Some Reflections on Dynamic Graphics for Data Exploration

Luke Tierney

Department of Statistics & Actuarial Science
University of Iowa

August 3, 2016
Thanks for the opportunity to help congratulate Bill on this well-deserved award.

Outline of this talk:
- background on early interactive graphics for data analysis
- contributions of Bill Cleveland and others at Bell Labs
- some of the work this influenced and motivated
- influence and impact of this work on my efforts
- some thoughts on current state and directions

Slides will be posted on http://www.stat.uiowa.edu/~luke
Early Work on Interactive Graphics

- Many early contributions, and more recent ones, are illustrated in the video library of the Statistical Computing and Statistical Graphics sections.
- Much early work focused on use of rotation to take visualization beyond two dimensions:
  - A very early contribution from Bell Labs is Chang, (1970)
  - An important milestone is PRIM-9 for projecting, rotating, isolating, masking (Fisher, Friedman, and Tukey, 1975)
- Another early development is interactive probability plots (Fowlkes 1971)
  - used knobs to control parameters of distributions and transformations
- These techniques and more are shown in the Bell Labs video Dynamic Displays of Data from 1985.
A Simulation of Interactive Rotation
Brushing Scatterplots

- A new idea from Bell Labs was **brushing linked scatterplots**.
- The initial implementation used a scatterplot matrix.
- The hardware used was the ATT Teletype Model 5620 (BLIT).

The ideas were described in “Brushing scatterplots” (Becker and Cleveland, 1987, Technometrics).

Brushing could be used in two modes:
- persistent
- transient

Persistent brushing allowed creating selections for subsetting or deletion.
Transient highlighting aided in exploring conditional relationships.
Interactive identification/labeling of points was also supported.
Later versions on more advanced hardware supported color brushing.
A Simulation of Scatterplot Brushing
Further Developments

- Availability of more powerful workstations lead to increased development of interactive systems
  - SGI Iris (Becker, Cleveland and Weil)
  - Apollo (Huber)
  - Lisp machines (Buja, McDonald, Oldford, Stuetzle, etc.)
- Personal computers also entered the picture:
  - MacSpin (Donoho, Donoho, and Gasco)
- Becker, Cleveland, and Wilks (1987, Statistical Science) provided an overview.
Further Developments

- Some theoretical work on linking, brushing, selection
  - Furnas and Buja (1984)
  - Wills (1996)
- Extending rotation: the Grand Tour
  - Early work of Asimov, Buja, Hurley, McDonald
  - Xgobi, ggobi software
  - Cook and Swayne (2007)
Rick Becker demonstrated brushing on a BLIT at a JSM in the mid 1980s.
The Macintosh was fairly new at the time.
I was intrigued and wrote simple stand-alone programs for rotation and scatterplot matrix brushing.
Early development for the Mac was an adventure, involving
- using the SUMMACC cross compiler on a VAX
- downloading over phone line on a 300-baud modem
- crash, start over
A clear need: framework for preparing, cleaning, transforming data for visualization.
Embedding a language, like S, seemed natural.
Xlisp source was available, which lead to XLISP-STAT.
Another advantage of language integration: customizable interactions
- Example: adding a smooth to highlighted points in a linked plot
Many commercial systems provide some interactive graphics support. Support within R includes iPlots, rggobi, RGL. Original thinking was interaction for exploration, static views for presentation, reporting. Electronic publishing allows interaction in presentation, reporting. Interactive presentations are very popular in journalism. Most common forms are animation over a variable, often time (e.g. Rosling, hottest year), labeling and drill down. Animation for presentation: shows what you want to show. Rotation, brushing not very common at the moment. Customized interaction is currently not well supported. Some promising efforts include ggvis, animint.
Congratulations and Thank You