REPRODUCING R

Scripts, Documents, and Packages

June 22, 2013

Jeff Allen, Dallas R Users Group

About Me

- MS Computer Science, SMU
- By day…
  - Computational Biologist at UT Southwestern
    - Use R to analyze biomedical data
    - Develop Java-based web applications
- By night…
  - Freelance consultant as Trestle Technology
    - Training
    - Web development
    - Data analysis
    - IT consulting

TRESTLE TECHNOLOGY

Overview

- Motivation
- Scripts
- Reproducible Documents
- R Packages

Motivation

- R coding will get lonely
- Eventually want to share with others
- How best to wrap up your R code for different audiences?
  - Non-technical managers?
  - Technical peers?
  - The public at large?

R Scripts

source()

R Scripts

- First step towards reproducibility
- Save your analysis in a file
  - Can be re-executed
- Call `source("filename.R")` to run a file

Uses

- Poor man's reproducibility
  - Certainly better than just interactive use
  - Will have *some* reference later on
- Only accessible to R programmers
- A lot of opportunities to do this poorly
  - Dependency on local files
  - Assume data has already been loaded
  - Assume packages have already been loaded
  - No documentation standards

Reproducible Documents

sweave, knitR

History

- Sweave came first
  - Embed R code in LaTeX
  - Execute R and embed input/output into LaTeX doc
  - Convert to PDF, etc.
- knitr came next
  - More flexible engine
  - Embed R code into whatever you want (slides, HTML, LaTeX, …)
  - Various improvements like caching

LaTeX

- "A document markup language and document preparation system."[1]

```
\documentclass[12pt]{article}
\usepackage{amsmath}
\title{\LaTeX}
\date{}
\begin{document}
  \maketitle
  \LaTeX{} is a document preparation system for the \TeX{}
  typesetting program. It offers programmable desktop publishing
  features and extensive facilities for automating most aspects of
  typesetting and desktop publishing, including numbering and
  cross-referencing, tables and figures, page layout, bibliographies,
  and much more. \LaTeX{} was originally written in 1984 by Leslie
  Lamport and has become the dominant method for using \TeX; few
  people write in plain \TeX{} anymore. The current version is
  \LaTeXe.

  % This is a comment; it will not be shown in the final output.
  % The following shows a little of the typesetting power of LaTeX:
  \begin{align}
    E &= mc^2                                    \\
    m &= \frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}}
  \end{align}
\end{document}
```

[1]LaTeX. (2013, June 14). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:35, June 19, 2013, from http://en.wikipedia.org/w/index.php?title=LaTeX&oldid=559858385

LaTeX

- "A document markup language and document preparation system."[1]

LaTeX

LaTeX is a document preparation system for the TeX typesetting program. It offers programmable desktop publishing features and extensive facilities for automating most aspects of typesetting and desktop publishing, including numbering and cross-referencing, tables and figures, page layout, bibliographies, and much more. LaTeX was originally written in 1984 by Leslie Lamport and has become the dominant method for using TeX; few people write in plain TeX anymore. The current version is LaTeX 2_ε.

$$E = mc^2 \tag{1}$$

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2}$$

[1]LaTeX. (2013, June 14). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:35, June 19, 2013, from http://en.wikipedia.org/w/index.php?title=LaTeX&oldid=559858385

Markdown

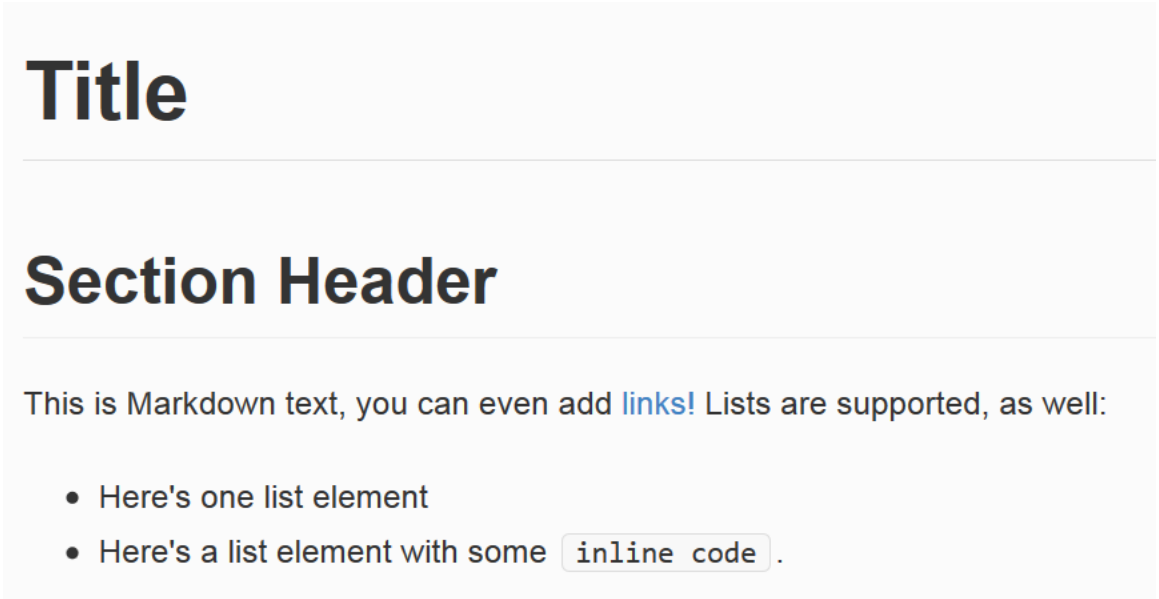
- Markup language to render HTML
- A variety of formats and subtle variants

```
# Title

## Section Header

This is Markdown text, you can even add [links!]
(http://trestletech.com) Lists are supported, as well:

 - Here's one list element
 - Here's a list element with some `inline code`.
```

