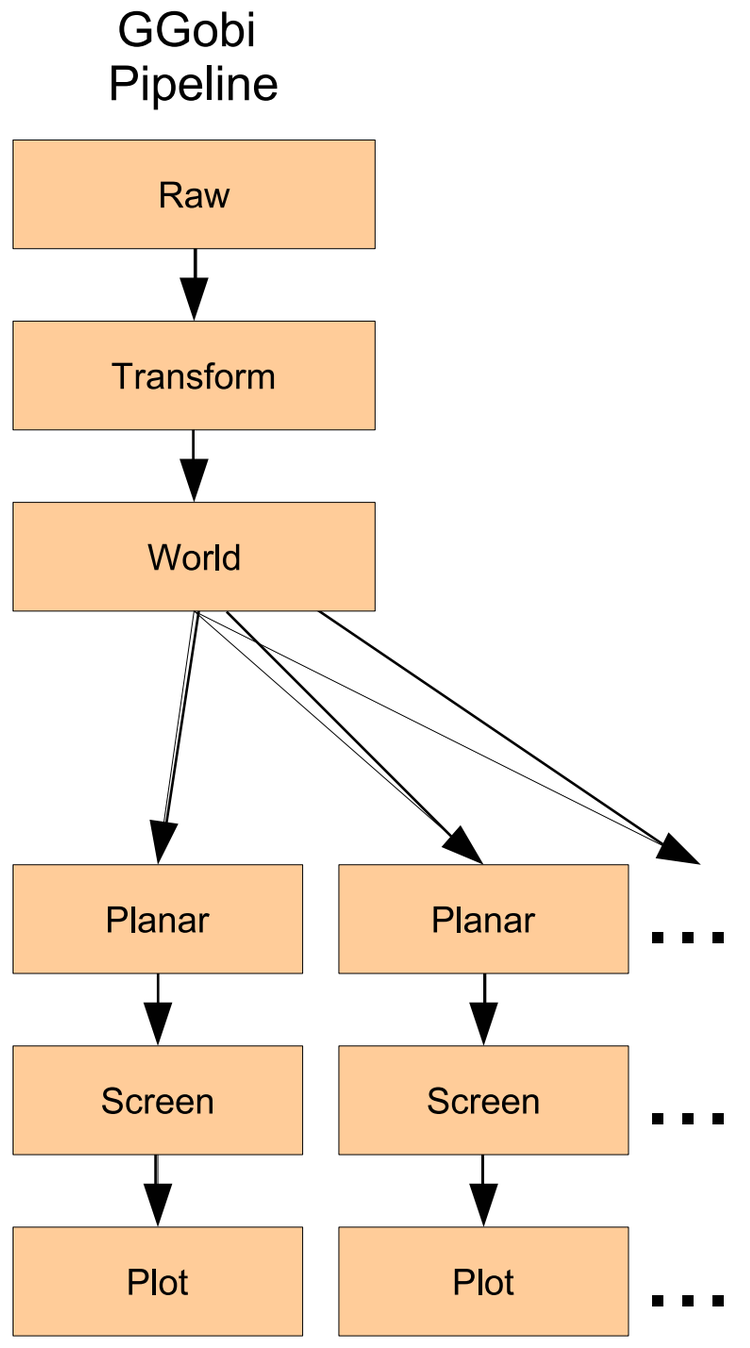
1990 1995 2000 2005 2010 2015

S -|

R - - - - - - - - - - - - - - - - - - -

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

Temple Lang, Lawrence

RGtk/2

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

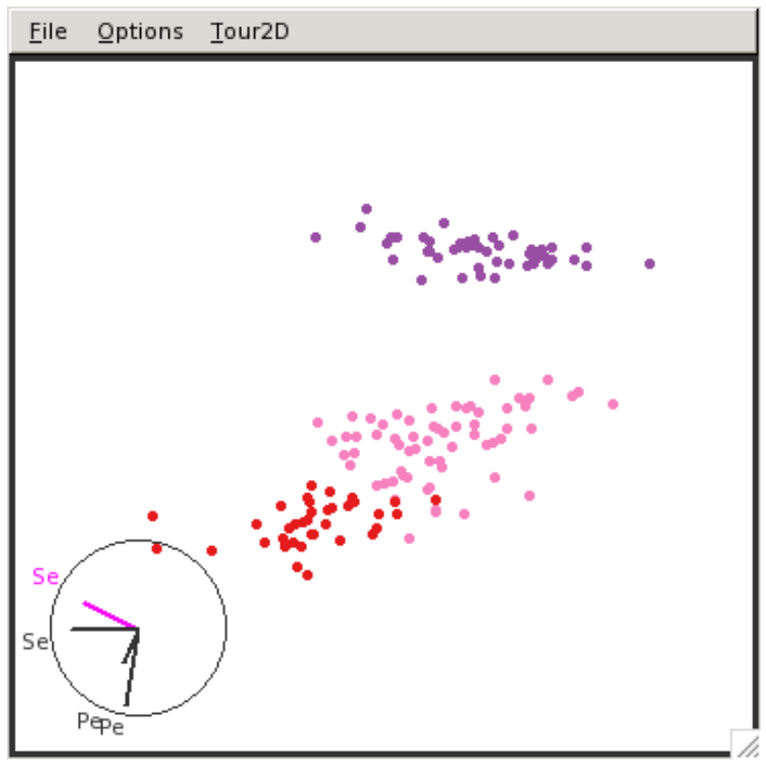
"RGtk2: The nature of the R event loop prevents the continuous execution of the GTK main loop, thus preventing things like timers and idle tasks from executing reliably. This manifests itself when using functionality such as `GtkExpander` *and* `GtkEntryCompletion`*"*
`http://www.ggobi.org/rgtk2/`

