

Tools and practices for open and reproducible science with R

Dr. Jennifer (Jenny) Bryan

Dept. of Statistics & Michael Smith Laboratories, UBC

2013 November

<http://www.stat.ubc.ca/~jenny/>

<https://github.com/jennybc>

jenny@stat.ubc.ca

[@JennyBryan](https://twitter.com/JennyBryan)



links, files, etc. available here

The screenshot shows a GitHub repository page for `jennybc/2013-11_sfu`. The page title is "jennybc / 2013-11_sfu". The repository summary indicates 22 commits, 1 branch, 0 releases, and 1 contributor. The master branch is selected. The repository description is "Supporting documents for talk and workshop for SFU Statistics and Actuarial Science — Edit". The file list includes .gitignore, 2013-11_sfu-bryan-abstract.txt, 2013-11_sfu.Rproj, README.md, links.md, simple-markdown.md, and simple-r-markdown.md. The right sidebar provides links to Code, Issues (1), Pull Requests (0), Wiki, Pulse, Graphs, Network, and Settings. A clone URL is provided at the bottom.

Supporting documents for talk and workshop for SFU Statistics and Actuarial Science — Edit

22 commits 1 branch 0 releases 1 contributor

branch: master 2013-11_sfu / +

Fetching latest commit...

File	Description	Last Commit
.gitignore	gitignore Keynote	November 28, ...
2013-11_sfu-bryan-abstract.txt	title and abstract	November 27, ...
2013-11_sfu.Rproj	title and abstract	November 27, ...
README.md	ggplot2 needs no explanation but still ...	November 28, ...
links.md	Adding to workshop page	November 29, ...
simple-markdown.md	Simple markdown and R markdown example files	November 28, ...
simple-r-markdown.md	Simple markdown and R markdown example files	November 28, ...

Code Issues (1) Pull Requests (0) Wiki Pulse Graphs Network Settings

HTTPS clone URL https://github.com/jennybc/2013-11_sfu

meta-issues re: scientific computing

how to organize your work?

how to make work more pleasant for you?

how to make it navigable by others?

how to reduce tedium and manual processes?

how to reduce friction for collaboration?

how to reduce friction for communication?

specific tools and habits help you build a lot of this
into your normal coding and analysis process

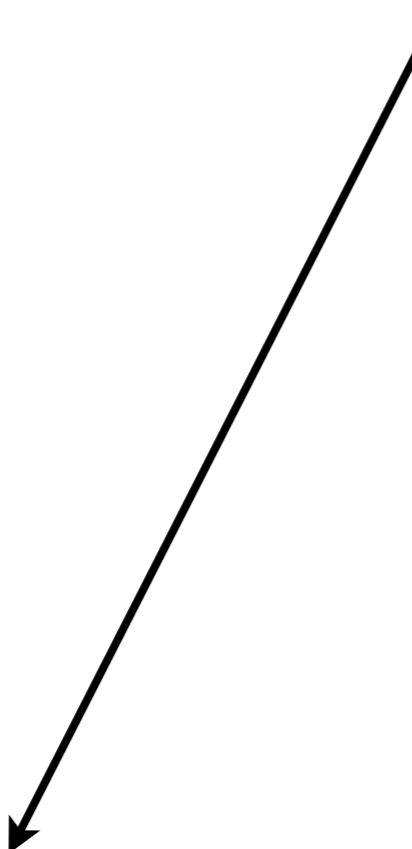
weak links in the chain: process, packaging and presentation



project organization
literate programming
reproducible research



collaboration / open science



version control / back up / archive

project organization / literate programming / reproducible research

Sweave
knitr
R markdown
R packages

RStudio

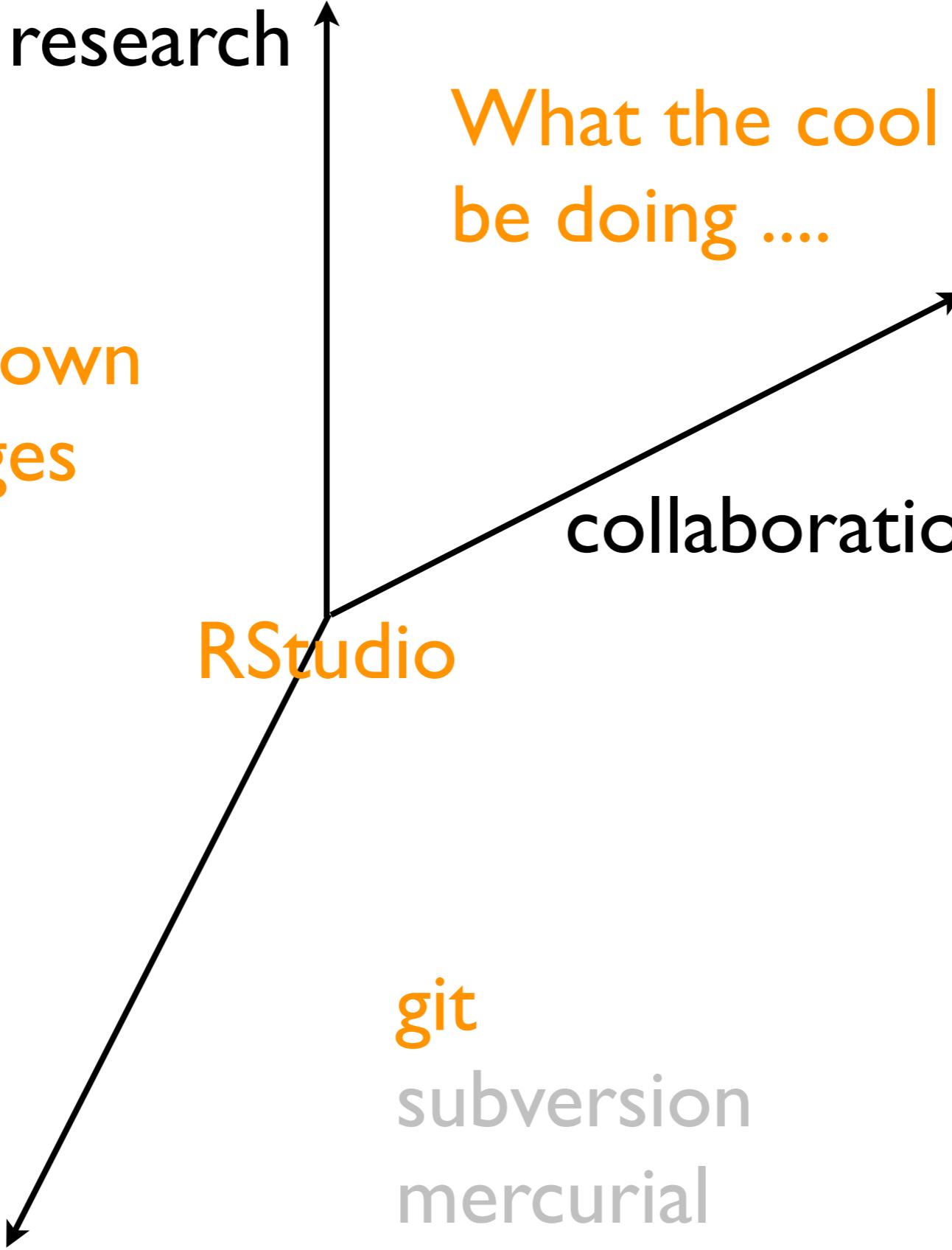
What the cool kids seem to
be doing

collaboration / open science

github
Rforge
sourceforge

git
subversion
mercurial

version control / back up / archive



RStudio is an integrated development environment (IDE) for R

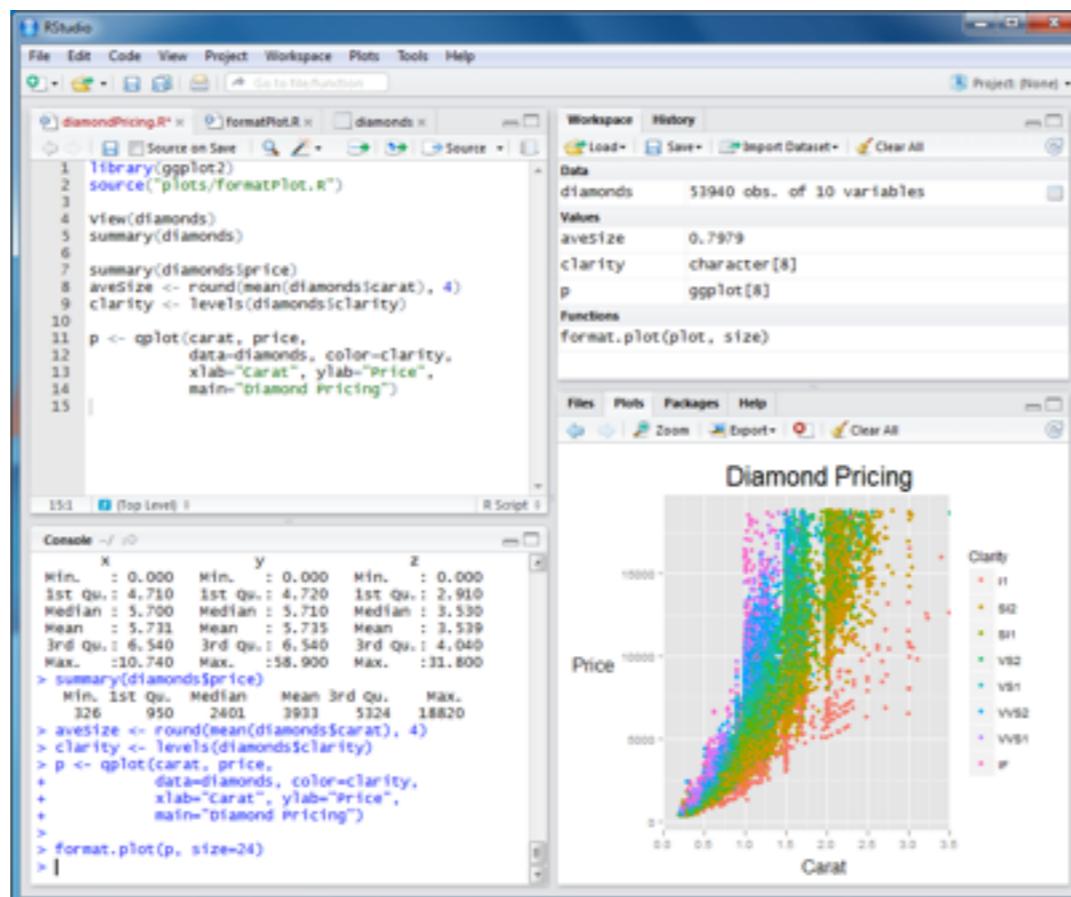
The screenshot displays the RStudio interface with the following components:

- Script Editor (Left Panel):** Shows an R script named "diamondPricing.R" with code for loading ggplot2, summarizing the diamonds dataset, and creating a scatter plot of Price vs. Carat by Clarity.
- Console (Bottom Left):** Displays the output of the R commands, including the summary statistics for the diamonds dataset and the resulting plot command.
- Workspace Browser (Top Right):** Shows the "diamonds" dataset with 53940 observations and 10 variables, and the created plot object "p".
- Plot Viewer (Bottom Right):** A scatter plot titled "Diamond Pricing" showing Price on the y-axis (0 to 15000) versus Carat on the x-axis (0.0 to 3.5). The data points are colored by Clarity, with a legend on the right mapping colors to clarity levels: I1 (red), SI2 (orange), SI1 (green), VS2 (light green), VS1 (cyan), VVS2 (blue), VVS1 (purple), and IF (pink).

R ≠ RStudio

RStudio mediates your interaction with R; it would replace Emacs + ESS or Tinn-R, but not R itself

Rstudio is a product of -- actually, more a driver of -- the emergence of R Markdown, knitr, R + Git(Hub)



Web Open

PDF → HTML
Latex → Markdown
Static → Interactive

Open source ↑↑
Open science ↑↑
Open research ↑↑

from Hadley Wickham's talk in the Simply Statistics Unconference on the Future of Statistics

Web Open

PDF → HTML
Latex → Markdown
Static → Interactive

Open source ↑↑
Open science ↑↑
Open research ↑↑

from Hadley Wickham's talk in the Simply Statistics Unconference on the Future of Statistics

Why Markdown?



