

sendplot: an R library to create portable platform independent interactive plots with tool-tip content

Daniel P. Gaile^{1,2,3*}, Lori A. Shepherd^{1,2}, Andrew E. Bruno⁴, Norma J. Nowak^{2,5}, Kenneth F. Manly^{1,2}, and Carl D. Morrison⁶, and Lara E. Sucheston^{1,2,3}

¹Department of Biostatistics, University at Buffalo, Buffalo, NY 14214-3000

²New York State Center of Excellence in Bioinformatics and Life Sciences, Buffalo, NY

³Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY

⁴Center for Computational Research, University at Buffalo, Buffalo, NY

⁵Department of Biochemistry, University at Buffalo, Buffalo, NY

⁶Department of Pathology, Roswell Park Cancer Institute, Buffalo, NY

Received on 02/14/08

ABSTRACT

Motivation:

The analysis of high dimensional data, particularly those generated by advanced biotechnologies such as expression and CGH array platforms, is not guaranteed to provide small dimensional results. Each component of a set of 'results' may have assay and sample specific attributes which are of potential interest to the collaborative research team investigating a scientific question. In addition, because current science necessitates multi-disciplinary and multi-institutional research teams, it is imperative results be communicated clearly and efficiently via a platform independent vehicle that allows for a custom viewing experience.

Results: We provide an R software library, sendplot, which allows users to generate interactive plots with tool-tip content. The software creates a pair of files: a Portable Network Graphics (PNG) file which is a bitmap image and an HTML file which contains embedded Javascript code for dynamically generating tool-tips. When the HTML file is opened within a supported browser, the PNG image is displayed and the user is able to mouse over and view tool-tip windows for user specified image locations. The information that appears in the tool-tip windows is user specified and easy to configure. The functions in sendplot provide a convenient means for R users to generate such output for both scatter-plot and heatmap graphical displays.

Availability: The sendplot library is available at:

<http://cran.r-project.org/src/contrib/Descriptions/sendplot.html>

<http://sphhp.buffalo.edu/biostat/research/software/sendplot/>

Supplementary Information: Example output and a detailed vignette can be found at:

<http://sphhp.buffalo.edu/biostat/research/software/sendplot/>

Contact: dpgaile@buffalo.edu

1 BACKGROUND

The sendplot R library provides interactive plots with tool-tip content for two generic graphical representations of data: 1) scatter-plots and 2) heatmaps (e.g., Eisen *et al.*, 1998). While the software has many generic applications, it is particularly useful for the analysis and visualization of array comparative genomic hybridization (aCGH) data (Albertson and Pinkel, 2003). For example, the software provides interactive heatmap representations of aCGH data matrices with matrix columns corresponding to assayed samples, matrix rows corresponding to particular array CGH assays, matrix entries corresponding to observed log₂ Tumor/Control values, and tool-tip data displays that are customizable. Surveys of software for the analysis of aCGH and microarray data in general are provided in Chari *et al.*, 2006, Hibbs *et al.*, 2005, and Prasad and Ahson, 2006. To our knowledge, no other software packages currently provide the functionality of sendplot.

2 DESCRIPTION

2.1 Graphics Generated by R

The graphical output used by sendplot is generated by R; it is outputted either directly to PNG format or it is generated in postscript and then converted to PNG (recommended). Sendplot functions allow for the graphics device to be partitioned via the 'layout' mechanism (Murrell, 1999) where it is assumed that the first figure (i.e., figure designated by a '1' in the layout design matrix) is the interactive plot. This approach allows the construction of interactive plots which contain any number of decorative plots of any level of complexity. The interactive plot is limited to being either a scatter-plot or heatmap image.

2.2 Interactive Tooltips

The sendplot functions generate an HTML file with embedded Javascript code to create an interactive web page for viewing the PNG graphic outputted from R. Using meta-data provided by R about the PNG image, a pixel coordinate mapping (image map) is

