# blogR: R for blogs

Shane M. Conway

December 13, 2009

**Abstract**

`blogR` is an R package to provide a standardized framework for online reproducible research through blogs. It aims to both simplify the blogging process for R users, while simultaneously improving the quality. The primary input format for `blogR` is a Sweave document, which should ultimately allow R researchers to publish their papers in multiple different formats and reuse existing writing.

## 1 Introduction

The statistical language `R` [1] has become the de-facto standard for statistical computing. `R` currently offers several packages for report writing, including `Sweave` [3] and `brew` [2]. Many R users also post reports using blogs.

According to wikipedia:

> The term "weblog" was coined by Jorn Barger on 17 December 1997. The short form, "blog," was coined by Peter Merholz, who jokingly broke the word weblog into the phrase we blog in the sidebar of his blog Peterme.com in April or May 1999.

There are many different blog platforms available. Of the top 100 blogs, Wordpress and Moveable Type combined cover 60%.

The goal of `blogR` is to make blogging with R easier and better. By lowering the transaction costs in blogging, `blogR` should both increase the number and quality of R blogs. At present, most R bloggers are either developing their own workflow or else are forced to follow several steps for each post: write the post in the blogging engine, create all R graphics, upload each graphic individually, and incorporate all media into the post. This can be a very time consuming process, and hence provides little incentive for people to make regular updates.

Another goal for the `blogR` package is reusability. By using `Sweave` as the primary document, the hope is to allow researchers to take existing latex papers and post them online as hypertext documents.

`blogR` currently supports posting R results to Wordpress, with plans to ultimately support other blog engines including Google's Blogger.

### 1.1 R Blogs

There are many active blogs that currently employ R for data analysis.

There are also two R blog aggregators (i.e. which pull from a number of different blogs: Planet R (http://planetr.stderr.org/) and R bloggers (http://www.r-bloggers.com/).

| | Blog Name | URL |
|---|---|---|
| 1 | Antrhospace | http://www.stanford.edu/~cengel/cgi-bin/anthrospace |
| 2 | Biogistic Reflections | http://blogisticreflections.wordpress.com |
| 3 | Cengel's Favorite Links from Diigo | http://www.diigo.com/user/Cengel |
| 4 | Cerebral Mastications | http://www.cerebralmastication.com/ |
| 5 | Dataninja | http://dataninja.wordpress.com |
| 6 | FOSS Trading | http://blog.fosstrading.com/ |
| 7 | Getting Genetics Done | http://gettinggeneticsdone.blogspot.com/ |
| 8 | Gregor Gorjanc | http://ggorjan.blogspot.com/ |
| 9 | Jeromy Anglim's Blog | http://jeromyanglim.blogspot.com/ |
| 10 | John Myles White: Die Sudelbücher | http://www.johnmyleswhite.com/ |
| 11 | LearnR | http://learnr.wordpress.com/ |
| 12 | neilkodner.com | http://www.neilkodner.com/ |
| 13 | One R Tip A Day | http://onertipaday.blogspot.com/ |
| 14 | Please Scoop Me! | http://pleasescoopme.com/ |
| 15 | Quantivity | http://quantivity.wordpress.com/ |
| 16 | R-statistics blog | http://www.r-statistics.com/ |
| 17 | Realizations in Biostatistics | http://realizationsinbiostatistics.blogspot.com/ |
| 18 | Romain Francois, Professional R Enthusiast | http://romainfrancois.blog.free.fr/ |
| 19 | simon jackman's blog | http://jackman.stanford.edu/blog |
| 20 | The Endeavour | http://www.johndcook.com/blog |
| 21 | REvolutions | http://blog.revolution-computing.com/ |
| 22 | Thinking inside the box | http://dirk.eddelbuettel.com/blog |
| 23 | Statistical Modeling, Causal Inference, and Social Science | http://www.stat.columbia.edu/~cook/movabletype/mlm/ |
| 24 | 'R' you ready? | http://ryouready.wordpress.com/ |
| 25 | Stats raving mad | http://statsravingmad.wordpress.com/ |
| 26 | Wine PhD | http://www.greghirson.com/blog |

## 2 Literate Programming and Online Reproducible Research

One of the major uses of R, especially given it's integration with LaTeX, is the ultimate goal of "reproducible research". This subject has been covered most extensively by ... and ... Ultimately, a research paper should be completely reproducible including both the software and data.

