

STAT 6560
Graphical Methods
Spring Semester 2009
Project Work

Rong Xia

Utah State University
Department of Mathematics and Statistics
3900 Old Main Hill
Logan, UT 84322-3900

Andreas Buja

A Modern Pioneer for Interactive Graphics

Brief Background



Figure 1: Andreas Buja (Picture comes from <http://www-stat.wharton.upenn.edu/~buja/>)

- Born in Switzerland
- Currently working as the Liem Sioe Liong/First Pacific Company Professor of Statistics at the Wharton School of the University of Pennsylvania
- Worked at Bell Communications Research, AT&T Bell Labs, Salomon Brothers Inc., University of Washington, Seattle and many other research institutions
- Served as Associate/Managing/Advisory Editor for *Journal of Computational and Graphical Statistics*, Associate Editor for *Journal of American Statistical Association*, *The Annals of Statistics* and several other journals and committees
- Obtained his Ph.D. in Mathematics/Statistics in 1980 from Swiss Federal Institute of Technology (ETH, Zurich)
- Research interests focus on data visualization and data mining
- Co-author of many interactive and dynamic data visualization systems, such as DataViewer, XGobi and GGobi software family

- Supervised several distinguished statisticians in the likes of Dianne Cook, Deborah Donnell, and Catherine Hurley as Ph.D. advisor
- “*He is my hero!*” - (Adele Cutler)

Buja’s Thoughts about Statistical Graphics

- One of the essential areas for statistical computing.
“I already listed several areas that I consider essential for statistical computing, there is also data organization, there is algorithms, there is numerics, most importantly optimization, simulation, and then of course, what is closest to my heart: visualization.” - (Symanzik (2008))
- The most prominent development of Computational Statistics during the 1980s.
“If you go back further, 1980s, well, visualization was big at the time in sort of an esoteric way in that only certain research groups had the capability of doing fancy graphics. It was ... at the time. Later the term visualization was imposed on us.” - (Symanzik (2008))
- Good at representing statistics and have great influences in industry.
“Pictures really, really speak. Graphics, graphics, graphics just teach them! I think most of us cannot teach that sort of thing, but if you can, it will really have the most impact out there in industry.” - (Symanzik (2008))
- Have a bright future in large data research.
“As long as there are computers, yes, of course (laughs). If anything, it becomes more important in certain segments of activities.” - (Symanzik (2008))

Buja’s Expectations on Graduate Education in Statistics

- Graduate students should own a solid foundation in theory, computations skills as well as good data senses.
“There is the traditional curriculum that gives Ph.D. students a solid foundation in theory, but then they also should acquire computational skills, they should become good applied statisticians who have good sense, good data sense.” - (Symanzik (2008))
- There are great job markets for statisticians!!
“All kinds of things.” - (Symanzik (2008))

Buja and GGobi

- **What is GGobi?**

- “GGobi is an open source visualization program for exploring high-dimensional data. It provides highly dynamic and interactive graphics such as tours, as well as familiar graphics such as the scatterplot, barchart and parallel coordinates plots. Plots are interactive and linked with brushing and identification.” - (<http://www.ggobi.org/>)
- “GGobi is a descendant of two earlier programs: XGobi and Dataviewer.” - (Cook & Swayne (2007))
- “GGobi is unique because it offers tours (rotations of data in higher than 3D), complex linking between plots using categorical variables, and the tight connection with R.” - (Cook & Swayne (2007))
- GGobi could be used to draw dotplots and scatterplots, barcharts, spineplots and histograms, parallel coordinate plots, scatterplot matrices.
- GGobi links data points and lines between plots using persistent or transient brushing, and identification.
- GGobi acts as a high-dimensional drawing tool, by adding, moving, and drawing lines between points.
- GGobi can be extended using modular plugins.
- R can be used in sync with GGobi via the package *rggobi*.
- GGobi is free!!

- **Buja’s Contribution to GGobi**

- Most of GGobi was written and designed by Debby Swayne, Di Cook, Duncan Temple Lang, Andreas Buja.
- Buja has worked consistently over 30 years on interactive and dynamic data visualization softwares, i.e. DataViewer (1980s), XGobi (1990s), GGobi (2000s). -(Symanzik (2004))
- Several graduate students also contribute to the development of GGobi. Many of them are Di Cook and Buja’s students.