*to whom correspondence should be addressed

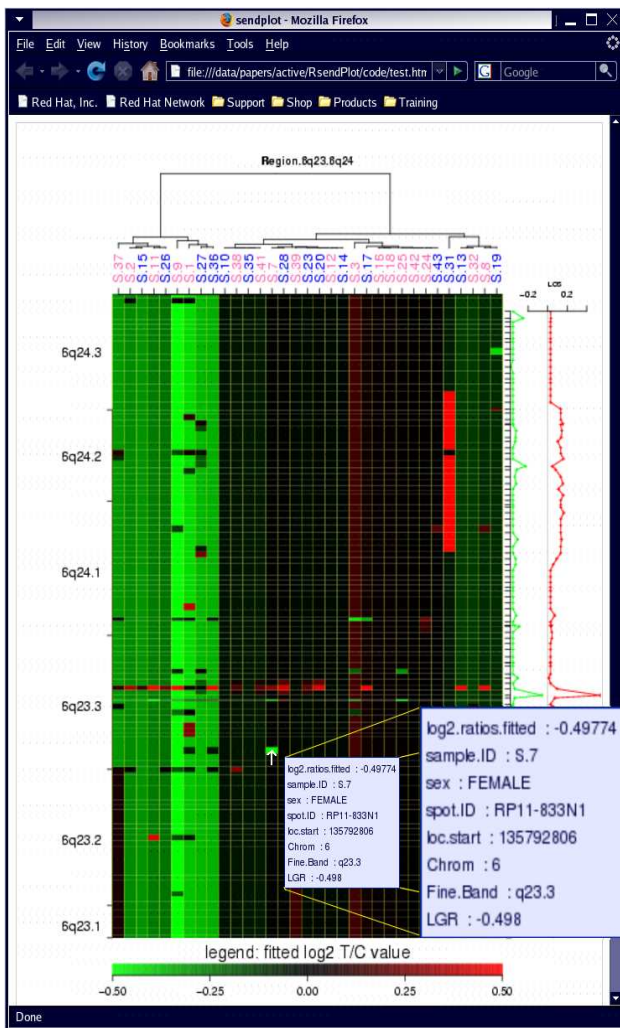


Fig. 1. A sendplot interactive heatmap for array comparative genomic hybridization (aCGH) data. The rows of the data matrix correspond to particular BAC aCGH assays spanning 6q23 and 6q24. The columns of the data matrix correspond to particular clinical samples which were assayed. The entries in the data matrix are color coded according to the fitted log₂ Tumor/Control value for the particular BAC assay and sample. Decorative (i.e., non-interactive) plots include a dendrogram displaying the manner in which the samples were clustered (top), a plot of the linear combination of order statistics across samples (right), and a legend (bottom). The cursor has been placed over the data matrix element corresponding to the sample S.7 and BAC assay RP11-833N1. The resulting tooltip (and magnification) is displayed. The contents of the tooltip are user specified and highly customizable.

generated using special HTML markup which links various areas of the PNG image to customized Javascript code for rendering tooltip windows. The Javascript code detects when the users mouse enters a specified area of the PNG image and displays a tooltip window containing the variable names and values defined for that specific location. The `wz.tooltip.js` Javascript library (Zorn 2007) is embedded within the HTML page and used for rendering the tool-tips.

2.3 Creating the Image Map

The sendplot functions are flexible with respect to the dimension of the graphical output and provide convenient mechanisms to create the pixel mappings. For example, coordinate markers can be included in the PNG output so that the pixels which bound the active region of the plot can be easily identified using `mspaint` or `kolourpaint`. Once these coordinates are identified, the sendplot functions map the plotted coordinates to pixel coordinates.

Tool-tip content is categorized as either x-specific, y-specific, or point specific (e.g., xy-specific) and convenient mechanisms exist to attach user-specified attributes to the interactive plot. For example, in Figure 1 the tool-tip contains x-specific information (i.e., `spot.ID`, `loc.start`, `Chrom`, and `Fine.Band`), y-specific information (i.e., `sample.ID` and `sex`), and point specific information (i.e., `log2.ratios.fitted` and `LGR`).

2.4 heatmap.send: a Convenient Wrapper Function

The sendplot library includes the convenient wrapper function `heatmap.send`. This allows users to create an interactive heatmap using a function that utilizes the same set of arguments as the standard heatmap function in R.

3 CONCLUSIONS

Although originally designed for applications related to the analysis and visualization of expression array and array CGH data, the sendplot functions are generally applicable to any dataset which can be displayed using scatter-plot of heatmap (i.e., 'image') functions in R. The sendplot library allows for quick and convenient generation of customizable interactive plots with tool-tip content.

FUNDING

This work was supported by grants from the National Cancer Institute (CA 27469, to the GOG Administrative Office; CA 37517, to the GOG Statistical Office; CA 113882, to Norma J. Nowak; and CA 16056-27-RPCI, Cancer Center Support Grant to Roswell Park Cancer Institute).

REFERENCES

- Albertson, Donna G, Pinkel, Daniel (2003). Genomic microarrays in human genetic disease and cancer. *Genomic microarrays in human genetic disease and cancer*, **12** Spec No 2: 142-152.
- Chari, R., Lockwood, W.W., Lam, W.L. (2006). Computational Methods for the Analysis of Array Comparative Genomic Hybridization. *Cancer Informatics*, **2**: 48-58.
- Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, **95** (25): 14863-14868.
- Hibbs, M.A., Dirksen, N.C., Li, K., Troyanskaya, O.G. (2005). Visualization methods for statistical analysis of microarray clusters. *BMC Bioinformatics*, **6**: 115.
- Murrell, P.R. (1999). Layouts: A mechanism for arranging plots on a page. *Journal of Computational and Graphical Statistics*, **8**: 121-134.
- Prasad, T.V., Ahson, S.I. (2006). Visualization of microarray gene expression data. *Bioinformatics*, **4**: 141-145.
- Zorn, W. (2007). JavaScript, DHTML Tooltips JavaScript Cross Browser Library. <http://www.walterzorn.com/tooltip/tooltip.e.htm>.