STAT 545A
Class meeting #3
Wednesday, September 11, 2013

Dr. Jennifer (Jenny) Bryan

Department of Statistics and Michael Smith Laboratories

any questions from tutorial re: Basic care and feeding of data in R?

HW #1 “Dry Run”. Due as soon as possible. Publish 1 page of something -- practically anything -- to Rpubs, then share that link w/ rest of the class. Instructions/”assignment” is posted but will get some editing. Content not important.

HW #2 “Dry Run 2 + Hello Gapminder”. Due 9:30am Monday Sept. 16. Perform basic data intake and smell testing for the Gapminder data. Write it up via R Markdown. Compile, publish, share. Content can be drawn from tutorials, your job is to pull together. Instructions coming.

today

data analysis vs programming vs developing

combo lecture and hands on computing:
R objects, beyond the data.frame
students will complete on their own

R Markdown

three different modes of working with R:

data analysis

programming

developing

three different modes of working with R:

data analysis ← I am mostly trying to prepare you for this.

programming

developing

three different modes of working with R:

data analysis

programming ← But we need to talk .... about some things more related to this.

developing

hands-on computing phase here

JB's typology of R objects, i.e. "flavors"

a simple view of simple R objects that will get you pretty far

Simple view	Technically correct R view		
	mode	class	typeof
character	character	character	character
logical	logical	logical	logical
numeric	numeric	integer or numeric	integer or double
factor	numeric	factor	integer

JB's typology of R objects, i.e. "flavors"

a simple view of simple R objects that will get you pretty far

Simple view	Technically correct R view		
	mode	class	typeof
character	character	character	character
logical	logical	logical	logical
numeric	numeric	integer or numeric	integer or double
factor	numeric	factor	integer

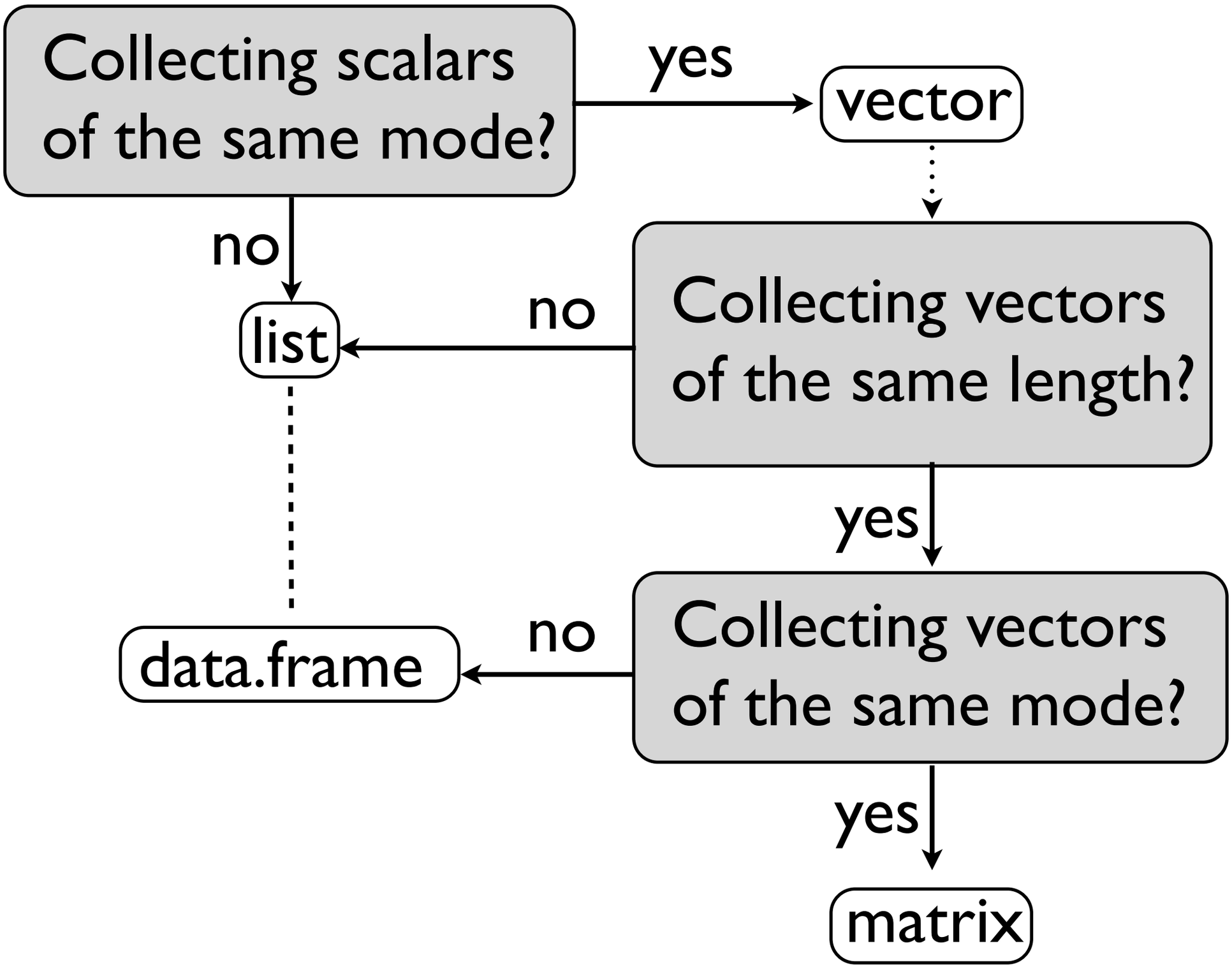
JB's typology of R objects, i.e. "flavors"

a simple view of simple R objects that will get you pretty far

Simple view	Technically correct R view		
	mode	class	typeof
character	character	character	character
logical	logical	logical	logical
numeric	numeric	integer or numeric	integer or double
factor	numeric	factor	integer

We will spend some time talking about factors in a future class.