Donald E. Knuth: "Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do." http://www.literateprogramming.com/knuthweb.pdf

"One line of processing is called weaving the web; it produces a document that describes the program clearly and that facilitates program maintenance. The other line of processing is called tangling the web; it produces a machine-executable program."

## 2.1 LaTeX

http://stackoverflow.com/questions/270121/best-latex-editor-for-windows
Mention miktex

## 2.2 Sweave

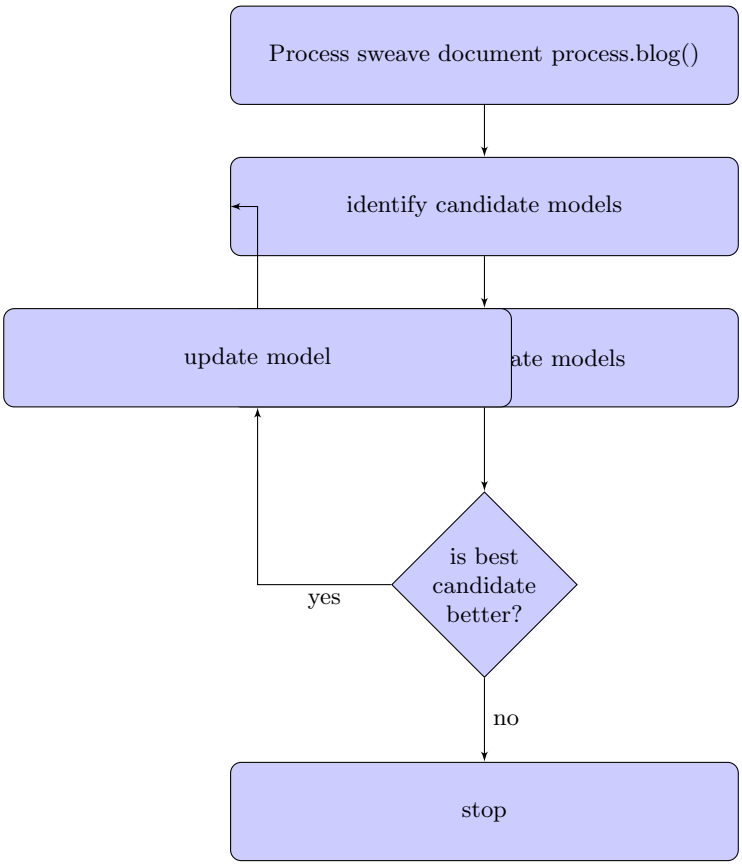
Sweave is a function in the statistical programming language R that enables integration of R code into LaTeX or LyX documents. The purpose is "to create dynamic reports, which can be updated automatically if data or analysis change".[1] The data analysis is performed at the moment of writing the report, or more exactly, at the moment of compiling the Sweave code with Sweave (i.e., essentially with R) and subsequently with LaTeX. This can facilitate the creation of up-to-date reports for the author. Because the Sweave files together with any external R files that might be sourced from them and the data files contain all the information necessary to trace back all steps of the data analyses, Sweave also has the potential to make research more transparent and reproducible to others[citation needed]. However, this is only the case to the extent that the author makes the data and the R and Sweave code available. If the author only publishes the resulting PDF document or printed versions thereof, a report created using Sweave is no more transparent or reproducible than the same report created with other statistical and text preparation software.

cran.r-project.org/web/packages/pgfSweave/ http://cran.r-project.org/web/packages/cacheSweave/ http://cran.r-project.org/web/packages/tikzDevice/ cran.r-project.org/web/packages/odfWeave http://cran.r-project.org/contrib/extra/lyx/ http://www.lyx.org/ http://texlipse.sourceforge.net/

# 3 Using the `blogR` package

blogR follows a simple workflow:

; ;




The blogR package uses XML-RPC to post blog entries (and associated media) onto supported blog sites.

```
> 1 + 1
> 1 + pi
> sin(pi/2)
```

## 3.1 Configuration

`blogR` can use a number of environmental varaibles to simplify its ongoing usage (see subsequent section on "simplified workflow").

configure.blogR

## 3.2 Creating new blog posts

In–line `Sweave` expressions are created using `\Sexpr`. Code chunks can be created with the noweb convention (using `<<>>=`). The LaTeX code chunk syntax is not currently supported.