Temple Lang, Swayne, Wickham, Lawrence

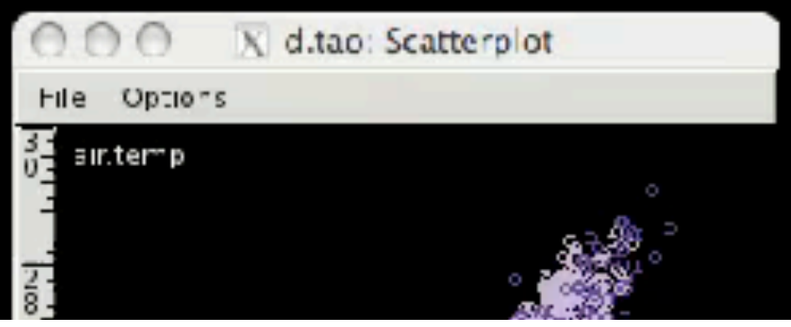
rggobi

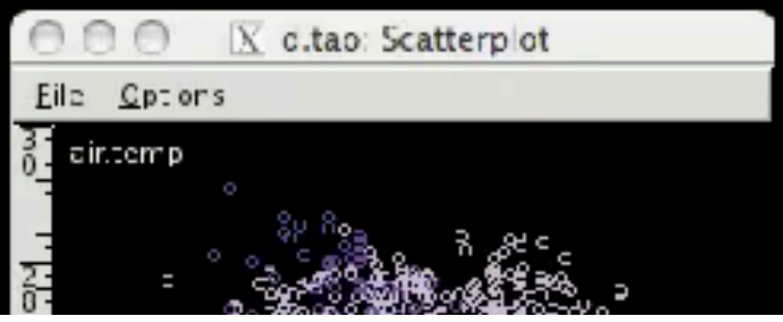
- Exposure of a limited set of internal ggobi data structures from R
- C functions
- R wrappers