"Simple view" of data collections



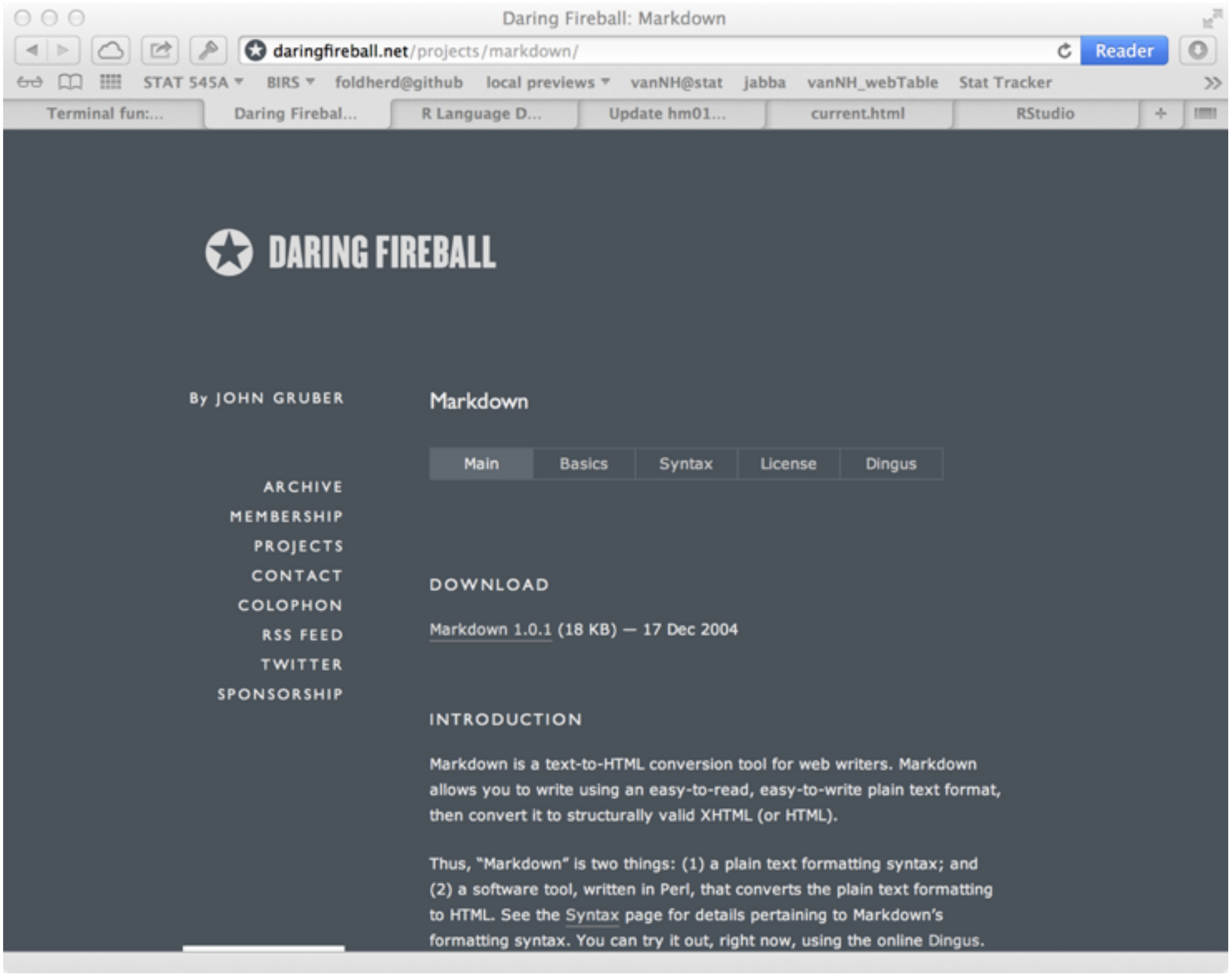
weak links in the chain: process, packaging and presentation

1. **Build your communication skills.** Unless you live in a plastic bubble, you are going to need to work with other people. You will be given tasks by other people, collaborate with other people to achieve those tasks, and ultimately have to report the results of your work to other people. You need to be able to speak clearly and concisely, listen carefully, write well (and quickly), and give informative and interesting presentations. Contrary to popular belief *the person who learns to do these things well will advance farther than someone who has better technical capabilities but poor communication skills!* Management usually can't tell the difference between a good statistician and a great one, but they can see immediately who communicates their results well and who does so poorly. Unfortunately, most university environments stress working alone and in isolation, completely the opposite of what life will be like on the other side of graduation.

   You need to take actions to ensure that your communication skills are sharp. These actions can include: (i) Taking a writing class, especially one that stresses technical writing, which has a completely different flavour from essay writing; (ii) Taking a class in verbal communication, and in particular one that covers the fine art of making and delivering presentations; (iii) Taking business courses, especially those in business communication and organizational structure and behaviour, so that you can better understand your audience and learn to arrange your communications accordingly; (iv) Seeking out courses that expressly advertise group project work and/or presentations, even (*especially!*) if these things scare you. All of our speakers indicate that anything that you can do to practice your communication skills will have a positive effect on your employability and advancement.

