A Minimal Demo of ggplyr/glyphmaps/modelglyphs

Garrett Grolemund

May 26, 2012

This document demonstrates the basic features of the `ggplyr` package (name may change). `ggplyr` was written to create plots like the ones below. Each plot was made with `ggplyr`. The first two plots recreate plots from Wickham (2011). The last two recreate plots from Wickham et al. (Submitted).

1 Glyphs

The common feature of these graphs is that they are built around glyphs. Glyphs are geometric objects (i.e, geoms) designed to display information within each geom. In other words, a glyph can display information even if it ids drawn by itself, without references to an external coordinate system. In reality, all geoms are a type of glyph, but the term glyph is usually reserved for complicated geoms, such as those that contain their own internal coordinate systems. The star glyphs in Figure **??** illustrate how glyphs can contain an internal (minor) coordinate system and can still be plotted in an external (major) coordinate system.

Glyphs reveal a heirarchical structure to graphics: every plot is a collection of geoms, each of which can be thought of as its own self contained plot. Sometimes these subplots are not very interesting, as in the subplot created by a single point geom. Sometimes, they are quite complex, as in the star glyph of Figure **??**.

Graphs inherit this heirarchical structure from the data they describe. Data is produced through an iterative process of collecting observations, grouping observations and summarizing groups of observations to create more compact, information dense sets of data. Humans innately perform this process when collecting data. It is a cognitive pattern of the human brain which I will write about in the second cognitive chapter of my thesis.

Glyph maps simultaneously expose data from multiple levels of this process. As a set of geoms, the glyphs reveal relationships between data points in the higher level, compact data set. As individual plots, each glyph retains information about the data points in the lower level group of data that it summarises. This dual display makes glyph maps particularly useful for certain data analysis tasks. It also provides two different approaches to constructing glyph maps.

Glyph maps can be built from the top down by treating each glyph as a geom within the plot of interest. `ggplyr` provides new geoms $\texttt{geom}_s tar, \texttt{geom}_r adar, \texttt{geom}_a art, and \texttt{geom}_p lyrwhichallowausertoquicklyincorporategeomsintoanexis$

`ggplyr's top down methods help users fit glyphs into existing methods of visualization. These methods provide new geoms (often based on new grobs) that can be used alongside existing geoms in ggplot2.`

`Top down methods are difficult to implement in ggplot2 for two reasons`

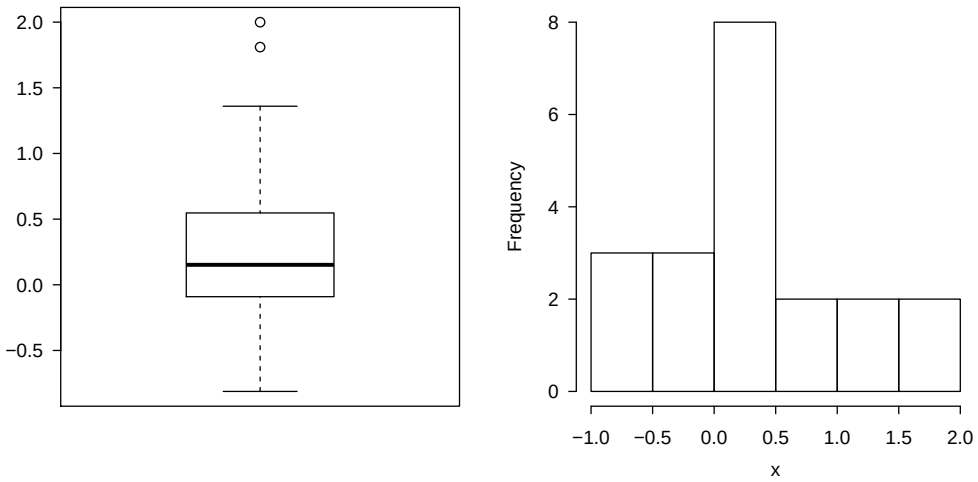
`First, ggplot2 calculates aesthetics on the entire data set at once. However, glyphs contain aesthetics that must be keyed to subsets of the data.`

`Second, the final width and height of individual glyphs often depends on non-position aesthetics, such as angle and length. In the ggplot2 pipeline, these aesthetics are scaled right before the plot ranges are trained. As a result, the final widths of glyphs must be computed at draw time and frequently place parts of the glyph outside the plot window.`

```
set.seed(1121)
(x <- rnorm(20))

##  [1]  0.14496  0.43832  0.15319  1.08494  1.99954 -0.81188  0.16027  0.58589  0.36009
## [10] -0.02531  0.15088  0.11008  1.35968 -0.32699 -0.71638  1.80977  0.50840 -0.52746
## [19]  0.13272 -0.15594

mean(x)

## [1] 0.3217

var(x)

## [1] 0.5715
```

The first element of x is 0.145. Boring boxplots and histograms recorded by the PDF device:

```
## two plots side by side (option fig.show='hold')
par(mar = c(4, 4, 0.1, 0.1), cex.lab = 0.95, cex.axis = 0.9, mgp = c(2, 0.7, 0),
    tcl = -0.3, las = 1)
boxplot(x)
hist(x, main = "")
```



Do the above chunks work? You should be able to compile the TEX document and get a PDF file like this one: https://github.com/downloads/yihui/knitr/knitr-minimal.pdf. The Rnw source of this document is at https://github.com/yihui/knitr/blob/master/inst/examples/knitr-minimal.Rnw.

References

Hadley Wickham. The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1âĂŞ29, 2011. URL http://www.jstatsoft.org/v40/i01/.

Hadley Wickham, Heike Hofmann, Charlotte Wickham, and Diane Cook. Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics*, Submitted.