TRESTLE TECHNOLOGY

Markdown

- Markup language to render HTML
- A variety of formats and subtle variants



Title

Section Header

This is Markdown text, you can even add links! Lists are supported, as well:

- Here's one list element
- Here's a list element with some `inline code`.

R + Markdown

.RMD

````
# Sample Markdown

This is an R Markdown document!

```{r}
x <- 1:10
x
```

You can also embed plots, for example:

```{r fig.width=7, fig.height=6}
plot(x, x^2)
```
````

R + Markdown

.MD

````
# Sample Markdown

This is an R Markdown document!

```r
x <- 1:10
x
```

```
## [1] 1 2 3 4 5 6 7 8 9 10
```

You can also embed plots, for example:

```r
plot(x, x^2)
```

![plot of chunk unnamed-chunk-2](figure/unnamed-chunk-2.png)
````

R + Markdown

.HTML

Sample Markdown

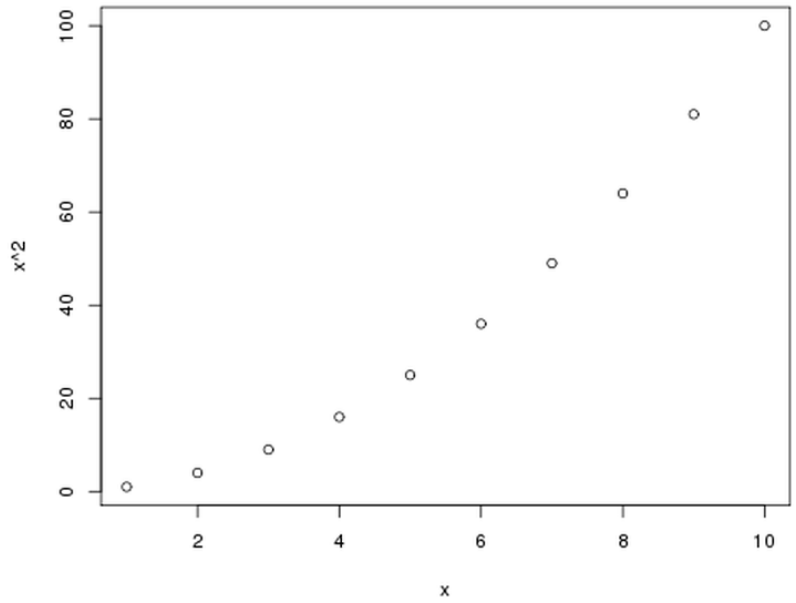
This is an R Markdown document!

```
x <- 1:10
x
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

You can also embed plots, for example:

```
plot(x, x^2)
```



TRESTLE TECHNOLOGY

Rstudio IDE

- Support for knitR & Sweave
  - R + Markdown
  - R + LaTeX
- One button push to generate either

Uses

- Great for sharing with non-technical audience
  - May or may not include the raw R code
  - They can follow the English
- Great for sharing data analysis with fellow R programmers
  - Encourages “literate programming”
  - Much easier to follow English than R
- Not great for sharing R functions

Packages

Outline

- Ingredients
- Tools
- Code

Publishing & Sharing

Ingredients

- /R
- /man
- /inst
- /src
- DESCRIPTION
- NAMESPACE

Ingredients

- /R
- /man
- /inst
- /src
- DESCRIPTION
- NAMESPACE

Out of Scope

Ingredients

- /R
- /man
- /inst
- /src
- DESCRIPTION
- NAMESPACE

Auto-Generated

Ingredients

- /R
- /man
- /inst
- /src
- DESCRIPTION
- NAMESPACE

Our responsibilities

DESCRIPTION

Package: RODProt

Type: Package

Title: JSON Table Schemas from R

Version: 0.1.3

Date: 2013-04-22

Author: Jeffrey D. Allen <jeffrey.allen@utsouthwestern.edu>

Maintainer: Jeffrey D. Allen <jeffrey.allen@utsouthwestern.edu>

Description: An R Client for Interacting with Data Encoded in one of the 'Open Data Protocols' Standards including JSON Table Schemas.

Imports:

rjson (>= 0.2.10),

RCurl,

License: MIT

Tools You'll Want

Packages

(Why this section is only 30 minutes long.)

roxygen2

- R Documentation (Rd) format
  - LaTeX-like, very fickle
  - Bad error checking

