Easier migration from Sweave to R Markdown using the texor package

Yinxiang Huang Abhishek Ulayil Heather Turner Dianne Cook

Keywords: Reproducible Research, LaTeX, R Markdown

Abstract

The **texor** package was developed to migrate legacy R Journal articles written in LaTeX to R Markdown and to generate an HTML version of these documents. The converted articles are now live and visible on The R Journal website. Building on this work, we have extended the package to convert Sweave files into R Markdown while retaining the executable code chunks. This makes it much easier for package authors to convert vignettes from Sweave to R Markdown, which they may wish to do for reasons of easier maintenance or due to the advantages of HTML over PDF format: superior accessibility, responsive layout on a range of devices or HTML-specific features such as dynamic graphics and tables.

Background

LaTeX's powerful typographic capabilities produce elegant PDF documents but as more documentation is viewed through a web browser rather than in print, there is a growing need for elegant HTML formats. For the R community, the first literate programming LaTeX articles were produced using Sweave to have code dynamically executed when documents were typeset. Then the community adopted a markdown-based, literate programming format for rendering to HTML or PDF. The **texor** package, was developed to convert legacy R Journal LaTeX articles to RMarkdown to deliver legacy articles in the new HTML and PDF formats. We have extended the package to convert Sweave articles, such as those used in package vignettes to RMarkdown. A single function does the conversion, and supports several different options for users. The goal for this work is to facilitate the transition of CRAN's legacy documents to HTML format to support the modernization of CRAN.

Sweave article specific features

In the pursuit of converting each and every aspect of the LaTeX based Sweave files we have implemented Lua filters for Pandoc and R function extensions to cover pre-processing of the articles.

1. Code blocks : We have written a Custom Pandoc Reader in Lua using LPEG to read and extract code blocks effectively. Next, we transform each code block into an R Markdown code block and merge it with the remainder of the Sweave article converted into markdown. Static code chunks are also retained.

2. Equations : To maintain parity with the standard LaTeX equation numbering system, we have developed a Lua filter to automatically label equations, which would be visible after compilation of R Markdown file to HTML.

3. Figures : We have an optional figure handling scheme, which can transform LaTeX figures into R Markdown code chunks to include figures using `knitr`.

4. Tables : Similar to figures, table data can also be transformed into a csv data file, which is included in the R Markdown article using a `kable()` in a new code chunk. For more complicated LaTeX tables, we can fall back to simple R Markdown tables.

5. Meta data : Sweave file is parsed for meta data like title, author info and abstract, which is added to the yaml header of the generated R Markdown file.

6. Bibliography : Our current package supports BibTeX files for including a bibliography. In case we have embedded bibliographic entries inside the Sweave article itself, then we can convert it into BibTeX using a sister package **rebib** developed specifically for this.

7. Output formats : We aim to bring multiple output formats for the generated R Markdown file, this should allow users to customize the final web document they intend to create.

Key Features of the package

1. Performance : The package is fast and efficient to convert the Sweave/LaTeX articles into R Markdown.

2. Easy to use : We have designed the package to be easy to use, requiring just a single function call for most use cases. There are more options available within the functions provided by the package for customizing the ouput.

3. Automation : The package is designed around utilizing a single function call to orchestrate the whole conversion process. This enables the package to be used in automation on a collection of articles.

Impact

1. Accessibility : One of our motivations for converting existing LaTeX/Sweave articles into R Markdown and subsequently HTML, is the better support for screen readers. This will have a great impact on visually impaired R users, to be able to use screen readers on a wider set of articles about R.

2. Encourage adoption of R Markdown : Our package will also serve as a transition for authors accustomed to LaTeX/Sweave formats and provide an easy way to get their existing articles into R Markdown and provide a pathway to encourage them to adopt R Markdown subsequently.

3. Ease of viewing : The web articles are responsive and will generally work well with mobile devices. This should improve the experience of readers who use mobile devices to read the articles/documents.