<http://cpsievert.github.io/slides/markdown/#/1>

What is Markdown?

- Markdown is a lightweight markup language for creating HTML (or XHTML) documents.
- Markup languages are designed produce documents from human readable text (and annotations).
- Some of you may be familiar with *LaTeX*. This is another (less human friendly) markup language for creating pdf documents.
- Why I love Markdown:
 - Easy to learn and use.
 - Focus on **content**, rather than **coding** and debugging **errors**.
 - Once you have the basics down, you can get fancy and add HTML, JavaScript & CSS.

<http://cpsievert.github.io/slides/markdown/#/5>

Markdown



HTML

foo.md



foo.html

**easy to write
(and read!)**

**easy to publish
easy to read in
browser**

Markdown



HTML

```
Title (header 1, actually)  
=====
```



This is a Markdown document.

```
## Medium header (header 2, actually)
```

It's easy to do *italics* or make things bold.

> All models are wrong, but some are useful. An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. Absolute certainty is a privilege of uneducated minds-and fanatics. It is, for scientific folk, an unattainable ideal. What you do every day matters more than what you do once in a while. We cannot expect anyone to know anything we didn't teach them ourselves.

Enthusiasm is a form of social courage.

Code block below. Just affects formatting here but we'll get to R Markdown for the real fun soon!

```
```  
x <- 3 * 4
```
```

I can haz equations. Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

```
$$  
\begin{equation*}  
|x| =  
\begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases} \\\\  
\end{cases}  
\end{equation*}  
$$
```

```
<!DOCTYPE html>  
<html>  
<head>  
<meta http-equiv="Content-Type" content="text/html;  
charset=utf-8"/>
```

```
<title>Title (header 1, actually)</title>
```

```
<!-- MathJax scripts -->  
<script type="text/javascript" src="https://  
c328740.ssl.cf1.rackcdn.com/mathjax/2.0-latest/  
MathJax.js?config=TeX-AMS-MML_HTMLorMML">  
</script>
```

```
<style type="text/css">  
body {  
    font-family: Helvetica, arial, sans-serif;  
    font-size: 14px;  
    ...
```

```
<body>  
<h1>Title (header 1, actually)</h1>
```

```
<p>This is a Markdown document.</p>
```

```
<h2>Medium header (header 2, actually)</h2>
```

```
<p>It's easy to do <em>italics</em> or  
<strong>make things bold</strong>.</p>
```

```
<blockquote>  
<p>All models are wrong, but some are...  
<p>Code block below. Just affects formatting here  
but we'll get to R Markdown for the real fun  
soon!</p>
```

```
<pre><code>x < 3 * 4  
</code></pre>
```



Markdown

HTML

Title (header 1, actually)

This is a Markdown document.

Medium header (header 2, actually)

It's easy to do *italics* or make things bold.

> All models are wrong, but some are useful. An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. Absolute certainty is a privilege of uneducated minds-and fanatics. It is, for scientific folk, an unattainable ideal. What you do every day matters more than what you do once in a while. We cannot expect anyone to know anything we didn't teach them ourselves.

Enthusiasm is a form of social courage.

Code block below. Just affects formatting here but we'll get to R Markdown for the real fun soon!

```
```  
x <- 3 * 4
```
```

I can haz equations. Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

```
$$  
\begin{equation*}  
|x| =  
\begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases} \\\\  
\end{equation*}  
$$
```



Title (header 1, actually)

This is a Markdown document.

Medium header (header 2, actually)

It's easy to do *italics* or **make things bold**.

All models are wrong, but some are useful. An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. Absolute certainty is a privilege of uneducated minds-and fanatics. It is, for scientific folk, an unattainable ideal. What you do every day matters more than what you do once in a while. We cannot expect anyone to know anything we didn't teach them ourselves. Enthusiasm is a form of social courage.

Code block below. Just affects formatting here but we'll get to R Markdown for the real fun soon!

```
x <- 3 * 4
```

I can haz equations. Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

Markdown

HTML

Title (header 1, actually)

This is a Markdown document.

Medium header (header 2, actually)

It's easy to do *italics* or make things bold.

> All models are wrong, but some are useful. An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. Absolute certainty is a privilege of uneducated minds-and fanatics. It is, for scientific folk, an unattainable ideal. What you do every day matters more than what you do once in a while. We cannot expect anyone to know anything we didn't teach them ourselves.

Enthusiasm is a form of social courage.

Code block below. Just affects formatting here but we'll get to R Markdown for the real fun soon!

```
```  
x <- 3 * 4
```
```

https://github.com/jennybc/2013-11_sfu/blob/master/simple-markdown.md

```
\begin{equation*}  
|x| =  
\begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases} \\\\  
\end{equation*}  
$$
```

Title (header 1, actually)

This is a Markdown document.

Medium header (header 2, actually)

It's easy to do *italics* or **make things bold**.

<http://rpubs.com/jennybc/simple-markdown>

privilege of uneducated minds-and fanatics. It is, for scientific folk, an unattainable ideal. What you do every day matters more than what you do once in a while. We cannot expect anyone to know anything we didn't teach them ourselves. Enthusiasm is a form of social courage.

Code block below. Just affects formatting here but we'll get to R

wn for the real fun soon!

```
3 * 4
```

I can haz equations. Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

How is the math getting typeset ?!?

Short answer: Mathjax

MathJax is an open source JavaScript display engine for mathematics that works in all browsers.

No more setup for readers. No more browser plugins. No more font installations... It just works.

How painful is that to use?

Not at all. Automagic with knitr and RStudio.

What happens to my beautiful equations if the reader is not connected to the internet?

The LaTeX is displayed. No great harm.

Code block below. Just affects formatting here but we'll get to R Markdown for the real fun soon!

```
x <- 3 * 4
```

I can haz equations. Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

```
\begin{equation*} |x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases} \end{equation*}
```

If I use Markdown, am I restricted to HTML output?

No. Use pandoc, the “swiss-army knife” for converting files from one format to another. For example, pandoc can work with a local LaTeX installation to convert Markdown to PDF.

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife. Pandoc can convert documents in markdown, reStructuredText, textile, HTML, DocBook, LaTeX, MediaWiki markup, OPML, or Haddock markup to

- HTML formats: XHTML, HTML5, and HTML slide shows using Slidy, reveal.js, Slideous, S5, or DZSlides.
- Word processor formats: Microsoft Word docx, OpenOffice/LibreOffice ODT, OpenDocument XML
- Ebooks: EPUB version 2 or 3, FictionBook2
- Documentation formats: DocBook, GNU TexInfo, Groff man pages, Haddock markup
- Outline formats: OPML
- TeX formats: LaTeX, ConTeXt, LaTeX Beamer slides
- PDF via LaTeX
- Lightweight markup formats: Markdown, reStructuredText, AsciiDoc, MediaWiki markup, Emacs Org-Mode, Textile
- Custom formats: custom writers can be written in lua.

How do I actually convert Markdown to HTML?

knitr and RStudio provide user friendly wrapper functions and buttons; deep under the hood, you're using the Sundown markdown rendering library

Stick around for the workshop to take a test drive!

Markdown

HTML

```
Title (header 1, actually)
=====
```

This is a Markdown document.

```
## Medium header (header 2, actually)
```

It's easy to do *italics* or make things bold.

> All models are wrong, but some are useful. An approximate answer to the right problem is worth a

good d

approx

privil

for sc

you do

once i

anythi

Enthus

Code b

we'll

x <- 3

```
library(markdown)
```

```
markdownToHTML( "simple-markdown.md" )
```

I can

the av

$\wedge\{n\}$ x

\$\$

\begin{

|x| =

\begin{

-x &\text{if \$x \leq 0\$}.

\end{cases}

\end{equation*}

\$\$

Title (header 1, actually)

This is a Markdown document.

Medium header (header 2, actually)

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

Markdown

Title (header 1, actually)

This is a Markdown document.

```
## Medium header (header 2, actually)
```

It's eas

> All mo
approxim
good dea
approxim
privileg
for scie
you do e
once in
anything
Enthusia