Some Old Pictures from XGobi

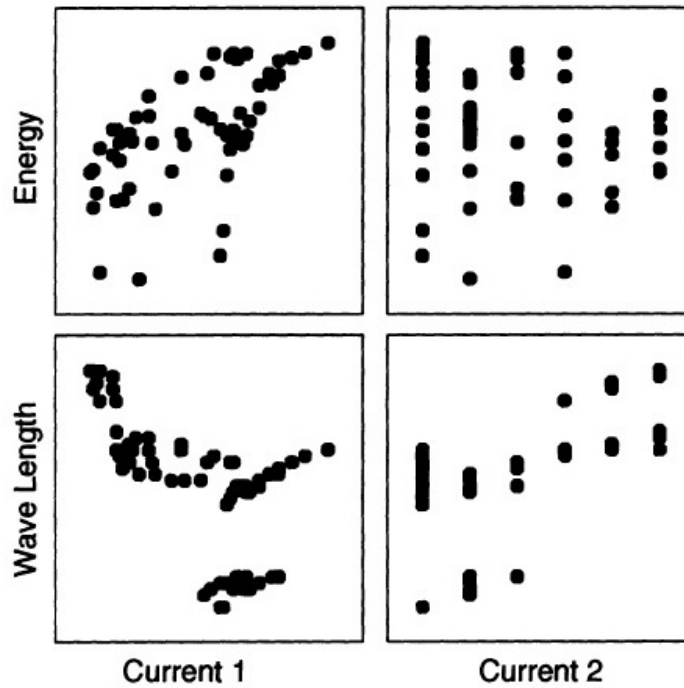


Figure 2: Scatterplot for the laser data Buja et al. (1996)

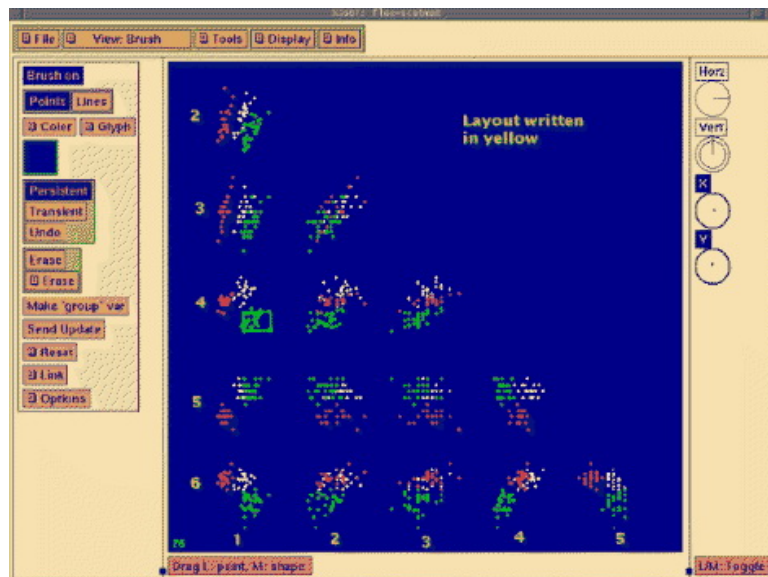


Figure 3: Scatter Plot Matrices for the flea beetle data - (<http://www.public.iastate.edu/~dicook/research/papers/manip/scatmat.html>)

Links to Demo Movies, Data Files and R-codes:

- <http://www.ggobi.org/demos/teaching-with-ggobi.html>
- <http://www.ggobi.org/demos/tour.html>
- http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/xia_rong_project1_laser.txt
- http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/xia_rong_project1_laser.R
- http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/xia_rong_project1_flea.csv
- http://www.math.usu.edu/~symanzik/teaching/2009_stat6560/RDataAndScripts/xia_rong_project1_flea.R

References

- Buja, A., Cook, D. & Swayne, D. F. (1996), ‘Interactive High–Dimensional Data Visualization’, *Journal of Computational and Graphical Statistics* **5**(1), 78–99.
- Cook, D. & Swayne, D. F. (2007), *Interactive and Dynamic Graphics for Data Analysis With R and GGobi*, Springer, New York, NY.
- Symanzik, J. (2004), Interactive and Dynamic Graphics, *in* J. E. Gentle, W. Härdle & Y. Mori, eds, ‘Handbook of Computational Statistics — Concepts and Methods’, Springer, Berlin, Heidelberg, pp. 293–336.
- Symanzik, J. (2008), ‘Interview with Andreas Buja’, *Computational Statistics* **23**(2), 177–184.