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

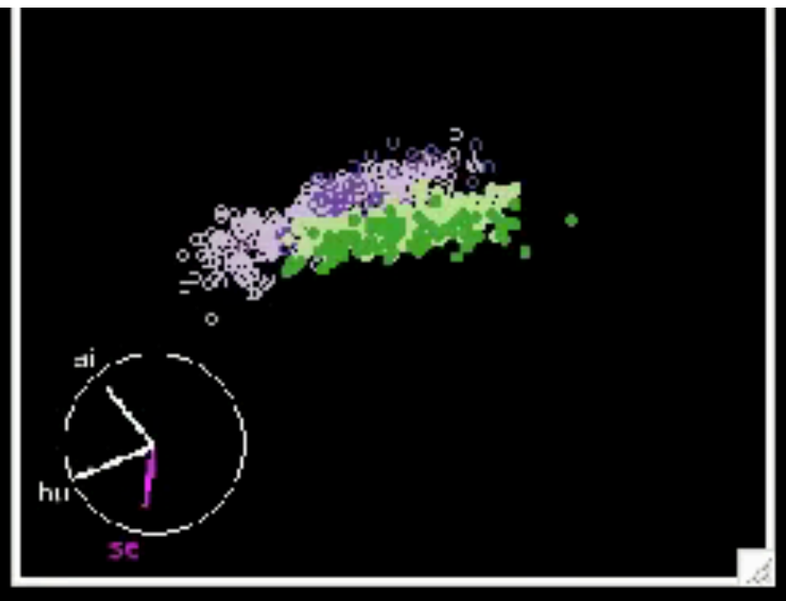


Multiple imputations

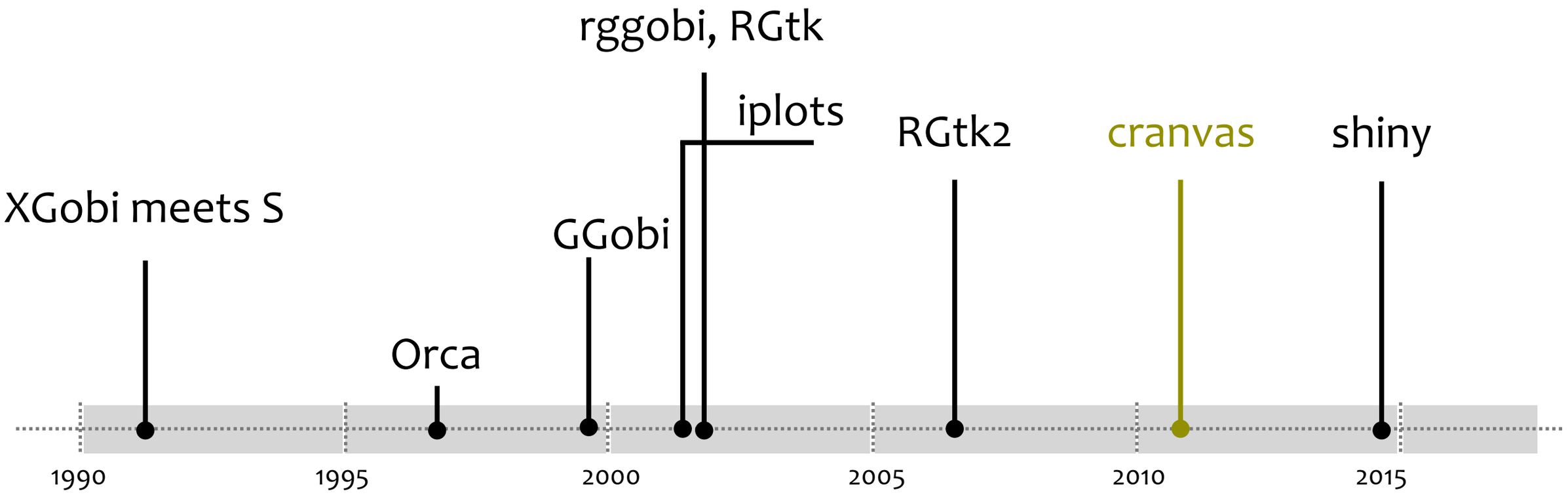





```
> library(norm)
…
> d.tao.impute.93 <- imp.norm(d.tao.nm.93, theta.93, d.tao.93)
> d.tao.impute.97 <- imp.norm(d.tao.nm.97, theta.97, d.tao.97)
> gd[,"sea.surface.temp"] <- c(
    d.tao.impute.97[,"sea.surface.temp"],
    d.tao.impute.93[,”sea.surface.temp"])
…
```



```
> 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
…
```


rggobi, RGtk
iplots
XGobi meets S
GGobi
RGtk2
cranvas
shiny
Orca
1990
1995
2000
2005
2010
2015

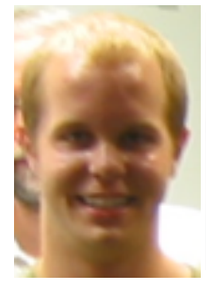
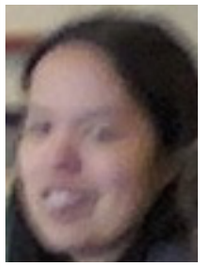
S -|

R - - - - - - - - - - - - - - - - - - -

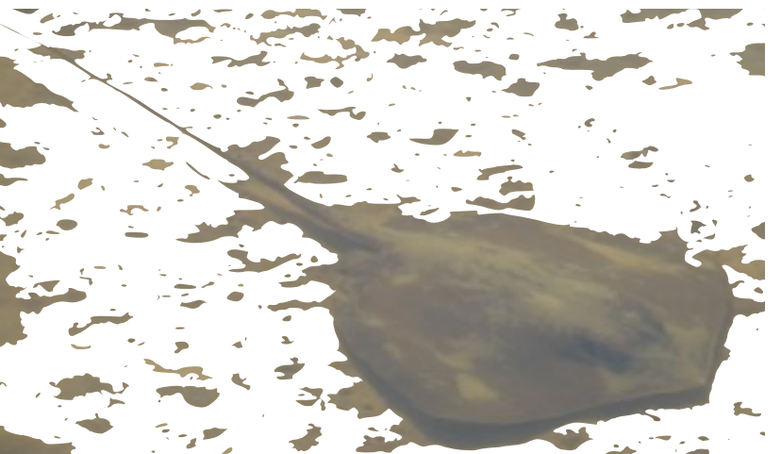
Xie, Cheng, Hofmann, Cook, Schloerke, Vendettuoli,
built on foundation by Lawrence, Wickham

cranvas

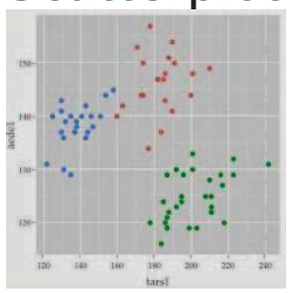
- qt widget set - speed, but trouble with portability
- R wrappers to qt functionality
- Fixed set of plot types, but new could be programmed directly in R

- Have to program plots from first principles
- Brushes and linking defined
- Linking by mutable objects (plumbr), and reference classes (ObjectSignals)



Scatterplot



Meta -- scatterplot

qscatter()

Mutaframe

DATA

Features: e.g. brushed color, size

listener

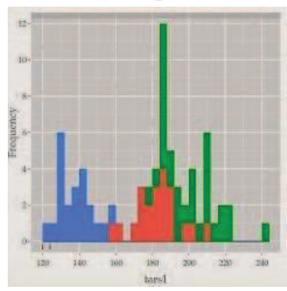
listener

listener

listener

qhist()

Histogram

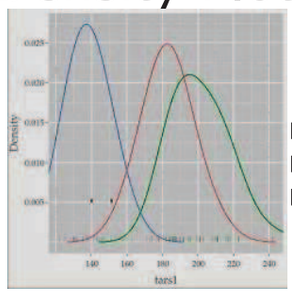


Meta -- histogram

qparallel()

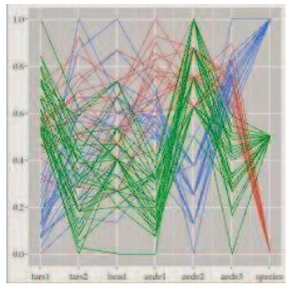
qdensity()