Code block

1

1

T-Scan b2

I can have
the aver-

$\hat{x}^n x_{-n}$

JOURNAL OF CLIMATE

2

```
\begin{e
```

$$|x| =$$

```
\begin{c
```

\begin{tikzpicture}

-x & \tex
\ - 1

\end{cases}

\end{equ}

\$\$

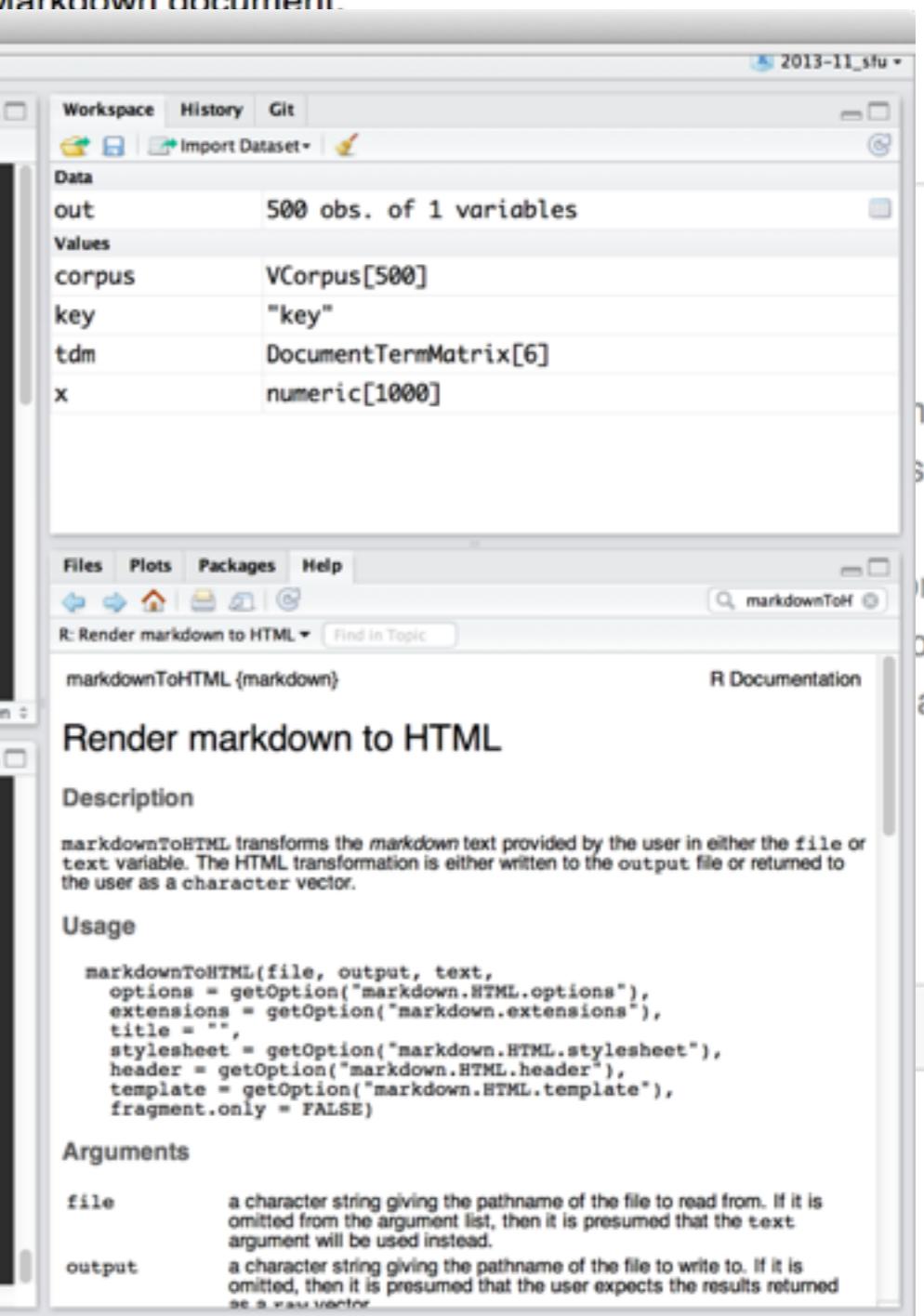
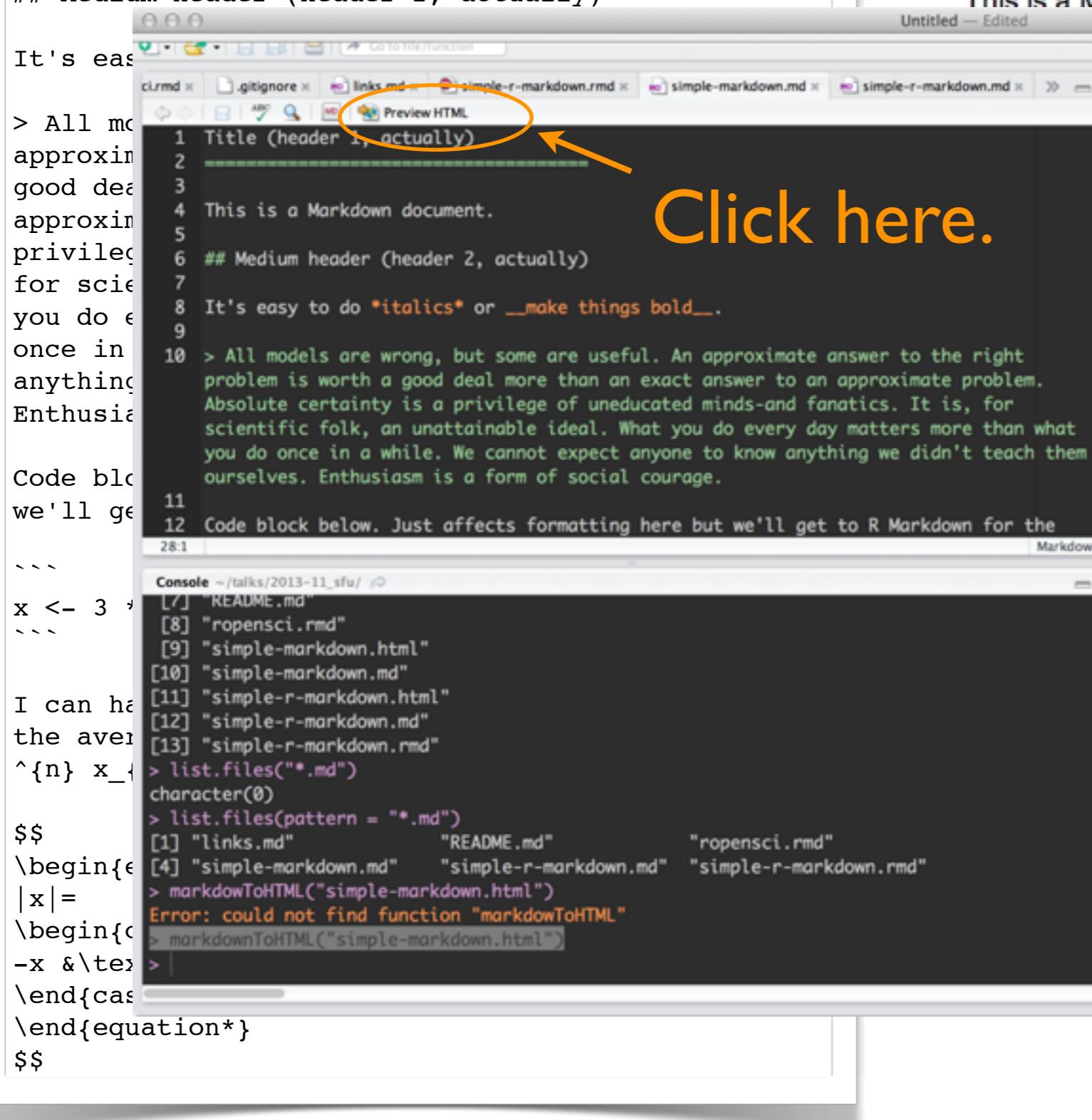
HTML

Title (header 1, actually)

This is a Markdown document

Untitled — Edited

Click here.



$$w = \begin{cases} -x & \text{if } x \leq 0. \end{cases}$$

If you have an annoying process for authoring for the web

or

If you avoid authoring for the web, because you're not sure how ...

start writing in Markdown.

R markdown

....

Code block below. Just affects
formatting here but we'll get to R
Markdown for the real fun soon!

```
```  
x <- 3 * 4
```
```

Remember this code block from the
plain Markdown example?

In R Markdown, we can insert real R code!

When processed by knitr, the output will be
inserted into the resulting Markdown
document. Which you know is easy to convert
to HTML or PDF or whatever.

R Markdown

Markdown

R Markdown rocks

This is an R Markdown document.

```
```{r}
x <- rnorm(1000)
head(x)
```
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the `r length(x)` random normal variates we just generated is `r round(mean(x), 3)`. Those numbers are NOT hard-wired but are computed on-the-fly. As is this figure. No more copy-paste ... copy-paste ... oops forgot to copy-paste.

```
```{r}
plot(density(x))
```
```

Note that all the previously demonstrated math typesetting still works. You don't have to choose between having math cred and being web-friendly!

Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

```
$$
\begin{aligned}
|x| = \\
\begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}
\end{aligned}
$$
```

R Markdown rocks

This is an R Markdown document.

```
```{r}
x <- rnorm(1000)
head(x)
```
```
[1] -1.3007 0.7715 0.5585 -1.2854 1.1973
2.4157
````
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the 1000 random normal variates we just generated is -0.081. Those numbers are NOT hard-wired but are computed on-the-fly. As is this figure. No more copy-paste ... copy-paste ... oops forgot to copy-paste.

```
```{r}
plot(density(x))
```
```

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

...

Markdown → HTML

R Markdown rocks

This is an R Markdown document.

```
```r
x <- rnorm(1000)
head(x)
```
```
[1] -1.3007 0.7715 0.5585 -1.2854 1.1973
2.4157
````
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the 1000 random normal variates we just generated is -0.081. Those numbers are NOT hard-wired but are computed on-the-fly. As is this figure. No more copy-paste ... copy-paste ... oops forgot to copy-paste.

```
```r
plot(density(x))
````
```

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

...

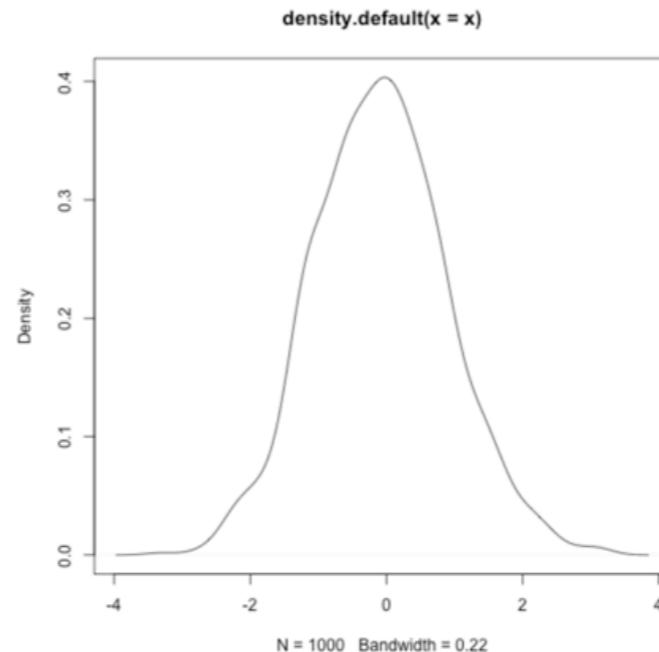
R Markdown rocks

This is an R Markdown document.

```
x <- rnorm(1000)
head(x)
```

```
## [1] -1.3007  0.7715  0.5585 -1.2854  1.1973  2.4157
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the 1000 random normal variates we just generated is -0.081. Those numbers are NOT hard-wired but are computed on-the-fly. As is this figure. No more copy-paste ... copy-paste ... oops forgot to copy-paste.



Note that all the previously demonstrated math typesetting still works. You don't have to choose between having math cred and being web-friendly!

Inline equations, such as ... the average is computed as $\frac{1}{n} \sum_{i=1}^n x_i$. Or display equations like this:

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

R Markdown → Markdown → HTML

foo.rmd → **foo.md** → **foo.html**

**easy to write
(and read!)**

**easy to publish
easy to read in
browser**

R Markdown

R Markdown rocks

This is an R Markdown document.

```
`` ` {r}  
x <- rnorm(1000)  
head(x)  
```
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the `r length(x)` random normal variates we just generated is `r round(mean(x), 3)`. Those numbers are NOT hard-wired but are computed on-the fly. As is this figure. No more copy-paste ... cop paste ... oops forgot to copy-paste.

```
```{r}  
plot(density(x))  
```
```

Note that all the previously demonstrated math typesetting still works. You don't have to choose between `\begin{array}{c} \frac{1}{2} \\ + \end{array}` and `\frac{1}{2} +`.

Inline  
computed  
display

## Master/Simple

# → HTML

# R Markdown rocks

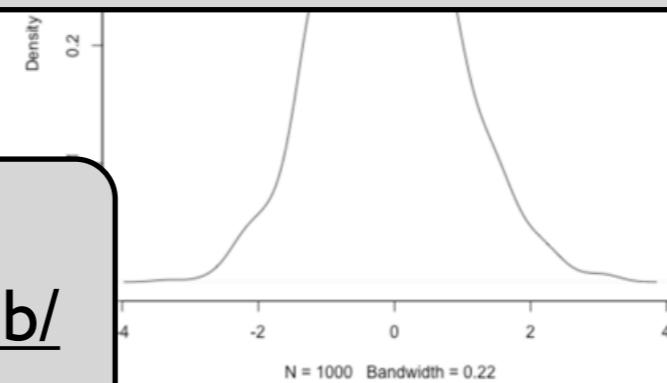
This is an R Markdown document

```
x <- rnorm(1000)
head(x)
```

```
[1] -1.3007 0.7715 0.5585 -1.2854 1.1973 2.4157
```

See how the R code gets executed and a representation thereof appears in the document? `knitr` gives you control over how to represent all conceivable types of output. In case you care, then average of the 1000 random normal variates we just generated is -0.081. Those numbers are NOT hard-wired but

<http://rpubs.com/jennybc/simple-r-markdown>



usly demonstrated math typesetting still works. You don't even have to be having math cred and being web-friendly!

Inline equations, such as ... the average is computed as  $\frac{1}{n} \sum_{i=1}^n x_i$ . Or display equations like this:

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

# How do I actually convert R Markdown to HTML?

knitr and RStudio provide user friendly wrapper functions and buttons; deep under the hood, you're using the Sundown markdown rendering library

Stick around for the workshop to take a test drive!