The post title should be contained in a comment at the top of the ascii file in the following format: `% title=`

An optional parameter will decide whether to update the users status on twitter using the twitteR package, includes a link to the new blog entry.

| | Variable | Definition |
|---|---|---|
| 1 | blogR.FILE | The file name for the blog post (a valid Sweave file with extension .Rnw) |
| 2 | blogR.PATH | The path to the location of the blog post files. |
| 3 | blogR.URL | The URL of the blog site. |
| 4 | blogR.USERNAME | The username with write access for posting to the blog. |
| 5 | blogR.PASSWORD | (optional) The password for the supplied username. |
| 6 | blogR.ARCHIVE | Logical value which determines whether blog posts are archived after being uploaded. |
| 7 | blogR.PUBLISH | Logical value which determines whether blog posts are immediately published or kept as a draft. |
| 8 | blogR.FORMAT | Format of the blog file: can be 'ascii', 'latex', or 'html' (ascii is currently the only supported format). |

Table 1: blogR configuration variables

### 3.3 Adding media to a blog

```
> data(iris)
> summary(iris)
```

```
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
 setosa    :50
 versicolor:50
 virginica :50
```

## 4 A simplified workflow

`blogR` can be used in an adhoc fashion as described above (with the new.post and new.media functions), or it can be used with a simplified workflow for repetitive blogging.

Repetitive blogging is done by setting up global environment variables using the configure.blogR function.

```
> library(graphics)
> boxplot(Sepal.Length ~ Species, data = iris)
```

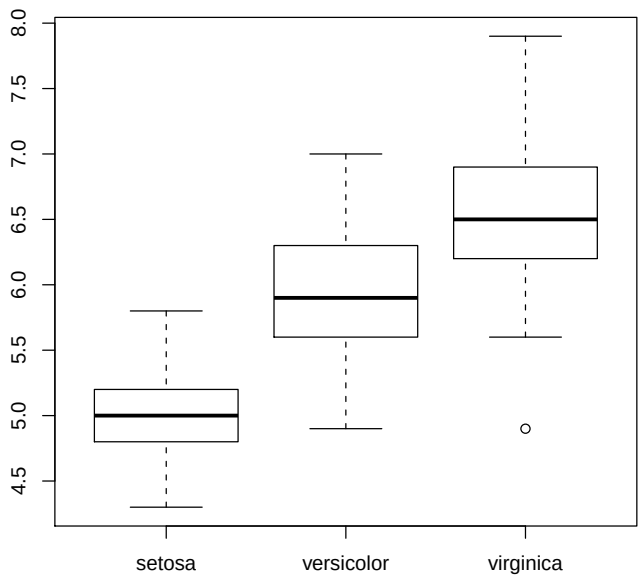



Figure 1: Boxplot of sepal length grouped by species.

# 5 Conclusion

The goal of the `blogR` package is to provide another means for reports from R, and particularly to improve further on R's goals for "reproducible research".

# References


[1] R Development Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

[2] Diethelm Wuertz, many others and see the SOURCE file (2007): *Rmetrics: Rmetrics - Financial Engineering and Computational Finance.* R package version 260.72. http://www.rmetrics.org

[3] Friedrich Leisch, *Dynamic generation of statistical reports using literate data analysis.* Proceedings in Computational Statistics, pages 575-580. Physika Verlag, Heidelberg, Germany, (2002). ISBN 3-7908-1517-9, URL http://www.stat.uni-muenchen.de/ leisch/Sweave