Density Plot



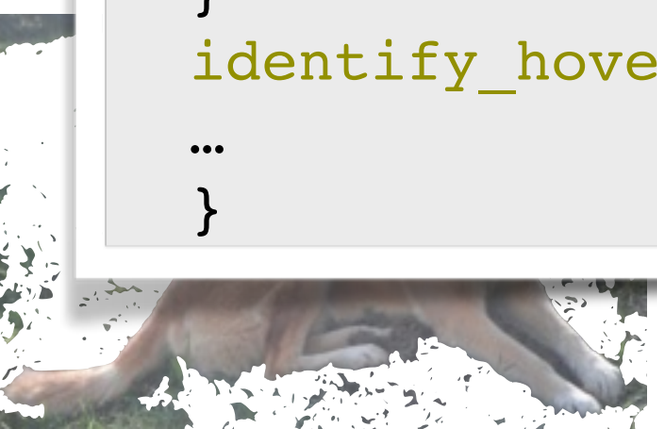
Meta -- density plot

Parallel Coordinates Plot



Meta -- para coord plot

```
## 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=cutr
                        cl3=cutree(n
                        cl4=cutree(n
                        cl5=cutree(n
                        cl6=cutree(n
                        cl7=cutree(n
                        cl8=cutree(n
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)
```

Load library, data, add better id variable

Set up data for cranvas

Use Euclidean distance

Open a pcp on ranking criteria, pcp for dendrogram, scatterplot, and label browser