# R Markdown

→ HTML

R Markdown rocks

This is an R Markdown document.

```
```{r}
x <- rnorm(1000)
head(x)
```
```

See how the R code gets executed and a

represen

`knitr`

conceal

averag

we jus

number

fly. A

paste

```
```{r}
```

plot(d

```
```
```

Note t

typese

betwee

Inline

comput

displa

\$\$

\begin{

|x| =

\begin{

-x & \text{if } x \leq 0\}

\end{cases}

\end{equation\*}

\$\$

# R Markdown rocks

This is an R Markdown document.

```
x <- rnorm(1000)
head(x)
```

```
[1] -1.3007 0.7715 0.5585 -1.2854 1.1973 2.4157
```

See how the R code gets executed and a representation thereof appears in the

## How to achieve at the command line:

```
library(knitr)
knit2html("simple-r-markdown.rmd")
```

FYI:

```
knit2html() = knit() + markdownToHTML()
```

equations like this:

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

# R Markdown

# HTML

R Markdown rocks  
=====

This is an R Markdown document.

```
```{r}
x <- rnorm(1000)
head(x)
```
See how the R code gets executed and a representation thereof appears in the document?
```

```
Note that typesetting between here and there is handled by knitr. It's a great way to include computed values in your documents without having to worry about the details of the computation. For example, if you want to include the results of a complex computation in your document, you can simply wrap it in a chunk like this:
```

```
```{r}
x <- rnorm(1000)
head(x)
```
See how the R code gets executed and a representation thereof appears in the document?
```

```

\begin{equation*}
|x| = \sqrt{x^2}
\end{equation*}
```

# R Markdown rocks

This is an R Markdown document.

```
x <- rnorm(1000)
head(x)
```

~/talks/2013-11\_sfu - RStudio

2013-11\_sfu

Workspace History Git

Data out 500 obs. of 1 variables

Values corpus VCorpus[500]

key "key"

tdm DocumentTermMatrix[6]

x numeric[1000]

Files Plots Packages Help

New Folder Delete Rename More

Home talks 2013-11\_sfu

| Name                                       | Size      | Modified               |
|--------------------------------------------|-----------|------------------------|
| ..                                         |           |                        |
| .gitignore                                 | 35 bytes  | Nov 28, 2013, 9:35 AM  |
| .Rhistory                                  | 105 bytes | Nov 28, 2013, 9:29 AM  |
| 2013-11_sfu-bryan-abstract.txt             | 988 bytes | Nov 13, 2013, 1:15 PM  |
| 2013-11_sfu.Rproj                          | 205 bytes | Nov 27, 2013, 10:59 AM |
| README.md                                  | 4.3 KB    | Nov 28, 2013, 9:08 AM  |
| ropensci.rmd                               | 2.3 KB    | Nov 28, 2013, 12:10 PM |
| 2013-11_sfu-rmarkdown-knitr-git-github.key | 40.5 MB   | Nov 28, 2013, 1:55 PM  |
| 2013-11_sfu-workshop.key                   | 30.5 MB   | Nov 28, 2013, 10:14 AM |
| links.md                                   | 1.8 KB    | Nov 28, 2013, 12:54 PM |
| simple-markdown.md                         | 1 KB      | Nov 28, 2013, 11:30 AM |
| simple-markdown.html                       | 7 KB      | Nov 28, 2013, 11:28 AM |
| simple-r-markdown.rmd                      | 986 bytes | Nov 28, 2013, 1:04 PM  |
| simple-r-markdown.html                     | 43.8 KB   | Nov 28, 2013, 1:54 PM  |
| simple-r-markdown.md                       | 1.1 KB    | Nov 28, 2013, 1:54 PM  |
| figure                                     |           |                        |

Click here.

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x \leq 0. \end{cases}$$

**Do I have to do everything in R markdown? What about plain R scripts?**

**Use `knitr::stitch()` or Rstudio's Compile Notebook button to get a satisfying stand-alone webpage based on an R script.**

**Stick around for the workshop to see how!**

# simple R script: toyline.R

```
1 a <- 2
2 b <- 7
3 sigSq <- 0.5
4 n <- 400
5
6 set.seed(1234)
7 x <- runif(n)
8 y <- a + b * x + rnorm(n, sd = sqrt(sigSq))
9
10 (avgX <- mean(x))
11
12 plot(x, y)
13 abline(a, b, col = "blue", lwd = 2)
```

→ **HTML**

## toyline.R

jenny — Sep 6, 2013, 3:15 PM

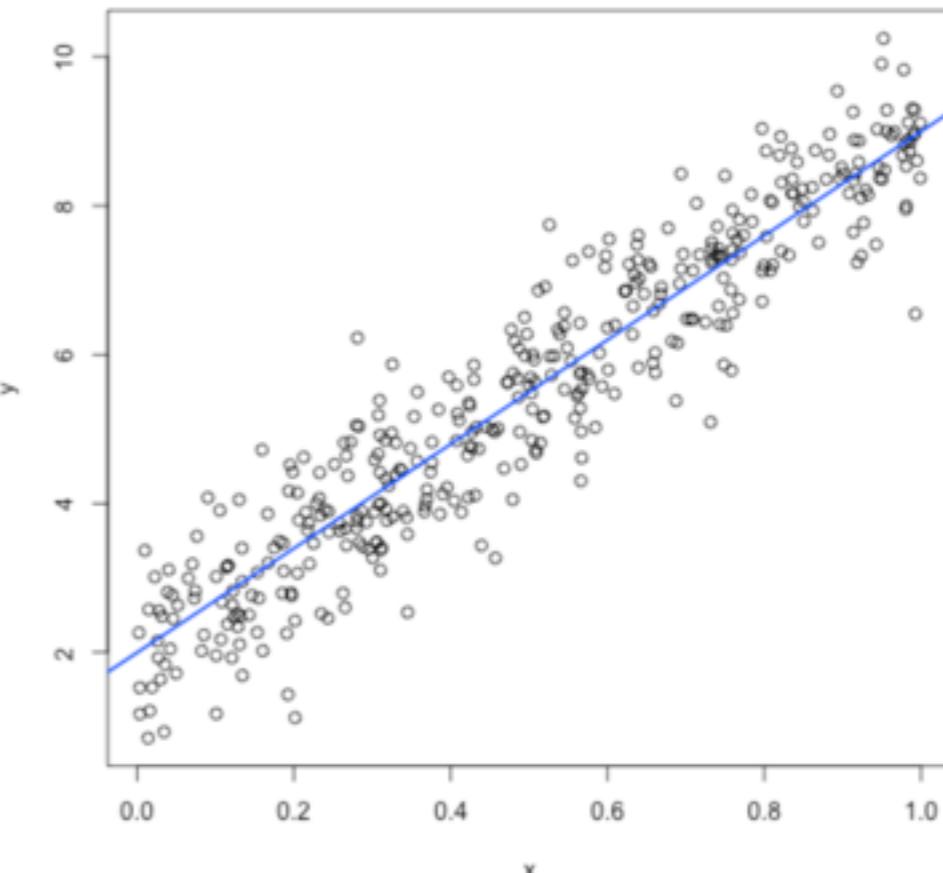
```
a <- 2
b <- 7
sigSq <- 0.5
n <- 400

set.seed(1234)
x <- runif(n)
y <- a + b * x + rnorm(n, sd = sqrt(sigSq))

(avgX <- mean(x))
```

```
[1] 0.4969
```

```
plot(x, y)
abline(a, b, col = "blue", lwd = 2)
```



How do I show the world all these awesome dynamic HTML reports I'm creating?

Easiest: Rpubs

Stick around for the workshop to see how!

Rest assured, all the usual ways to get stuff on the web still apply. Also stayed tuned to the GitHub bit later. You can still choose to make things public or private.

# How students submitted their homework to me this fall:

A screenshot of a GitHub Gist page titled "Toy example used in STAT 545A". The page shows an R script named "toyline.R" with the following code:

```
a <- 2
b <- 7
sigSq <- 0.5
n <- 400

set.seed(1234)
x <- runif(n)
y <- a + b * x + rnorm(n, sd = sqrt(sigSq))

(avgX <- mean(x))

plot(x, y)
abline(a, b, col = "blue", lwd = 2)

sessionInfo()
```

The page includes a "Gist Detail" sidebar with options like "Revisions" and "Download Gist". Below the code editor is a "Comments" section with "Write" and "Preview" buttons. A plus sign (+) is overlaid on the bottom left of the image.

a .r or .rmd  
file, posted on  
the web; default  
is to publish as  
a gist

A screenshot of an RPubs page titled "RPubs - toyline.R". The page displays the same R script as the GitHub Gist, followed by its output:

```
a <- 2
b <- 7
sigSq <- 0.5
n <- 400

set.seed(1234)
x <- runif(n)
y <- a + b * x + rnorm(n, sd = sqrt(sigSq))

(avgX <- mean(x))

[1] 0.4969

plot(x, y)
abline(a, b, col = "blue", lwd = 2)
```

Below the code is a scatter plot of y vs x with a blue regression line. A plus sign (+) is overlaid on the bottom left of the image.

a readable report,  
posted on the  
web; default is to  
use Rpubs

+  
21 Please add your link in this bulleted list:  
22  
23 \* Jenny Bryan: toyline [script](<https://gist.github.com/jennybc/6520226>) |  
[report](<http://rpubs.com/jennybc/toyline>)  
24 \* Matthew Gingerich: [TestFile](<http://rpubs.com/majugi/TestFile>)  
25 \* Justin Chu: [toylineTest](<http://rpubs.com/cjustin/8316>)  
--

Please add your link in this bulleted list:

- Jenny Bryan: toyline [script](#) | [report](#)
- Matthew Gingerich: [TestFile](#)
- Justin Chu: [toylineTest](#)

send instructor  
the links



Big picture, so far:

web-friendly is good

various hosting platforms make it easy to share web-ready products with minimal effort

embedding analysis and logic in source document for a report is good

- huge win for reproducibility
- also excellent for communication and documentation

(R) Markdown + knitr (+ RStudio) make it very easy to author dynamic reports that are ready for the web

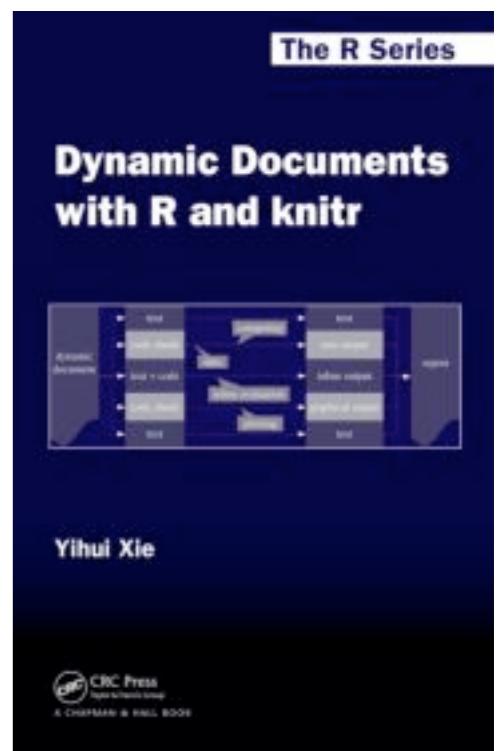
disclaimer:

knitr is **not limited** to executing R code

knitr is **not limited** to processing R  
Markdown

I chose to focus on R and R Markdown

Read more in the book or [on the web](#):  
*Dynamic documents with R and knitr* by Yihui Xie, part of the CRC Press / Chapman & Hall R Series (2013). ISBN: 9781482203530.



how do I put my source on the web?  
for the world or select collaborators?

maybe I'll share my data and my notes too ....  
how should I marshall all of that stuff?

how can I collaborate with others on an analysis  
or package development?

how do I put my source on the web?  
for the world or select collaborators?

maybe I'll share my data and my prose too ....  
how should I marshall all of that stuff?

how can I collaborate with others on an analysis  
or package development?

Advice to preserve sanity:

Stop doing this via email, attachments, and tracking changes in Word. Get that stuff into plain text, put it under version control and get it out on the web.

# "FINAL".doc



FINAL.doc!



FINAL\_rev.2.doc



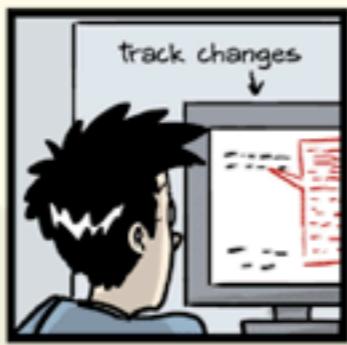
↑  
FINAL\_rev.6.COMMENTS.doc



FINAL\_rev.8.comments5.  
CORRECTIONS.doc



JORGE CHAM © 2012



FINAL\_rev.18.comments7.  
corrections9.MORE.30.doc



FINAL\_rev.22.comments49.  
corrections.10.#@\$%WHYDID  
ICOMETOGRADSCHOOL????.doc



WWW.PHDCOMICS.COM

<http://www.phdcomics.com/comics/archive.php?comicid=1531> via Ram, 2013 doi:10.1186/1751-0473-8-7

Version control systems (VCS) were created to help groups of people develop software

Developers needed a tool to facilitate structured editing of a group files, usually text (source code)

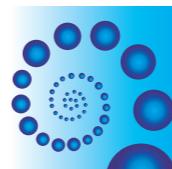
Well known VCSs: Subversion (SVN), Mercurial (hg), Git

Git, in particular, is being “repurposed” for activities other than pure software development ... like the messy hybrid of writing, coding and data wrangling most of us engage in 24/7

“Git, provides a lightweight yet robust framework that is ideal for managing the full suite of research outputs such as datasets, statistical code, figures, lab notes, and manuscripts.”