Excerpt from "Real Advice from Real People" by Tom Loughin, Statistical Society of Canada Liaison, Vol. 22 No. 4 (November 2008).

http://daringfireball.net/projects/markdown/

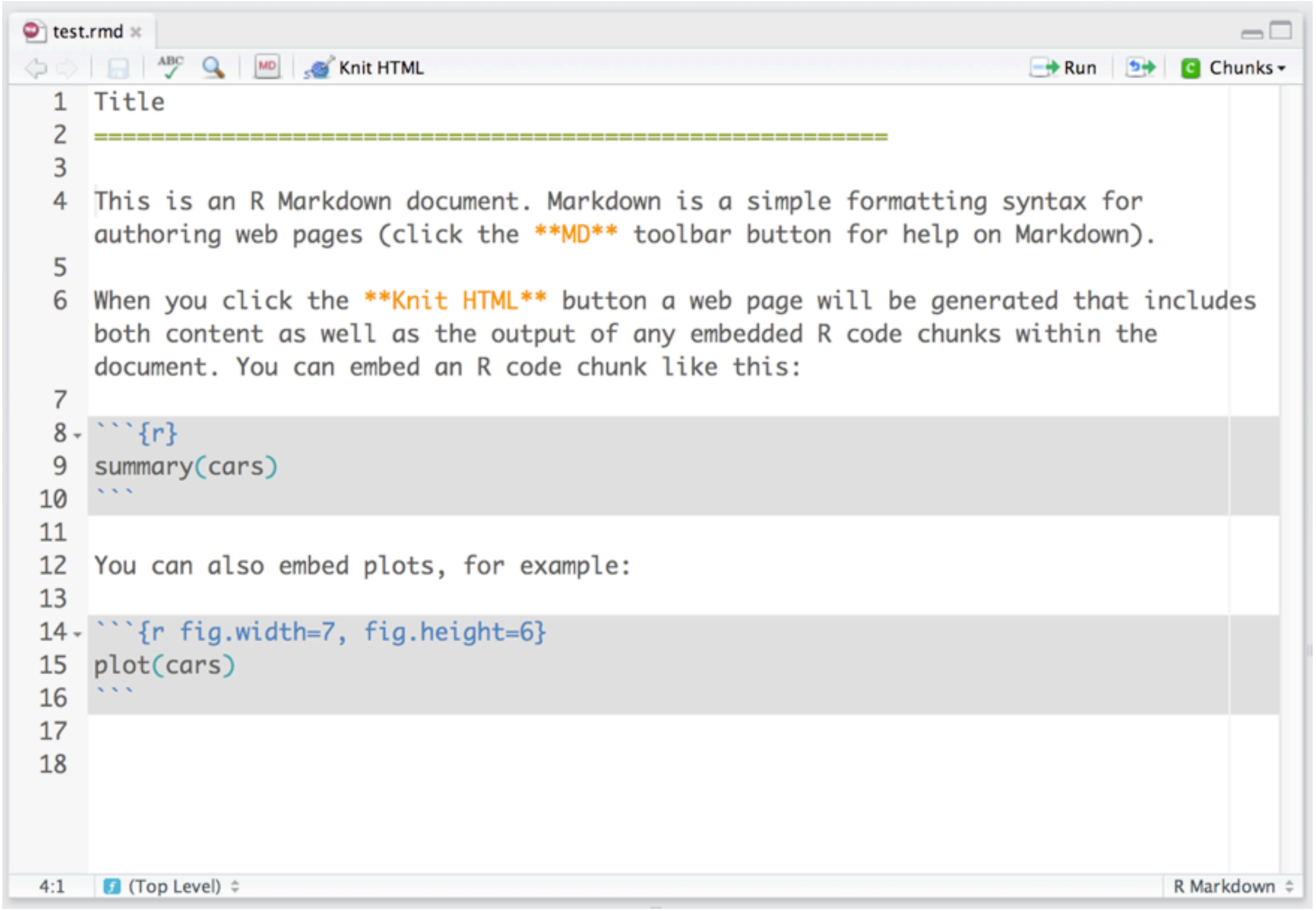


http://www.rstudio.com/ide/docs/r_markdown

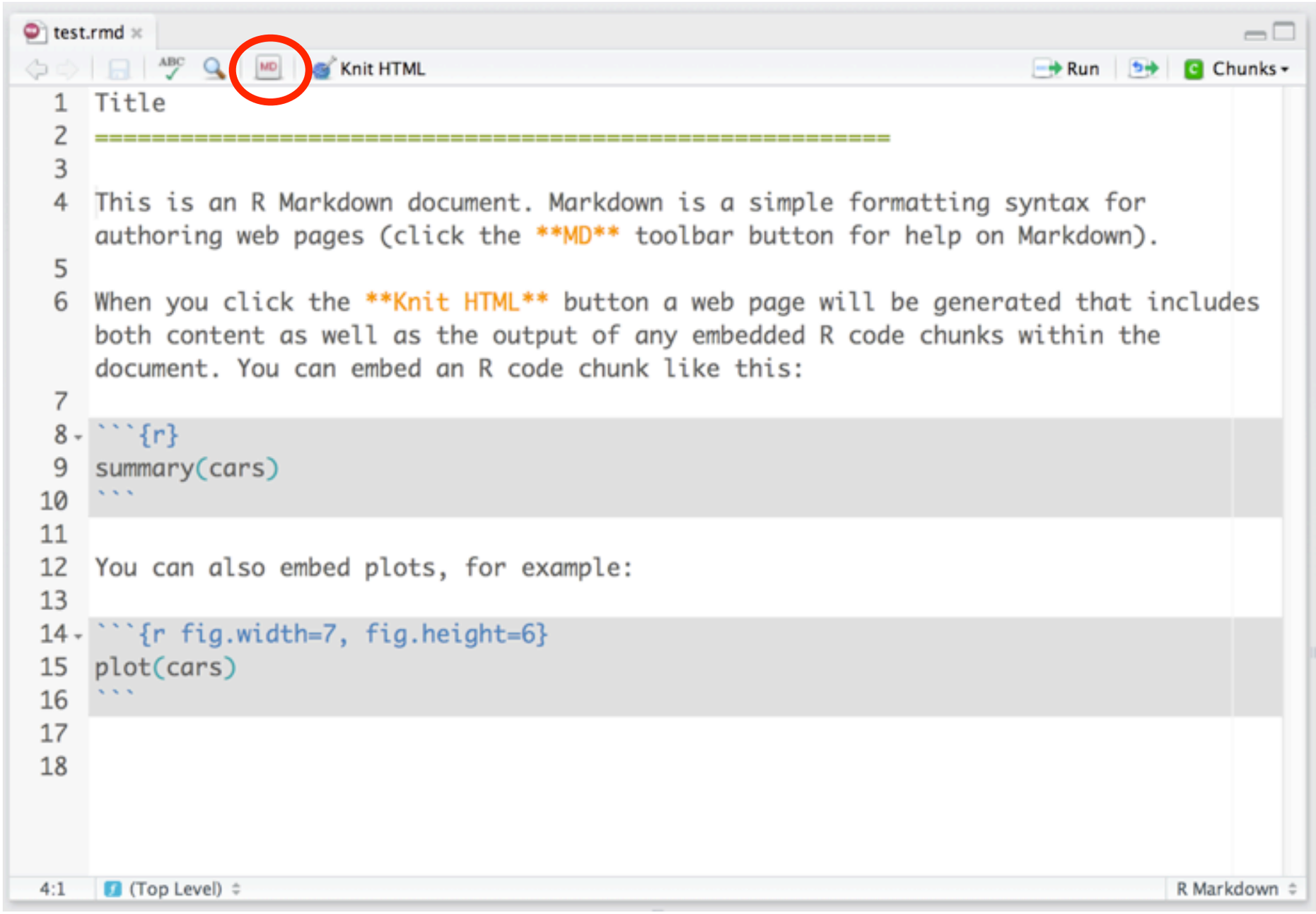
http://www.rstudio.com/ide/docs/authoring/using_markdown