iShowU-Capture71.mov

Parallel coordinate plot of ranking criteria

Dendrogram displaying results from 2-8 clusters

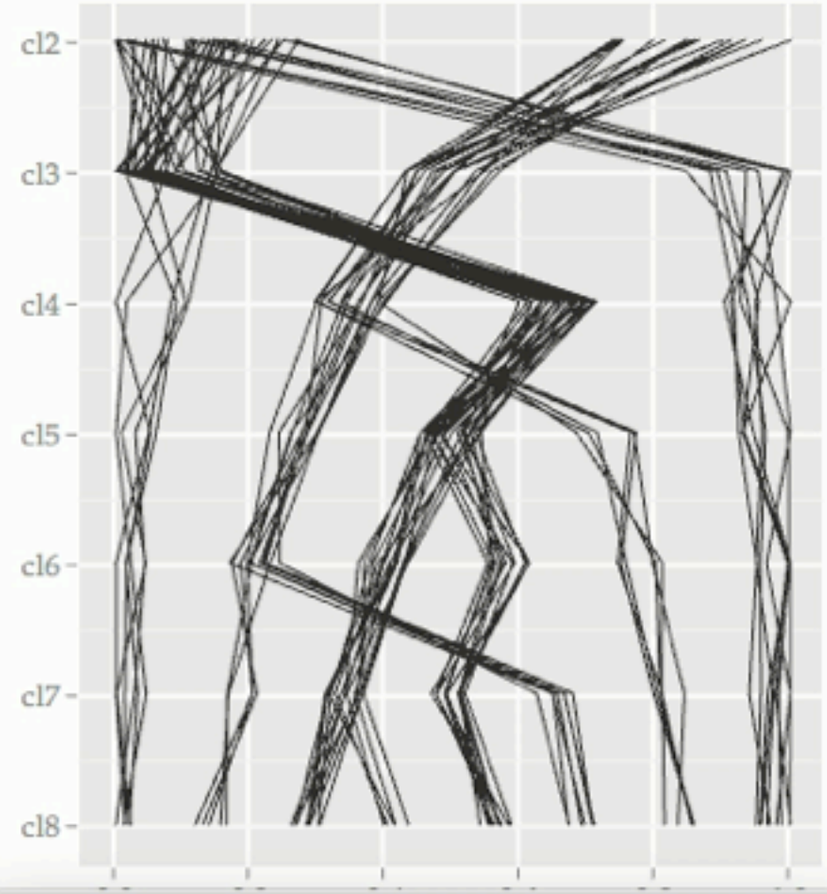
List of programs

Scatterplot of ranks by two different schemes

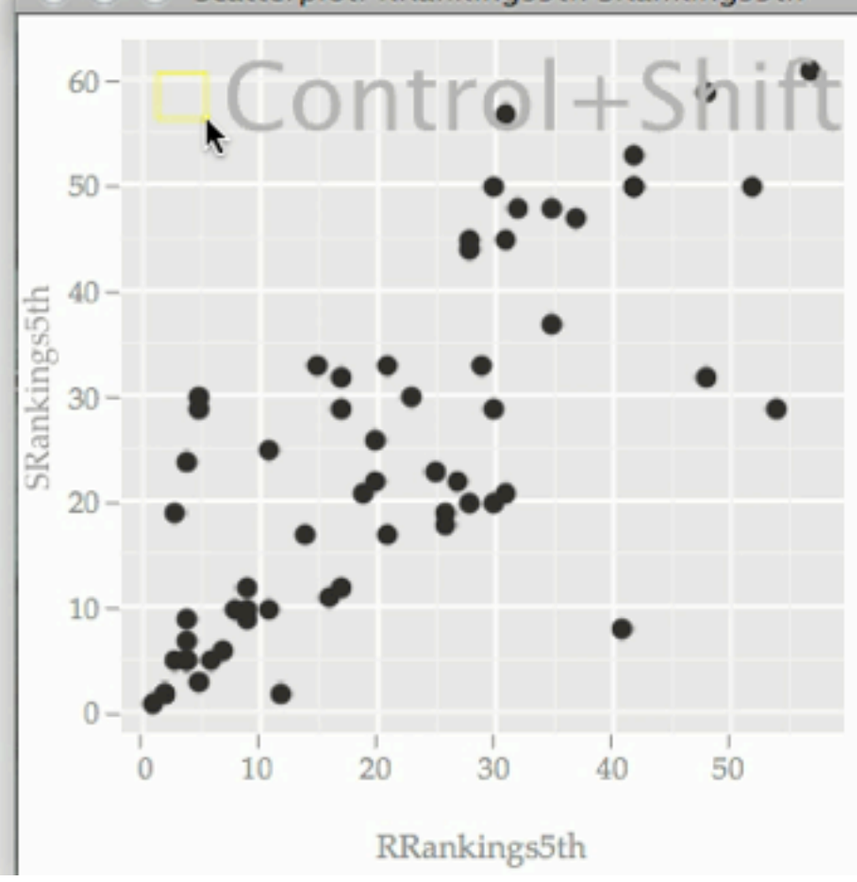
00:00:00 -00:01:27

Par-coords plot: AvNumPubs, AvCitations, MedianTimetoDegr...

Par-coords plot: cl8, cl7, cl6, cl5, cl4, cl...