“... this tool can be leveraged to make science more reproducible and transparent, foster new collaborations, and support novel uses.”

Ram *Source Code for Biology and Medicine* 2013, **8**:7  
<http://www.scfbm.org/content/8/1/7>



SOURCE CODE FOR  
BIOLOGY AND MEDICINE

BRIEF REPORTS

Open Access

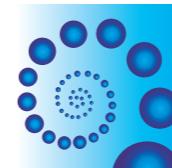
Git can facilitate greater reproducibility and increased transparency in science

Karthik Ram

“Git, provides a lightweight yet robust framework that is ideal for managing the full suite of research outputs such as datasets, statistical code, figures, lab notes, and manuscripts.”

“... this tool can be leveraged to make science more reproducible and transparent, foster new collaborations, and support novel uses.”

Ram *Source Code for Biology and Medicine* 2013, **8**:7  
<http://www.scfbm.org/content/8/1/7>



SOURCE CODE FOR  
BIOLOGY AND MEDICINE

BRIEF REPORTS

Open Access

## Git can facilitate greater reproducibility and increased transparency in science

Karthik Ram

GitHub repository for this paper: [https://github.com/karthik/smb\\_git](https://github.com/karthik/smb_git)

Ram: Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine* 2013 **8**:7. doi:10.1186/1751-0473-8-7

collaboration = the “killer app” of version control

Learning Git has been -- and continues to be -- painful. But not nearly as crazy-making as the alternatives:

- documents as email attachments
- uncertainty about which version is “master”
- am I working with the most recent data?
- archaeological “digs” on old email threads
- uncertainty about how/if certain changes have been made or issues solved
- hair-raising ZIP archives containing file salad

Git **repository** = a bunch of files you want to manage in a sane way

repo = repository

you can set up repo ... then start your work

or you can make a set of existing files and make them into a repo

# GitHub = a place to host Git repositories on the web

## GitHub ≠ Git

The screenshot shows the GitHub repository page for `RcppCore/Rcpp`. The top navigation bar includes icons for file operations, a GitHub logo, and the URL `github.com/RcppCore/Rcpp`. The main header displays the repository name `RcppCore / Rcpp`, a watch count of 5, a star count of 2, and a fork count of 1. Below the header, a summary box shows 2,595 commits, 1 branch, 0 releases, and 6 contributors. A dropdown menu indicates the current branch is `master`. The main content area lists recent commits from `eddelbuettel`, such as fixing unused variable warnings and updating build scripts. On the right side, there's a sidebar with links for `Code`, `Issues` (47), `Pull Requests` (0), `Wiki`, `Pulse`, `Graphs`, and `Network`. At the bottom, there are options to clone the repository via `HTTPS`, `Clone in Desktop`, or `Download ZIP`.

Seamless R and C++ Integration

2,595 commits · 1 branch · 0 releases · 6 contributors

branch: master · Rcpp / +

suppress two unused variable warnings from g++-4.8  
eddelbuettel authored 2 days ago · latest commit b843b2e1e1

R include the package file first. closes #64 4 days ago

debian minor cosmetics for the Debian build a month ago

inst suppress two unused variable warnings from g++-4.8 2 days ago

man export Rcpp.plugin.maker from the NAMESPACE. closes #65 4 days ago

src mark functions as registered and make sure they don't throw. closes #73 3 days ago

tests fix R CMD check not finding sourceCpp files a year ago

vignettes correct use of href 3 days ago

.Rbuildignore added travis.yml to enable continuous integration on github 23 days ago

.Rinstignore added to exclude inst/doc/{Makefile,jss.bst} from making it into the ... 2 years ago

.gitignore ignoring inst/lib 20 days ago

.travis.yml no mas – per packages.ubuntu.com g++-4.7 appeared with release 12.10 5 days ago

ChangeLog expand unit tests for pt() and correct use of pt() with ncp argument 6 days ago

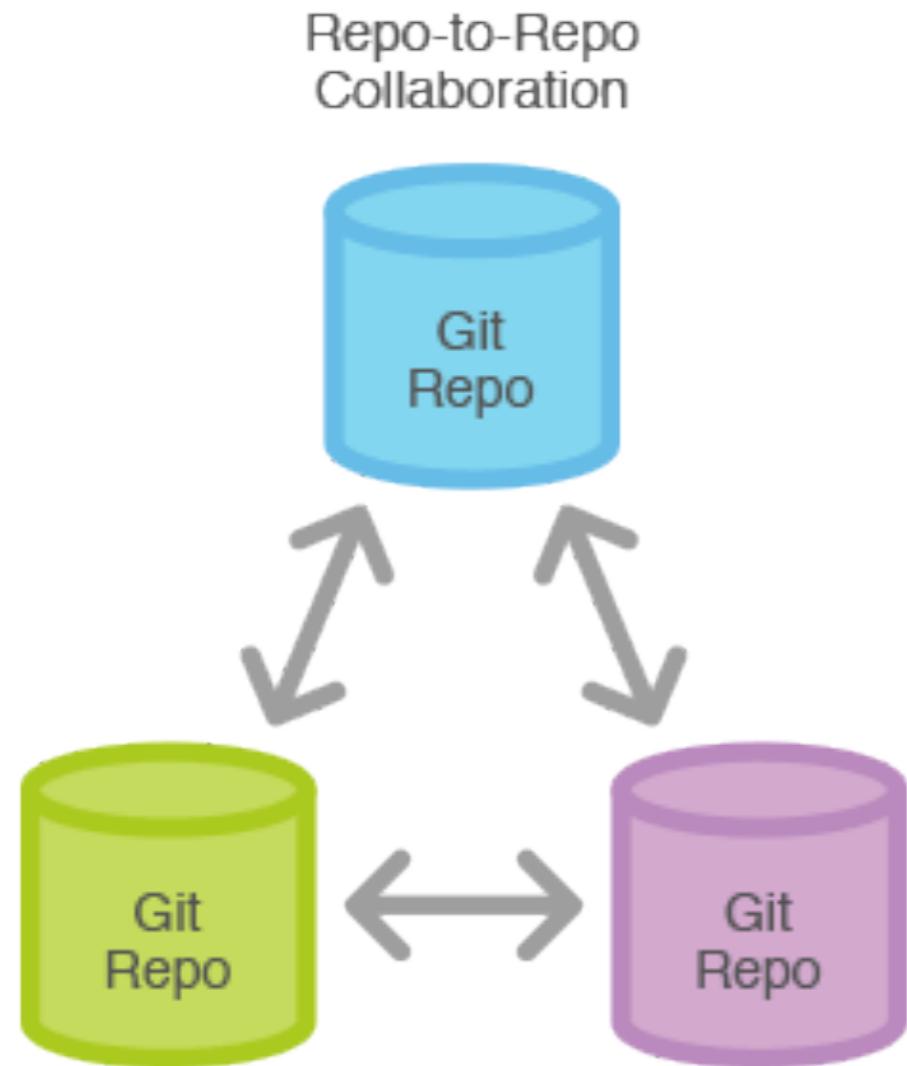
Code Issues (47) Pull Requests (0) Wiki Pulse Graphs Network

HTTPS clone URL  
<https://github.com>

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop Download ZIP

# possible, in theory



# more typical

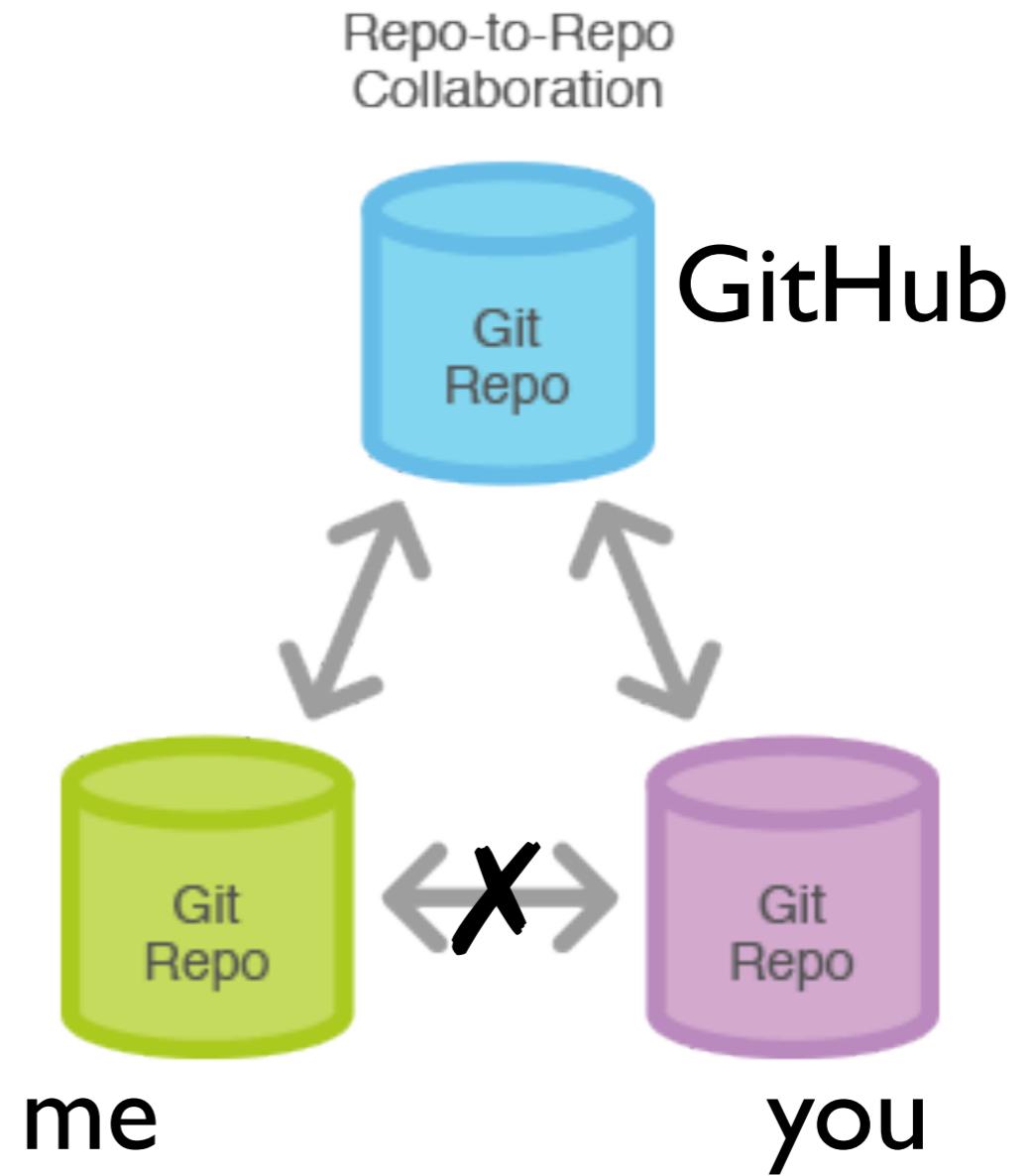


Image from <https://www.atlassian.com/git/tutorial/git-basics#!clone>

yihui/knitr

Syllabus and lecture pages | Br...

github

Search or Type a Command

Explore Gist Blog Help

yihui / knitr

PUBLIC

Code Network

A general-purpose tool for dynamic report generation.