http://yihui.name/knitr/

File > New > R Markdown
Save As ... test.rmd

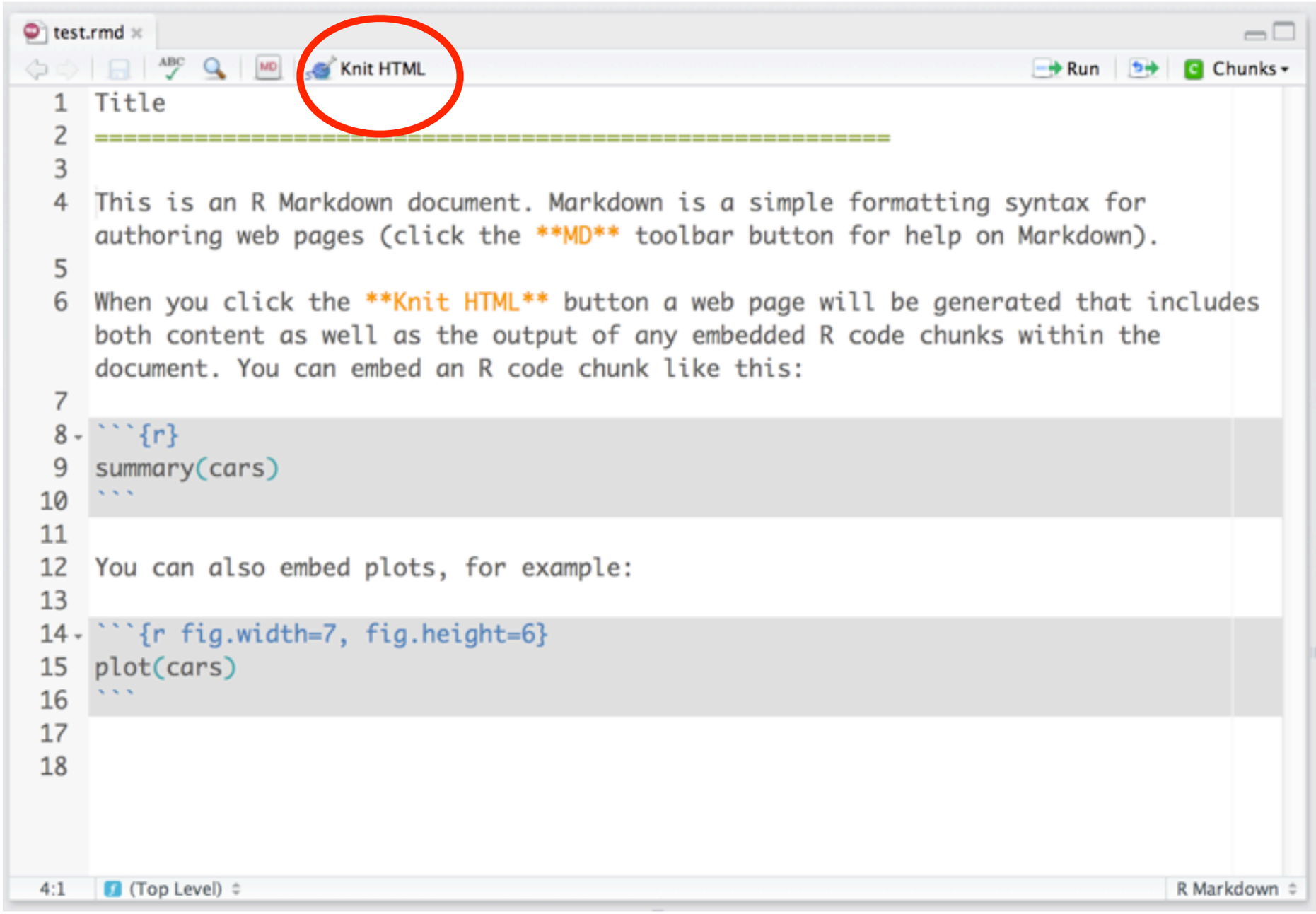


Displays Markdown Quick Reference in the Help tab/pane



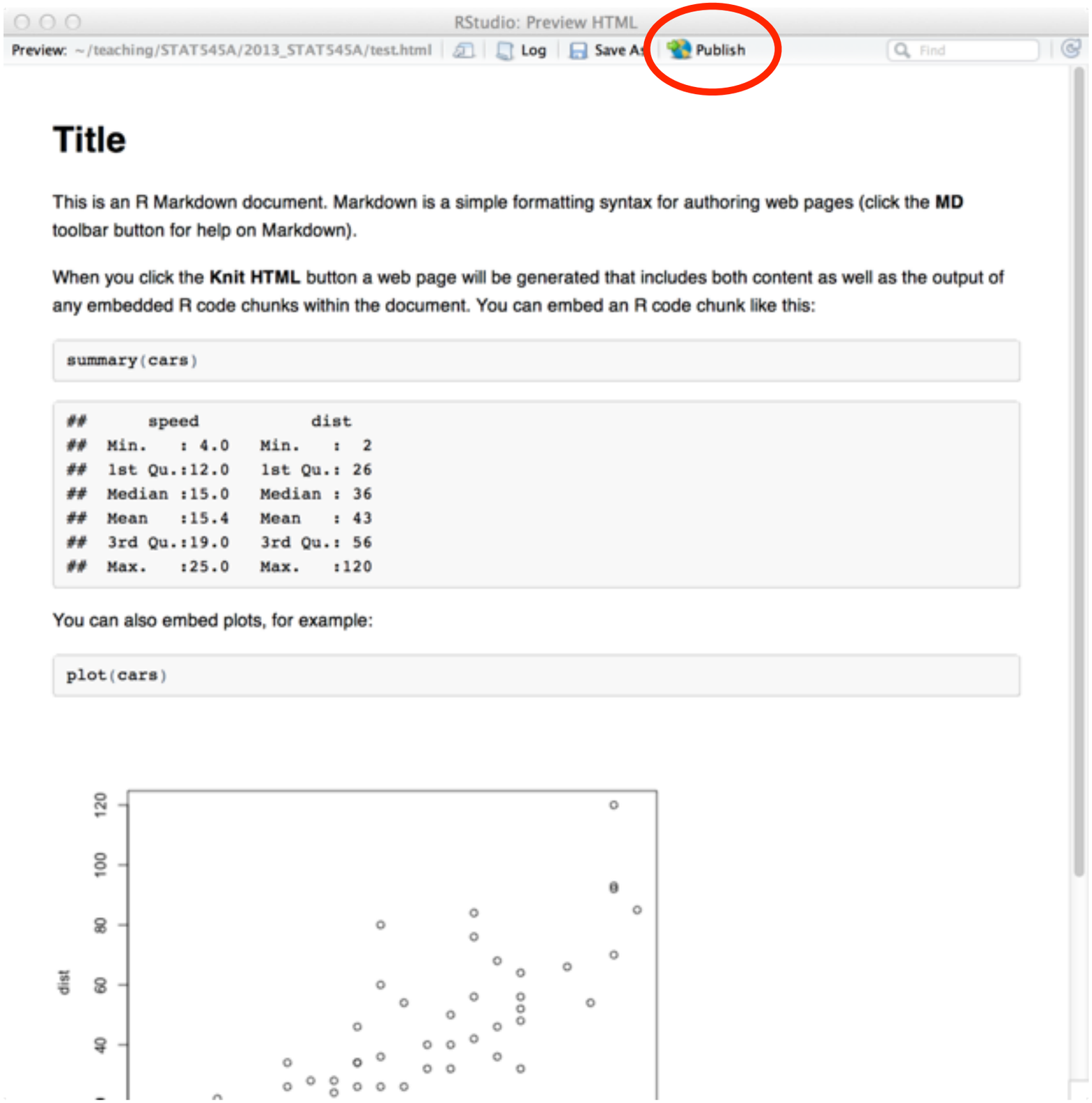
File > New > R Markdown

Save As ... test.rmd



Knit HTML.