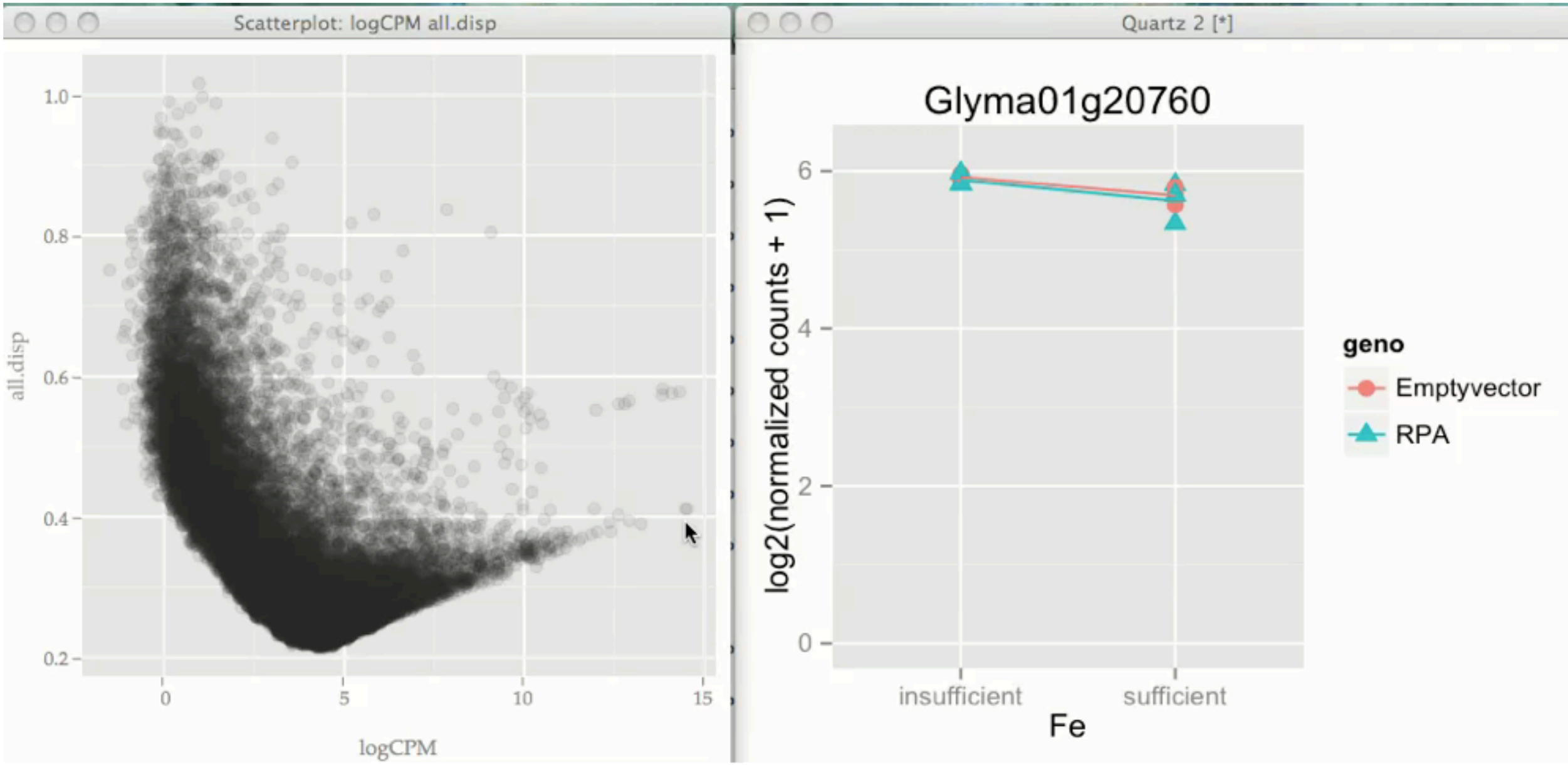


Scatterplot: RRankings5th SRankings5th

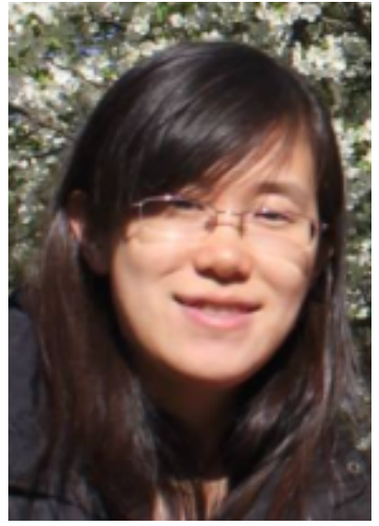


Record Selector: Inst.Prg

BOSTON UNIVERSITY Biostatistics
CARNEGIE MELLON UNIVERSITY Statistics
CASE WESTERN RESERVE UNIVERSITY Stat
COLORADO STATE UNIVERSITY Statistics
COLUMBIA UNIVERSITY IN THE CITY OF N
CORNELL UNIVERSITY Biometry
CORNELL UNIVERSITY Statistics
DUKE UNIVERSITY Statistics and Decision
FLORIDA STATE UNIVERSITY Statistics
GEORGE WASHINGTON UNIVERSITY Statis
HARVARD UNIVERSITY Statistics
IOWA STATE UNIVERSITY Statistics
JOHNS HOPKINS UNIVERSITY Applied Mat
KANSAS STATE UNIVERSITY Statistics
MICHIGAN STATE UNIVERSITY Statistics
NORTH CAROLINA STATE UNIVERSITY Sta
NORTH DAKOTA STATE UNIVERSITY MAI
NORTHWESTERN UNIVERSITY Statistics
OHIO STATE UNIVERSITY MAIN CAMPUS E
OHIO STATE UNIVERSITY MAIN CAMPUS S
OKLAHOMA STATE UNIVERSITY MAIN CA
OREGON STATE UNIVERSITY Statistics
PENN STATE UNIVERSITY Statistics
PURDUE UNIVERSITY MAIN CAMPUS Statis
RICE UNIVERSITY Statistics
RUTGERS THE STATE UNIVERSITY OF NEW
SOUTHERN METHODIST UNIVERSITY Stati
STANFORD UNIVERSITY Statistics
TEMPLE UNIVERSITY Statistics
TEXAS A & M UNIVERSITY Statistics
UNIVERSITY OF CALIFORNIA-BERKELEY St
UNIVERSITY OF CALIFORNIA-DAVIS Statis
UNIVERSITY OF CALIFORNIA-LOS ANGELE
UNIVERSITY OF CALIFORNIA-RIVERSIDE A
UNIVERSITY OF CALIFORNIA-SANTA BARE
UNIVERSITY OF CHICAGO Statistics
UNIVERSITY OF CONNECTICUT Statistics
UNIVERSITY OF FLORIDA Statistics
UNIVERSITY OF GEORGIA Statistics



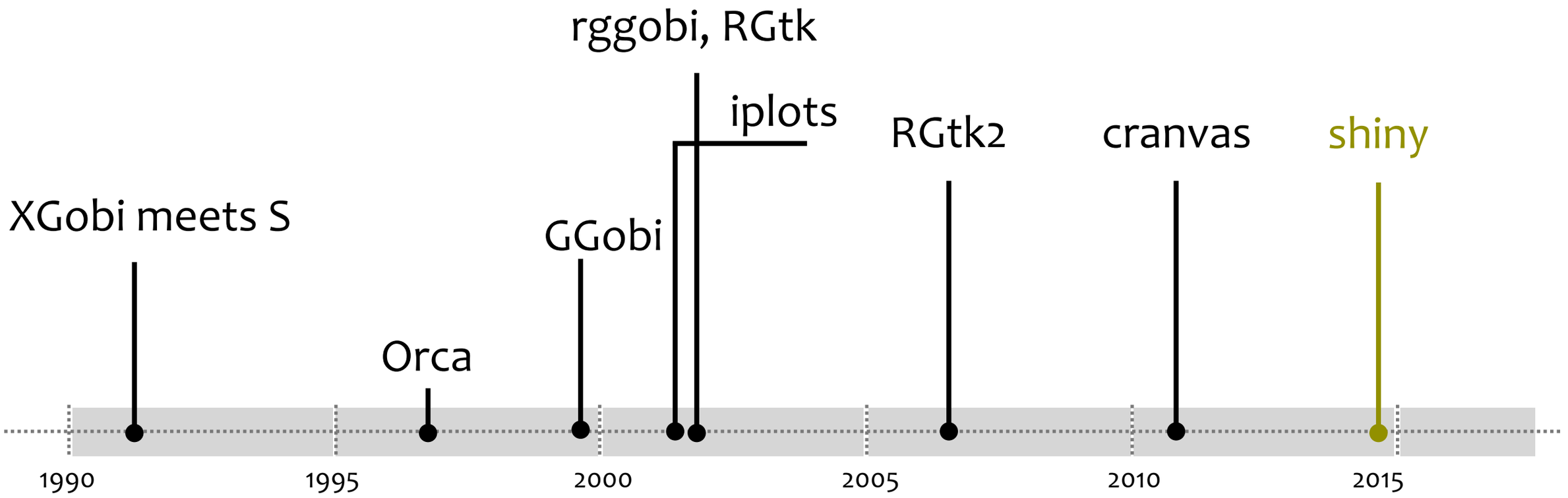

```
qd <- qdata(data, alpha = alpha, size = size)

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)


rggobi, RGtk
iplots
XGobi meets S
GGobi
RGtk2
cranvas
shiny
Orca
1990
1995
2000
2005
2010
2015

S -|

R - - - - - - - - - - - - - - - - - - -

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)
```