<http://yihui.name/knitr>

Clone In N

branch: mas

Latest commit to the master branch

bump version

yihui authored

knitr /

name

R

inst

man

tests

.Rbuildignore

.Rinstignore

.gitignore

Syllabus and lecture pages | Br...

github

Search or Type a Command

Explore Gist Blog Help

hadley / ggplot2

PUBLIC

Code Network Pull Requests 11 Issues 114

hadley/plyr

GitHub, Inc. [github.com/hadley/plyr](https://github.com/hadley/plyr)

Syllabus and lecture pages | Br...

github

Search or Type a Command

Explore Gist Blog Help

hadley / plyr

PUBLIC

Code

A R package for splitting, applying, mapping, and combining data structures.

<http://plyr.had.co.nz>

Clone in Mac ZIP

branch: master

Latest commit to the master branch

Version bump

hadley authored 9 months ago

jennybc

hadley/

Many R packages are developed in the open on GitHub

Nice option when someone tells you to “read the source”!

You can see exactly how files have changed, when, and by whom. If commit message is good, you'll see why.  
Commit = a formal “checkpoint” or snapshot of the state of the repository

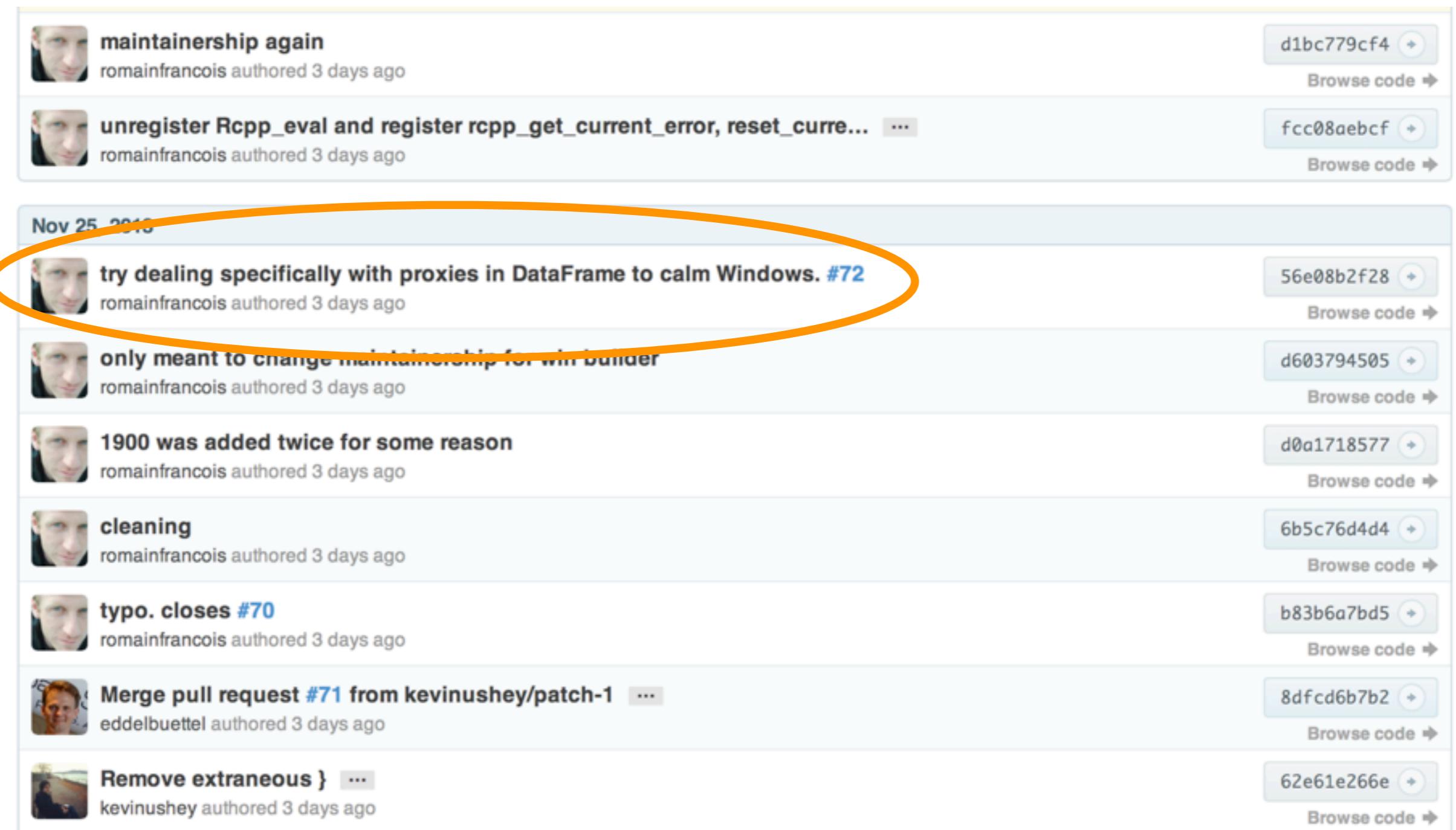
The screenshot shows the GitHub repository page for `RcppCore/Rcpp`. At the top, there's a navigation bar with icons for back, forward, search, and user profile. Below it is a header with the repository name, a search bar, and links for Explore, Gist, Blog, and Help. On the right, there are buttons for Watch (5), Star (2), Fork (1), and a user profile for `jennybc`.

The main content area displays the repository's activity. A yellow circle highlights the "2,595 commits" statistic. Below this, a timeline shows recent commits:

- suppress two unused variable warnings from g++-4.8 (eddelbuettel, 2 days ago)
- include the package file first. closes #64 (R, 4 days ago)
- minor cosmetics for the Debian build (debian, a month ago)
- suppress two unused variable warnings from g++-4.8 (inst, 2 days ago)
- export Rcpp.plugin.maker from the NAMESPACE. closes #65 (man, 4 days ago)
- mark functions as registered and make sure they don't throw. closes #73 (src, 3 days ago)
- fix R CMD check not finding sourceCpp files (tests, a year ago)
- correct use of href (vignettes, 3 days ago)

On the right side, there are links for Code, Issues (47), Pull Requests (0), Wiki, Pulse, Graphs, and Network. At the bottom, there's an HTTPS clone URL field with the URL `https://github.com/jennybc/RcppCore`, a note about cloning with HTTPS, SSH, or Subversion, and a "Clone in Desktop" button.

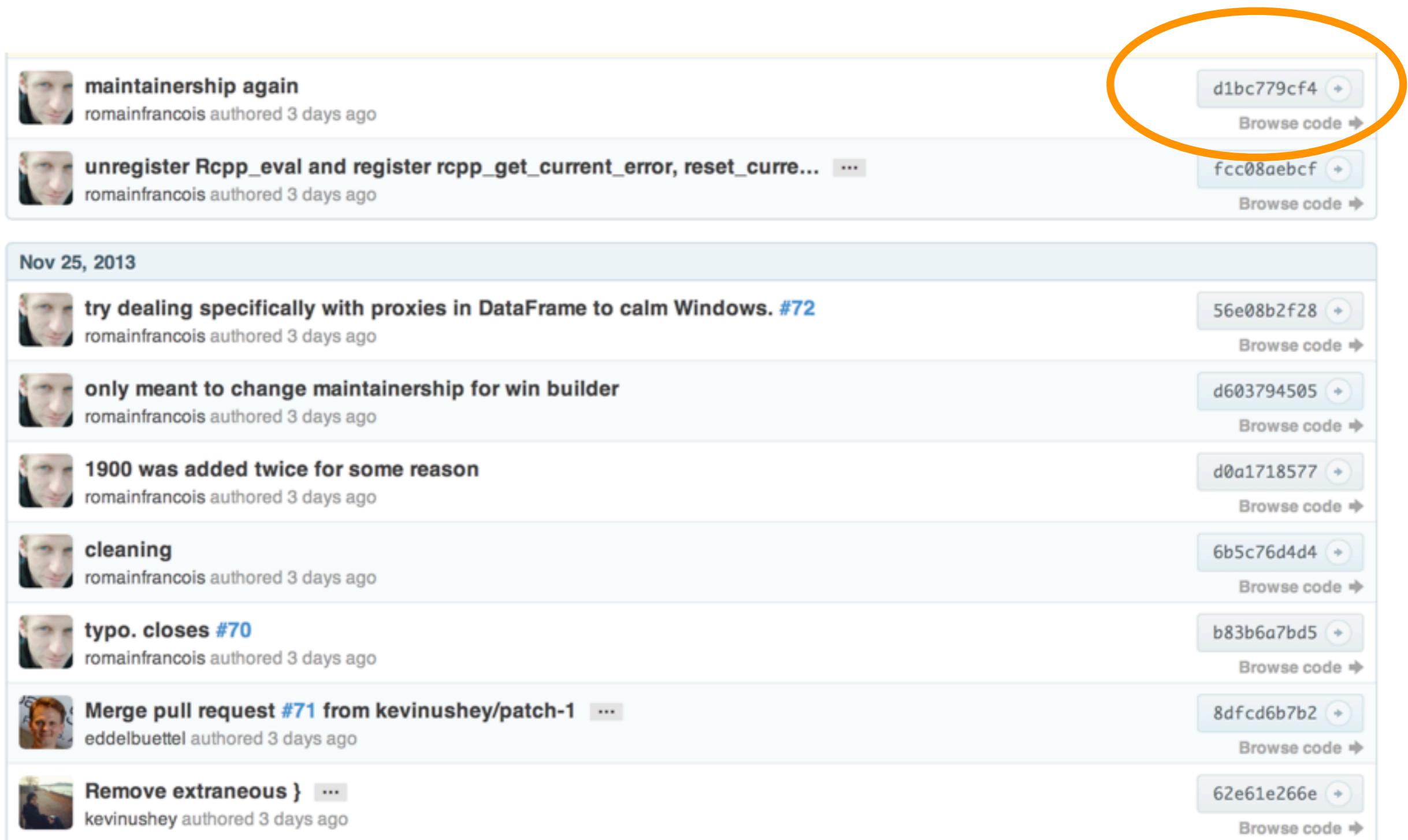
# Commit message = short description of what's changed



Nov 25, 2013

-  **maintainership again** d1bc779cf4 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **unregister Rcpp\_eval and register rcpp\_get\_current\_error, reset\_curre...** fcc08aebcf ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **try dealing specifically with proxies in DataFrame to calm Windows. #72** 56e08b2f28 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **only meant to change maintainership for win builder** d603794505 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **1900 was added twice for some reason** d0a1718577 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **cleaning** 6b5c76d4d4 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **typo. closes #70** b83b6a7bd5 ↗  
Browse code ➔  
romainfrancois authored 3 days ago
-  **Merge pull request #71 from kevinushey/patch-1** 8dfcd6b7b2 ↗  
Browse code ➔  
eddelbuettel authored 3 days ago
-  **Remove extraneous }** 62e61e266e ↗  
Browse code ➔  
kevinushey authored 3 days ago

# Click on a commit and GitHub provides a fantastic visual “diff” view of exactly what changed. Incredibly useful.



The screenshot shows a list of GitHub commits from November 25, 2013. The commits are listed in chronological order from top to bottom. Each commit includes the author's profile picture, the commit message, the date it was authored, and two buttons: one for the commit hash and one for 'Browse code'. A large orange circle highlights the commit message and the 'Browse code' button for the first commit in the list.

| Date         | Author         | Commit Message                                                           | Commit Hash | Action      |
|--------------|----------------|--------------------------------------------------------------------------|-------------|-------------|
| Nov 25, 2013 | romainfrancois | maintainership again                                                     | d1bc779cf4  | Browse code |
| Nov 25, 2013 | romainfrancois | unregister Rcpp_eval and register rcpp_get_current_error, reset_curre... | fcc08aebcf  | Browse code |
| Nov 25, 2013 | romainfrancois | try dealing specifically with proxies in DataFrame to calm Windows. #72  | 56e08b2f28  | Browse code |
| Nov 25, 2013 | romainfrancois | only meant to change maintainership for win builder                      | d603794505  | Browse code |
| Nov 25, 2013 | romainfrancois | 1900 was added twice for some reason                                     | d0a1718577  | Browse code |
| Nov 25, 2013 | romainfrancois | cleaning                                                                 | 6b5c76d4d4  | Browse code |
| Nov 25, 2013 | romainfrancois | typo. closes #70                                                         | b83b6a7bd5  | Browse code |
| Nov 25, 2013 | eddelbuettel   | Merge pull request #71 from kevinushey/patch-1                           | 8dfcd6b7b2  | Browse code |
| Nov 25, 2013 | kevinushey     | Remove extraneous }                                                      | 62e61e266e  | Browse code |

# GitHub provides a fantastic visual “diff” view of exactly what changed. Incredibly useful.

**mark functions as registered and make sure they don't throw. closes #73** [Browse code](#)

>Loading branch information...