Yes this can be accomplished outside of RStudio, using knitr functions at the command line, so we are not creating unhealthy dependency on RStudio.



This is where you'll need that RPubs account.

What's actually happening?

You author this. You've seen it already.

knit() function from knitr package does R Markdown to Markdown conversion

R chunks are run and replaced by their output.
Yes that's oversimplified but you get the idea.

test.rmd

````
Title
=============================================================

This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the **MD**
toolbar button for help on Markdown).

When you click the **Knit HTML** button a web page will be
generated that includes both content as well as the output
of any embedded R code chunks within the document. You can
embed an R code chunk like this:

```{r}
summary(cars)
```

You can also embed plots, for example:

```{r fig.width=7, fig.height=6}
plot(cars)
```
````

test.md

````
Title
=============================================================

This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the **MD**
toolbar button for help on Markdown).

When you click the **Knit HTML** button a web page will be
generated that includes both content as well as the output
of any embedded R code chunks within the document. You can
embed an R code chunk like this:


```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```


You can also embed plots, for example:


```r
plot(cars)
```

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

markdownToHTML() function from markdown package does Markdown to HTML conversion

Pandoc, the Swiss army knife of document conversion, is another option for Markdown to HTML conversion and, importantly, for Markdown to PDF conversion.

You can set things up so that RStudio and knitr use Pandoc for the second conversion step.

test.md

````
Title
=====================================================

This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the **MD**
toolbar button for help on Markdown).

When you click the **Knit HTML** button a web page will be
generated that includes both content as well as the output
of any embedded R code chunks within the document. You can
embed an R code chunk like this:


```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```


You can also embed plots, for example:


```r
plot(cars)
```

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

test.html

```
                page-break-inside: avoid;
        }
        pre {
                word-wrap: break-word;
        }
}
</style>



</head>

<body>
<h1>Title</h1>

<p>This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the
<strong>MD</strong> toolbar button for help on Markdown).</
p>

<p>When you click the <strong>Knit HTML</strong> button a
web page will be generated that includes both content as
well as the output of any embedded R code chunks within the
document. You can embed an R code chunk like this:</p>

<pre><code class="r">summary(cars)
</code></pre>

<pre><code>##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
</code></pre>

<p>You can also embed plots, for example:</p>

<pre><code class="r">plot(cars)
</code></pre>

<p><img src="
+AAAEJGlDQ1BJQ0MgUHJvZmlsZQAAOBGFVd9v21QUPolvUqQWPyBYR4eKxa9
VU1u5GxqtxgZJk6XtShal6dgqJOQ6N4mpGwfb6baqT3uBNwb8AUDZAw9IPCE
NBmJ72fbAtElThyqqSUh76MQPISbtBVXhu3ZiJ1PEXPX6yznfOec7517bRD1
fabWaGVWIlquunc8klZOnFpSeTYrSs9RLA9Sr6U4tkcvNEi7BFffO6+Edigj
L7ZHu/k72I796i9zRiSJPwG4VHX0Z
+AxRzNRrtksUvwf7+Gm3BtzzHPDTNgQCqwKXfZwSeNHHJz1OIT8JjtAq6xWt
CLwGPLzYZi+3YV8DGMiT4VVuG7oiZpGzrZJhcs/hL49xtzH/Dy6bdfTsXYNY
+5yluWO4D4neK/ZUvok/17X0HPBLsF+vuUlhfwX4j/
rSfAJ4H1H0qZJ9dN7nR19frRTeBt4Fe9FwpwtN
+2p1MXscGLHR9SXrmMgjONd1ZxKzpBeA71b4tNhj6JGoyFNp4GHgwUp9qplf
mnFW5oTdy7NamcwCI49kv6fN5IAHgD
```

Which would you rather read and write?

test.rmd

```
Title
=============================================================

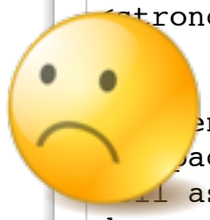
This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the **MD**
toolbar button for help on Markdown).

When you click the **Knit HTML** button a web page will be
generated that includes both content as well as the output
of any embedded R code chunks within the document. You can
embed an R code chunk like this:

```{r}
summary(cars)
```

You can also embed plots, for example:

```{r fig.width=7, fig.height=6}
plot(cars)
```
```



test.html

```
                page-break-inside: avoid;
        }
        pre {
                word-wrap: break-word;
        }
}
</style>



</head>

<body>
<h1>Title</h1>

<p>This is an R Markdown document. Markdown is a simple
formatting syntax for authoring web pages (click the
<strong>MD</strong> toolbar button for help on Markdown).</

<p>When you click the <strong>Knit HTML</strong> button a
web page will be generated that includes both content as
well as the output of any embedded R code chunks within the
document. You can embed an R code chunk like this:</p>

<pre><code class="r">summary(cars)
</code></pre>

<pre><code>##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
</code></pre>

<p>You can also embed plots, for example:</p>

<pre><code class="r">plot(cars)
</code></pre>

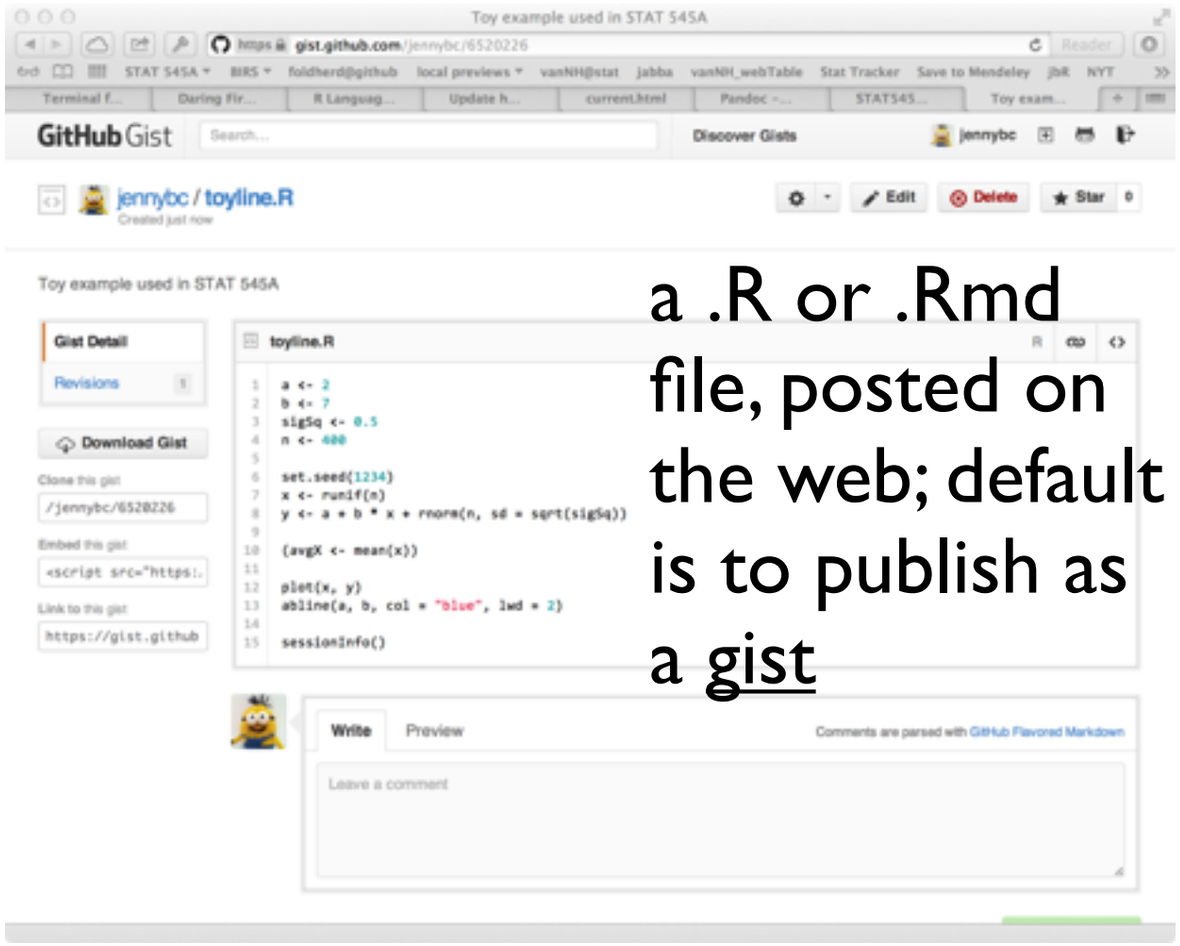
<p><img src="
+AAAEJGlDQ1BJQ0MgUHJvZmlsZQAAOBGFVd9v21QUPolvUqQWPyBYR4eKxa9
VU1u5GxqtxgZJk6XtShal6dgqJOQ6N4mpGwfb6baqT3uBNwb8AUDZAw9IPCE
NBmJ72fbAtElThyqqSUh76MQPISbtBVXhu3ZiJ1PEXPX6yznfOec7517bRD1
fabWaGVWIlquunc8klZOnFpSeTYrSs9RLA9Sr6U4tkcvNEi7BFffO6+Edigj
L7ZHu/k72I796i9zRiSJPwG4VHX0Z
+AxRzNRrtksUvwf7+Gm3BtzzHPDTNgQCqwKXfZwSeNHHJz1OIT8JjtAq6xWt
CLwGPLzYZi+3YV8DGMiT4VVuG7oiZpGzrZJhcs/hL49xtzH/Dy6bdfTsXYNY
+5yluWO4D4neK/ZUvok/17X0HPBLsF+vuUlhfwX4j/
rSfAJ4H1H0qZJ9dN7nR19frRTeBt4Fe9FwpwtN
+2p1MXscGLHR9SXrmMgjONd1ZxKzpBeA71b4tNhj6JGoyFNp4GHgwUp9qplf
mnFW5oTdy7NamcwCI49kv6fN5IAHgD
```

knit2html() function from knitr package does R Markdown to HTML conversion, i.e. hides the two separate steps