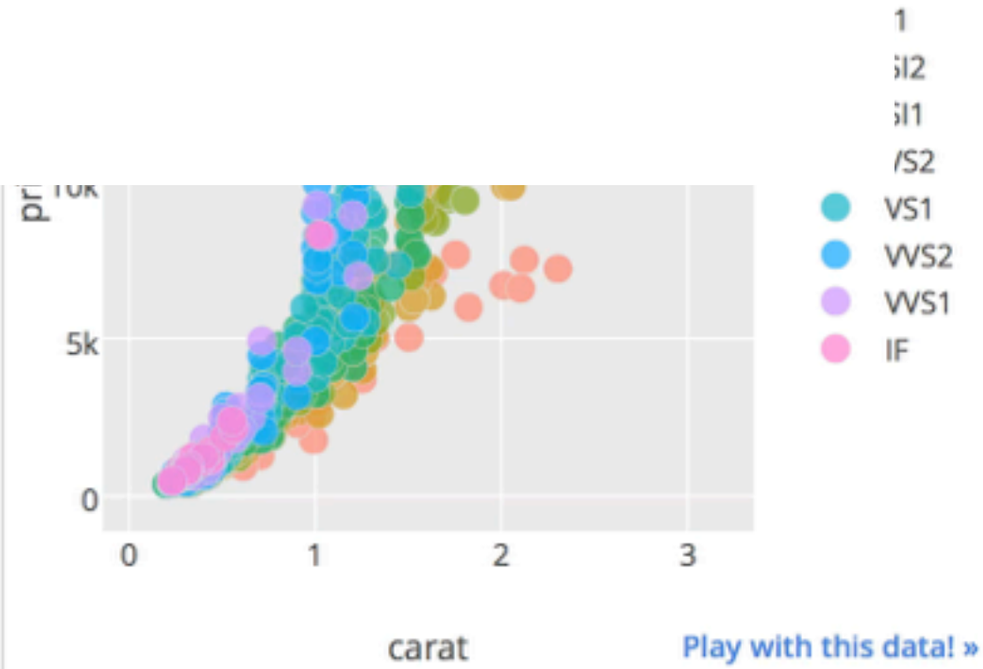
```
library(tourr)
library(animint)
mat <- rescale(as.matrix(flea[1:6]))
tour <- new_tour(mat, grand_tour(), NULL)
tour_dat <- function(step_size) {
  step <- tour(step_size)
  proj <- center(mat %*% step$proj)
  data.frame(x = proj[,1], y = proj[,2], species = flea$species)
}

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