 **romainfrancois** authored November 26, 2013 1 parent **d1bc779** commit **3331e7b48934936aa8c4782ba6ec06714da7f670**

 Showing 2 changed files with 29 additions and 11 deletions. [Show Diff Stats](#)

| 19  | 227 | src/api.cpp                                                                                                      | <a href="#">View file @ 3331e7b</a> |
|-----|-----|------------------------------------------------------------------------------------------------------------------|-------------------------------------|
| ... | ... | @@ -227,10 +227,11 @@ SEXP stack_trace( const char* file, int line ){                                            |                                     |
| 227 | 227 | #if defined(__GNUC__)                                                                                            |                                     |
| 228 | 228 | #if defined(WIN32)    defined(__FreeBSD__)    defined(__NetBSD__)    defined(__OpenBSD__)    defined(__CYGWIN__) |                                     |
| 229 | 229 | // Simpler version for Windows and *BSD                                                                          |                                     |
| 230 | -   | Rcpp::List trace = Rcpp::List::create(                                                                           |                                     |
| 231 | -   | Rcpp::Named( "file" ) = file,                                                                                    |                                     |
| 232 | -   | Rcpp::Named( "line" ) = line,                                                                                    |                                     |
| 233 | -   | Rcpp::Named( "stack" ) = "C++ stack not available on this system" ;                                              |                                     |
| 230 | +   | List trace = List::create(                                                                                       |                                     |
| 231 | +   | _[ "file" ] = file,                                                                                              |                                     |
| 232 | +   | _[ "line" ] = line,                                                                                              |                                     |
| 233 | +   | _[ "stack" ] = "C++ stack not available on this system"                                                          |                                     |
| 234 | +   | ) ;                                                                                                              |                                     |

# GitHub repositories can have *issues*: think “bug tracker”.

The screenshot shows a GitHub repository page for 'RcppCore/Rcpp'. At the top, there's a navigation bar with icons for back, forward, search, and user profile. The URL 'github.com/RcppCore/Rcpp' is visible. Below the navigation is a header with a GitHub icon, a 'This repository' dropdown, a search bar, and links for 'Explore', 'Gist', 'Blog', and 'Help'. A user profile for 'jennybc' is shown with a picture, a 'Watch' button (5), a 'Star' button (2), and a 'Fork' button (1). The main content area has a title 'RcppCore / Rcpp' with a 'Code' icon. Below it is a summary bar with metrics: 2,595 commits, 1 branch, 0 releases, and 6 contributors. A progress bar indicates activity levels. Underneath is a commit history for the 'master' branch, showing three recent commits by 'eddelbuettel': one from two days ago and two from four days ago. To the right of the commit history is a sidebar with links: 'Code' (highlighted with an orange oval), 'Issues' (with a count of 47), 'Pull Requests' (0), 'Wiki', 'Pulse', and 'Graphs'. The 'Issues' link is specifically circled in orange.

# GitHub repositories can have *issues*: think “bug tracker”.

RcppCore / Rcpp

Watch 5 Star 2 Fork 1

Browse Issues Milestones New Issue

Everyone's Issues 47 47 Open 27 Closed Sort: Newest ▾ 1 2

Created by you 0

Mentioning you 0

No milestone selected

Labels

| Label         | Count |
|---------------|-------|
| Testing       | 7     |
| api           | 15    |
| attributes    | 1     |
| bug           | 7     |
| documentation | 4     |
| enhancement   | 2     |
| modules       | 3     |
| question      | 1     |
| sugar         | 11    |
| duplicate     | 0     |

const ness problem with sapply bug api #74  
Opened by romainfrancois November 27, 2013

checking for interupts api #69  
Opened by romainfrancois November 25, 2013 2 comments

Rcpp 0.10.6.2 dies on unit tests bug #67  
Opened by eddelbuettel November 25, 2013 9 comments

Export `test` and `unit\_test\_setup` Testing #66  
Opened by romainfrancois November 24, 2013

Rcpp breaks updates bug #63  
Opened by eddelbuettel November 23, 2013 4 comments

Rcpp.package.skeleton bug #61  
Opened by romainfrancois November 22, 2013

Convert uses of inline::cxxfunction to attributes. documentation #56  
Opened by romainfrancois November 14, 2013

unquarantine module tests in runit.wrap modules Testing #52  
Opened by romainfrancois November 14, 2013

# How to use issues for a GitHub repository ...

Think outside the “software development” box.

Your repository could be for writing a manuscript or performing an analysis.

An issue is a great way to keep “to do’s” on your radar, gather notes and links and brainstorm with collaborators.

Examples: track down a citation for <x>, re-develop the notation for model <y>, create a webpage with the results of simulation study <z>, etc.

Close an issue = cross something off your to do list  
(feels good!)

If a commit message says “fixes #72”, GitHub closes issue #72 and establishes a link between the issue (the problem) and the commit (the fix).

**mark functions as registered and make sure they don't throw. closes #73**

Loading branch information...

 **romainfrancois** authored November 26, 2013

1 parent [d1bc779](#) commit [3331e7b](#)48934936aa8c4782ba6ec06714da7f670

 Showing 2 changed files with 29 additions and 11 deletions.

[Show Diff Stats](#)

 19 src/api.cpp [View file @ 3331e7b](#)

... ... @@ -227,10 +227,11 @@ SEXP stack\_trace( const char\* file, int line ){  
227 227 #if defined(\_\_GNUC\_\_)

 **romainfrancois** closed this issue from a commit

 **romainfrancois** mark functions as registered and make sure they don't throw. closes #73 [3331e7b](#)

**Closed**  **romainfrancois** closed the issue in 3331e7b November 26, 2013

[Browse code](#)

November 26, 2013

# Markdown files are automatically rendered nicely in GitHub repositories

Example: `links.md` in the repo I created for today!

The screenshot shows a GitHub repository page for `jennybc / 2013-11_sfu`. The repository has 2 stars and 0 forks. The current branch is `master`, and the file `links.md` is being viewed. The file was last updated 6 hours ago by `jennybc`, with notes about the RPubs SSL certificate fiasco and how to solve it on Windows. There is 1 contributor listed. The file itself contains 47 lines (31 sloc) and is 3.786 kb in size. The file content is rendered as follows:

# Links

Hadley Wickham's talk in the [Simply Statistics Unconference on the Future of Statistics](#)

Daring Fireball's [markdown page](#). Kind of where it all begins, but other references that are more recent and specific to our context are more relevant.

Carson Sievert's talk [Reproducible web documents with R, knitr & Markdown](#)

[MathJax](#) is an open source JavaScript display engine for mathematics that works in all browsers ... It just works.

You can see the raw Markdown too!

jennybc / 2013-11\_sfu

branch: master 2013-11\_sfu / links.md

jennybc 6 hours ago Notes about the RPubs SSL certificate fiasco and how to solve on Windows

1 contributor

file | 47 lines (31 sloc) | 3.786 kb

Open Edit Raw Blame History Delete

# Links

Hadley Wickham's talk in the :)

Daring Fireball's markdown page are more relevant.

Carson Sievert's talk Reproducible

MathJax is an open source Ja

https://raw.githubusercontent.com/jennybc/2013-11\_sfu/master/links.md

faculty.washington.edu/fredrik  
https://raw.githubusercontent.com/fredrikhult/STAT545A/master/links.md  
STAT545A/homeworks/links.md  
karthik/smb\_git

Links

[Hadley Wickham's talk](https://dl.dropboxusercontent.com/u/41902/future-of-statistics.pdf)  
the [Simply Statistics Unconference on the Future of Statistics]  
(http://simplystatistics.org/unconference/)

Daring Fireball's [markdown page](http://daringfireball.net/projects/markdown/)  
it all begins, but other references that are more recent and specific to R relevant.

Carson Sievert's talk [Reproducible web documents with R, knitr & Markdown](http://cpsievert.github.io/slides/markdown/)

# Comma (.csv) and tab (.tsv) delimited files are automatically rendered nicely in GitHub repositories

## Example: Lord of the Rings data I found for STAT 545A

jennybc / **lotr**

branch: master · **lotr / lotr\_clean.tsv** ·

 **jennybc** 2 months ago Add early exploration/cleaning

1 contributor

file | 684 lines (683 sloc) | 42.64 kb

Open Edit Raw Blame History Delete

Search this file...

|   | Film                       | Chapter                | Character   | Race   | Words |
|---|----------------------------|------------------------|-------------|--------|-------|
| 1 | The Fellowship Of The Ring | 01: Prologue           | Bilbo       | Hobbit | 4     |
| 2 | The Fellowship Of The Ring | 01: Prologue           | Elrond      | Elf    | 5     |
| 3 | The Fellowship Of The Ring | 01: Prologue           | Galadriel   | Elf    | 460   |
| 4 | The Fellowship Of The Ring | 02: Concerning Hobbits | Bilbo       | Hobbit | 214   |
| 5 | The Fellowship Of The Ring | 03: The Shire          | Bilbo       | Hobbit | 70    |
| 6 | The Fellowship Of The Ring | 03: The Shire          | Frodo       | Hobbit | 128   |
| 7 | The Fellowship Of The Ring | 03: The Shire          | Gandalf     | Wizard | 197   |
| 8 | The Fellowship Of The Ring | 03: The Shire          | Hobbit Kids | Hobbit | 10    |

# Use case: course materials and website, 3 profs, 2 TAs

The screenshot shows a GitHub repository page for `jennybc / stat540`. The repository is marked as **PRIVATE** and has a lock icon. It contains 493 commits, 2 branches, and 0 releases. The code editor shows a file named `faculty.was...` with content related to STAT540. A screenshot of a password-protected webpage titled "STAT540 Statistical methods for high dimensional biology (3 credits)" is shown, featuring course information, instructors, TAs, Google Group, time and location, prerequisites, evaluation, and syllabus. An orange arrow points from the repository details to the course webpage.

private repository contains source  
for password protected webpage  
on UBC Stat server

STAT540 Statistical methods for high dimensional biology (3 credits)

- Course Information
  - Instructors
  - TA(s)
- Google Group for Q & A (TAs will add students to group in due course)
- Time and Location
  - Lecture (Sec 201)
  - Seminar / computing lab (S2A) -- REGISTRATION IS REQUIRED!
- Prerequisites.
- Evaluation
- Syllabus
  - Week 1
  - Week 2
  - Week 3
  - Week 4
  - Week 5

# Use case: course materials and website, 1 prof

The screenshot shows a GitHub repository page for [jennybc / STAT545A](#). The repository is public, has 388 commits, 1 branch, and 0 releases. A progress bar indicates 'Fetching contributors'. The master branch is selected. The repository description states: 'Small fixes/additions since the course meetings ended.' A commit by jennybc from November 19, 2013, is listed under '.gitignore' with the note 'FINALLY creating Makefile to compile webpage'. A blue arrow points from this commit to the course website link in the browser tab.