```
\alias{get_resource}
\title{Download Data Package Resource}
\usage{
  get_resource(dataPkg, resource, cache = TRUE, ...)
}
\arguments{
  \item{dataPkg}{The Data Package (as loaded from
  \code{\link{read_data_package}}) to use.}

  \item{resource}{A character string specifying the names
  of the resource to retrieve from the Data Package. At the
  time of development, it was unclear exactly how a
  resource would be indexed (see
  \url{https://github.com/dataprotocols/dataprotocols/issues
  thus the package supports a variety of indexing
  mechanisms for resources. The order of precedence is as
  follows: \enumerate{ \item Keys from the \code{resources}
  hash - If the \code{resources} value was provided as a
  JSON hash instead of an array, the required name
  prefacing each resource will be used as that resource's
  canonical name. \item The \code{id} field within each
  resource will be checked for any exact matches. \item The
  \code{name} field within each resource will be checked
  for any exact matches. \item The \code{url} and
  \code{path} fields within each resource will be checked
  for any exact matches. \item The \code{url} and
  \code{path} fields within each resource will be checked
  for partial matches on the extracted \code{<resource>}
  portion of a \code{/<path>/<resource>} reference. The
  path is extracted using \code{\link{dirname}} and ignored
  to see if the remainder of the string matches (with the
  possible exclusion of a prefacing slash). }}
```

https://github.com/klutometis/roxygen

roxygen2

- Roxygen2 is an R package
- Documentation is in your R code
- Compiles Rd files for you
- Integrated into RStudio IDE

```
#' @param cache Whether or not to
#' function will check this Data
#' it has, it will simply return
#' retrieve the remote resource t
#' involve storing an extra copy
#' function will retrieve the
#' resource remotely, but not sto
#' is to set cache to \code{"flus
#' resource remotely, ignoring an
#' for future use.
#' @param ... Arguments to be pas
#' @importFrom digest digest
#' @author Jeffrey D. Allen \emai
#' @export
get_resource <- function(dataPkg,
  if (cache != TRUE && cache != F
    stop("Invalid cache value. Mu
  }

  if (resource %in% names(dataPkg
    ind <- which(resource == name
  } else if (resource %in% lapply
```

https://github.com/klutometis/roxygen

TRESTLE TECHNOLOGY

devtools

- An R package
- Designed to make package authorship simpler
- Integrated into RStudio IDE

https://github.com/hadley/devtools/

Walkthrough

Packages

(Let's do it!)

Steps

1. Install devtools and roxygen2
2. Create an RStudio project with “package” build mode.
    1. Set to roxygenize on all three possibilities
    2. Set to build NAMESPACE, collate, and man fields
3. Create a package directory
4. Create an “R” directory
5. Put some Roxygenized R code into “./R”
6. Load/Install Devtools
7. load_all()
8. Fill out the description File

Steps

9. Build and Reload
10. Test & debug R code, go back to step 9
11. Check package before submitting
    - --no-manual if you don't have pdflatex

Resources

Packages

(How do I …?)

How to Get Help

- Authoritative Manual
  - http://cran.us.r-project.org/doc/manuals/R-exts.html
- Other Presentations
  - http://dl.dropboxusercontent.com/u/41902/easy-packages3.pdf
- StackOverflow
  - http://stackoverflow.com/questions/tagged/r
  - Q & A site with active R community
- Mailing Lists
  - R-devel, Bioconductor
  - http://www.r-project.org/posting-guide.html#which_list
  - Do your research; they bite

Distribution

Packages

(… Now what?)

Publishing and Sharing

CRAN

- Any domain
- Any code that passes

  R CMD CHECK
- Available online in a matter of days

Bioconductor

- Biological applications
- More rigorously curated
- 6 month release cycle

GitHub

- Social coding platform based on Git
  - Great collaboration tool
- Free for open-source usage, a few $/month for private repositories
- Use devtools' `install_github()` function to install an R package from a GitHub repo.

Tips

- Study and emulate the experts
  - https://github.com/hadley/stringr
  - https://github.com/yihui/knitr
- Learn Roxygen2
  - https://github.com/klutometis/roxygen
- Learn devtools
  - https://github.com/hadley/devtools/

Questions?

- Slides at
  - http://trestletech.com/blog/

Find me on:


@TrestleJeff


http://trestletech.com/feed


https://www.facebook.com/trestletech


http://www.linkedin.com/in/jeffreydallen1

TRESTLE TECHNOLOGY

Outstanding Question

(How do I make use of other packages?)

Using Other Packages

- Short answer: `imports`
- Long answer:
  - http://cran.r-project.org/doc/manuals/R-exts.html#Package-namespaces
  - http://stackoverflow.com/questions/9893791/imports-and-depends
  - http://stackoverflow.com/questions/8637993/better-explanation-of-when-to-use-imports-depends
- Rule of thumb: Don't use "`depends`" in DESCRIPTION. There's a better way now.