```
libr
dsar
qplo
py <
py$g
```



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/

Copy Number Variation | Barcode Plots | Search CNVs | Field Trials | Genealogy | Methodology

Clear Selections

Number of Generations to Show

1 3 10

1 2 3 4 5 6 7 8 9 10

Type the name of the variety or click on the text box for options

Choose Varieties

Amsoy

Variety Li

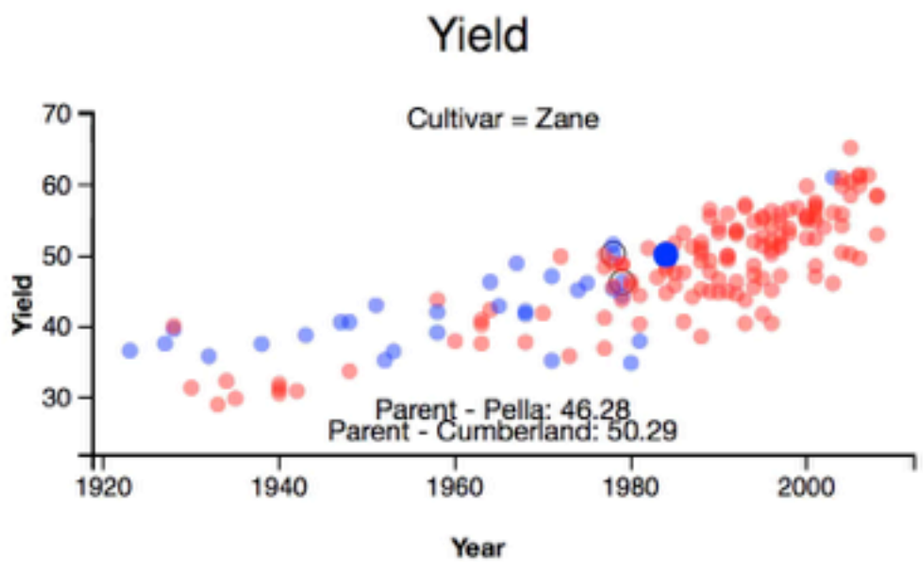
Vanderplas, Graham, Cook Work funded for the USDA examining soybean milestone cultivars

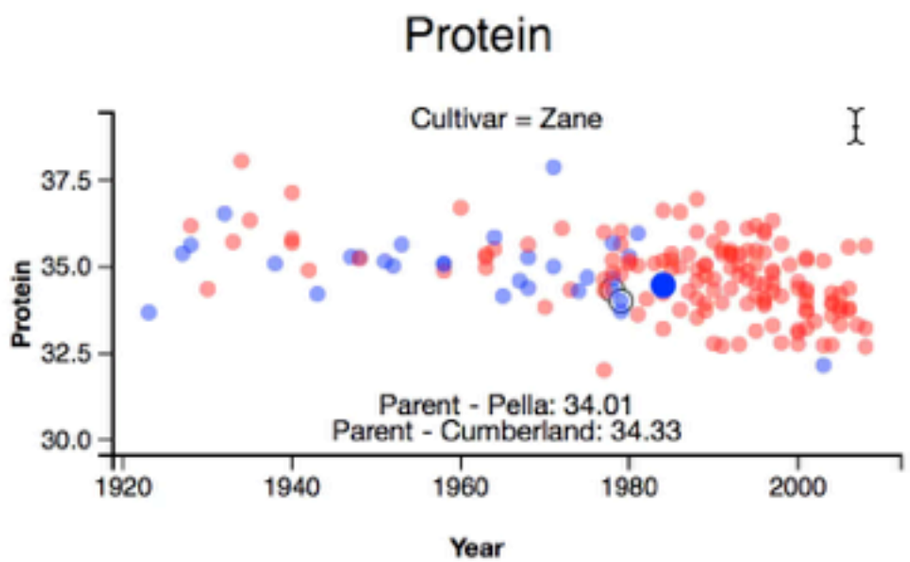
Click on data points in the plot to see field trial data.

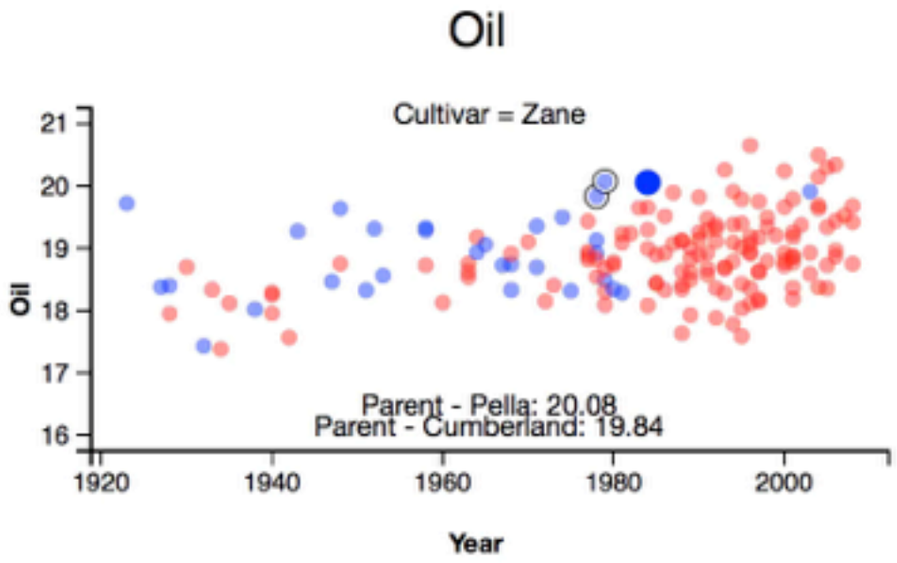
Yield, Protein, and Oil by Year
Show/Hide

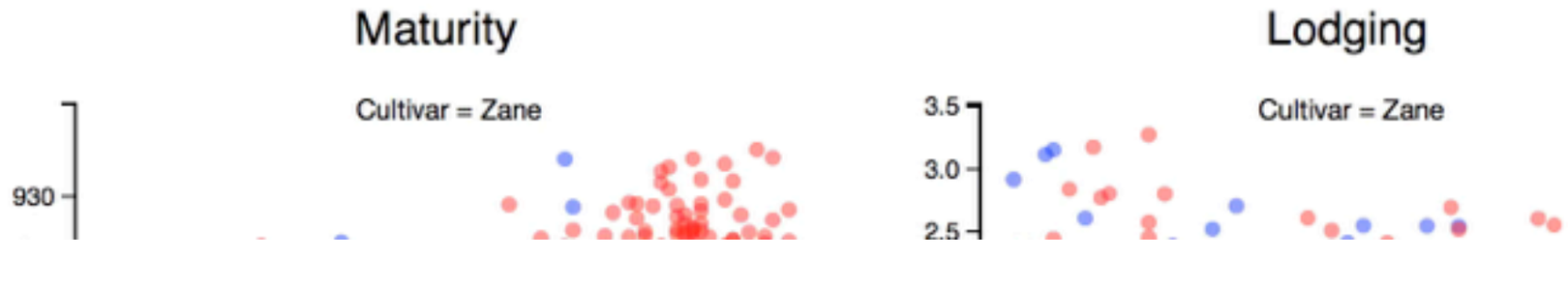
More Field Trial Data by Year
Maturity, Lodging, Seeds Show/Hide

Yield, Protein, Oil Pairwise Plots
Show/Hide









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.