UBC grad course in data analysis with R <http://www.stat.ubc.ca/~jenny/STAT545A/current.html> — Edit

388 commits 1 branch 0 releases Fetching contributors

branch: master STAT545A /

Small fixes/additions since the course meetings ended.

jennybc authored November 19, 2013

- 2012-lectures Link to 2013 material from the 2012 index
- 2013-lectures Add PDF slides and link for cm11 on coding
- courseAdmin Creating bullet list with placeholders to stop 1
- hw06\_scaffolds Fix links in README
- .gitignore FINALLY creating Makefile to compile webpage
- LICENSE Adding a license
- LICENSE.md Adding a license

Code Issues

www.stat.ubc.ca/~jenny/STAT545A/current.html

STAT 545A Exploratory Data Analysis

- cm 01 | Wednesday Sept 04 | Introduction to the course (slides as PDF)
- cm 02 | Monday Sept 09 | Create first report, Deep Thoughts, Basic care and feeding of data in R (slides as PDF)
- cm 03 | Wednesday Sept 11 | R objects (beyond data.frames) and R Markdown (slides as PDF)
- cm 04 | Monday Sept 16 | Data aggregation (slides as PDF)
- cm 05 | Monday Sept 23 | Explore a quantitative variable, visuals via lattice (slides as PDF)
- cm 06 | Wednesday Sept 25 | Explore a quantitative variable, visuals via lattice .cont'd (slides as PDF)
- cm 07 | Monday Sept 30 | Two quantitative variables + lattice .details + writing figures to file (slides as PDF)
- cm 08 | Wednesday Oct 02 | ggplot2
- cm 09 | Monday Oct 07 | Colors
- cm 10 | Wednesday Oct 09 | More ggplot2
- cm 11 | Wednesday Oct 16 | Coding style, project organization, version control (slides as PDF)

Go back to [STAT545A home](#)

public repository contains  
source for open webpage  
on UBC Stat server

## STAT 545A Exploratory Data Analysis

1.5 credits

04 September 2013 - 16 October 2013

Mon Wed 9:30 - 11am in ESB 1042, a computing lab on the main ground floor of the [Earth Sciences Building \(ESB\)](#) at

2207 Main Mall Instructor: Jennifer (Jenny) Bryan [jenny@stat.ubc.ca](mailto:jenny@stat.ubc.ca)

TA: Song Cai [scai@stat.ubc.ca](mailto:scai@stat.ubc.ca)

Google Group for Q & A: [STAT545A\\_2013](#)

github repository for course materials: <https://github.com/jennybc/STAT545A>

cm = class meeting

Monday Sept 02 is a statutory holiday. No class.

[cm 01 | Wednesday Sept 04 | Introduction to the course \(slides as PDF\)](#)

# Use case: source for website about why/how to use Git(Hub)

The screenshot shows a GitHub repository page for 'kbroman/github\_tutorial'. The repository has 97 commits, 1 branch, and 0 releases. It is currently fetching contributors. The branch dropdown shows 'branch: gh-pages'. The sidebar lists files like '\_includes', '\_layouts', '\_plugins', 'assets', 'pages', '.gitignore', 'License.md', 'README.md', 'Rakefile', '\_config.yml', 'favicon.ico', and 'index.md'. The main content area displays the 'git/github guide' page.

[http://kbroman.github.io/github\\_tutorial/](http://kbroman.github.io/github_tutorial/)

The screenshot shows the 'git/github guide' website at kbroman.github.io/github\_tutorial/. The page title is 'git/github guide' and it is described as 'Karl's minimal tutorial'. The content includes a blurb about the importance of git and github, a statement about providing a minimal guide, and a bulleted list of topics: Why git and github?, Your first time, Typical use, Start a new repository, Contribute to someone's repository, Handling merge conflicts, Oops; that last commit message was wrong, and Exploring the code and its history.

Files in a Git repo, even one hosted on GitHub, still reside on your computer

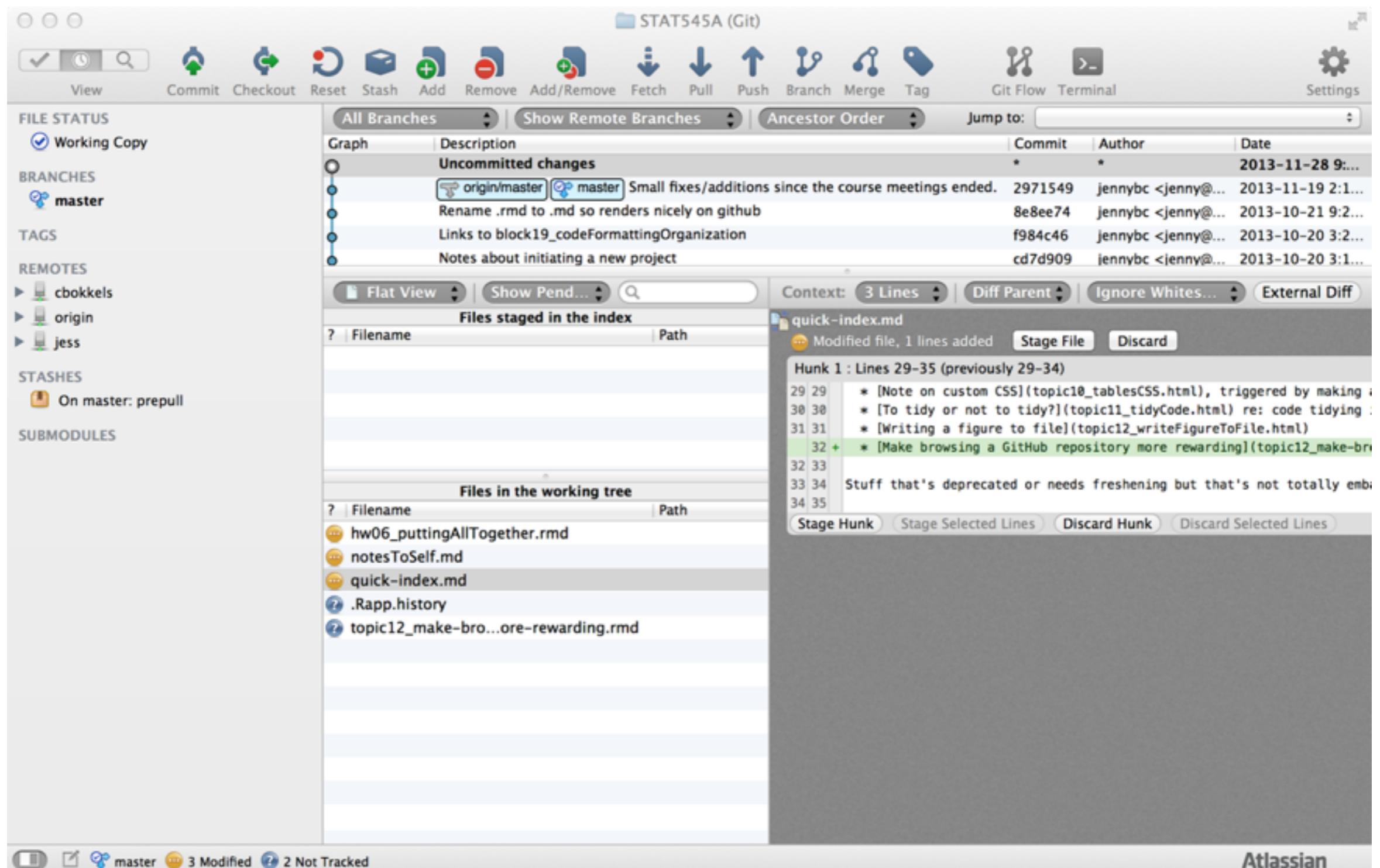
Browse and edit them all you want

Git has commands for communicating with the remote repository, e.g. the GitHub repo (push, pull, fetch, clone)

I highly recommend using a Git GUI on your computer for making commits, syncing with the remote, etc.

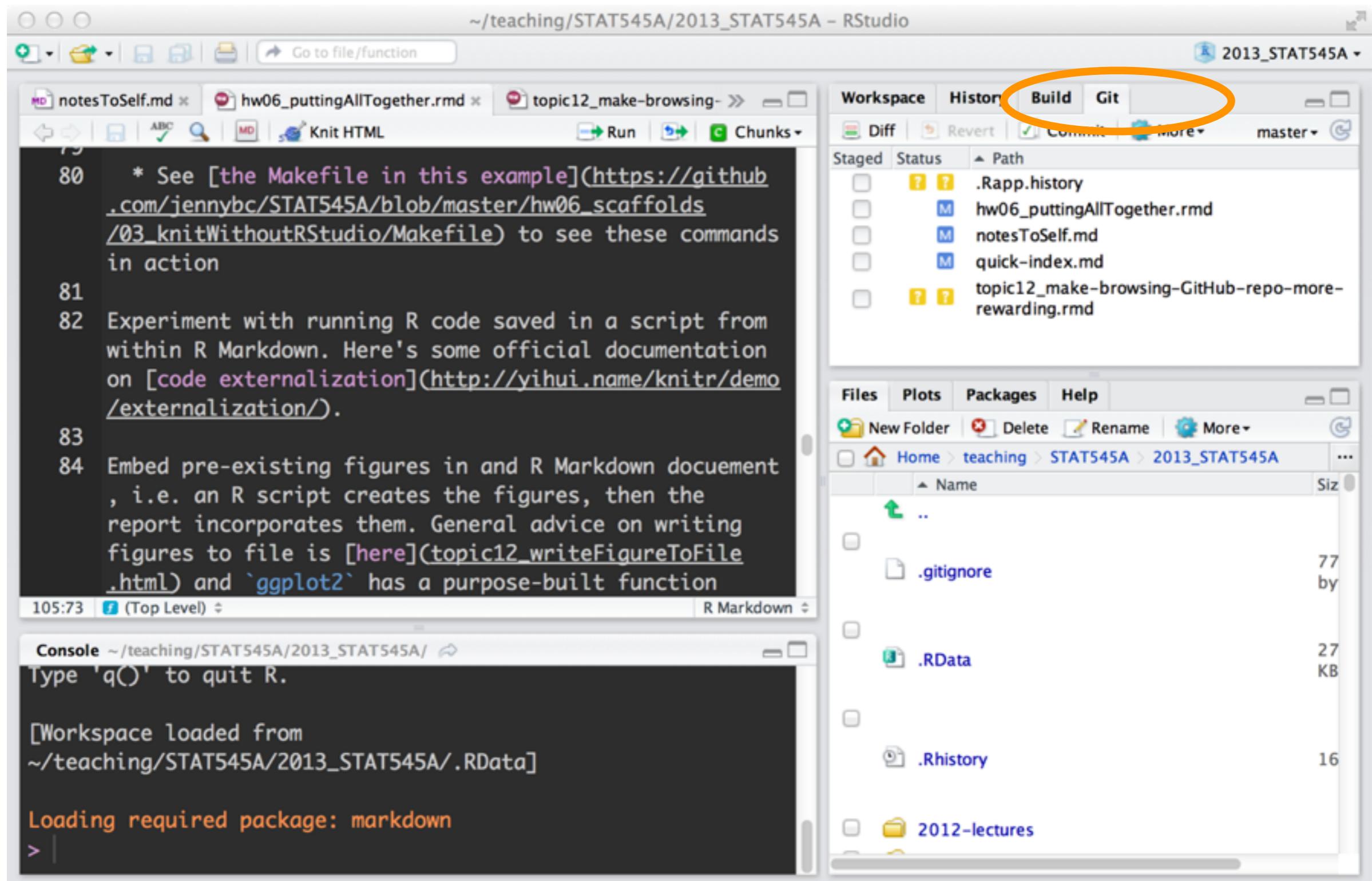
Reconciling and merging changes when two people make conflicting commits is not fun, but better than the alternatives

# I recommend SourceTree, a free Git client for Windows and Mac.



# RStudio can also act as your Git(Hub) client

[http://www.rstudio.com/ide/docs/version\\_control/overview](http://www.rstudio.com/ide/docs/version_control/overview)



Big picture, second half:

sane file and project management is good  
that's what version control does

distributed file management is good  
excellent for 2+ people collaborating

ability to browse something on the web is unreasonably  
powerful

Git + GitHub provide a compelling solution for  
collaborative file wrangling; (R) Markdown and RStudio  
play well with Git(Hub)

**Bottom line: do something deliberate that has a good hassle: result ratio for you.**

**Be open to upgrading your approach as time goes on.**

**Keep your eyes and ears open re: new developments.**

project organization / literate programming /  
reproducible research

knitr

R markdown

R packages

RStudio

version control / back up / archive

Know you what the fuss is about!

collaboration / open science

github

Rforge

sourceforge

git

subversion

mercurial