RStudio's Knit HTML button does practically the same thing

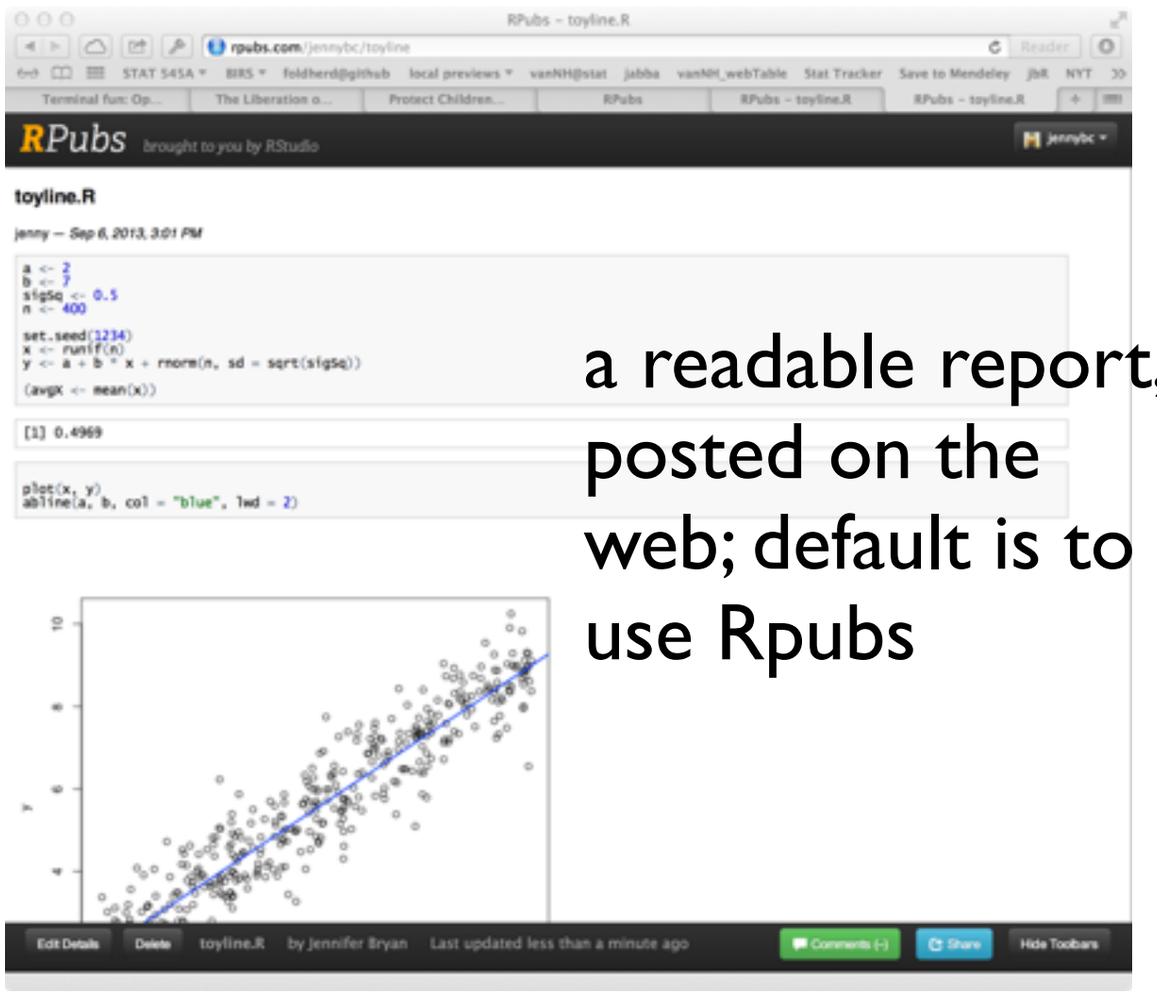
Admire, publish, etc. as you did before with the "notebook" you compiled from a plain .R script.

What a finished piece of coursework will look like:



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

+



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

a line in course repository giving your name and links

=

Instructions for how to submit coursework. Draft exists. Will be revised today.

Homework

I am testing a browser-based workflow for students to "turn in" the homework they have posted on RPubs (or elsewhere).

I'm asking keeners / existing githubbers to try it first. I've done some testing in another repo with a cooperative friend, which yielded these draft instructions. Please help me refine them further before we inflict them on the whole class!

How to add a link to your published homework:

1. Go to github and sign in.
2. Go to the repository for this course: https://github.com/jennybc/STAT545A
3. Go to the Markdown file (e.g. hm01_hw-submission-dryrun.md) you need to edit. Your browser will be pointed at something like `https://github.com/jennybc/STAT545A/blob/master/hm01_hw-submission-dryrun.md` .
4. Click "Edit" to put file into an editing environment.
5. Make your edit. DO NOT GET CREATIVE OR SLOPPY. Follow the instructions. Look around at what everyone else is doing and make your entry follow the pattern, with only the minimal changes needed to customize for *your* homework.
6. At the bottom of the page, fill in the description and click on "Propose file change" button. (Under the hood, this will fork the course repository and create a branch for your change, by default called "patch-1").
7. A page will open that informs you whether there are any merge issues with the change. **Here's where I am interested to see if all hell breaks loose when multiple people edit the same bulleted list.**
8. (Assuming the previous issue is survivable,) click on "Send pull request".
9. Tidy up. **This is where I have more uncertainty.** A student should now have a fork of the course repo, with 2 branches (master and patch-1). Should we help them tidy up? Deleting the patch-1 branch seems very desirable. I think this blog post about tidying up after pull requests deals with what I, the repo owner, can do, not the student / repo forker? I wonder if leaving the student's fork and master branch is good for future edits, but I don't know of any way to make it actually track the course repo, so these steps might need to be repeated every time. Here are more links that look relevant: Fork A Repo and Syncing a fork. I note that the instructions about syncing a fork only show command line git, whereas I want all this happening in a broswer.

Please add your link in this bulleted list:

- Jenny Bryan: toyline script | report
- Matthew Gingerich: TestFile
- Justin Chu: toylineTest