STAT 545A
Class meeting #2
Monday, September 9, 2013

Dr. Jennifer (Jenny) Bryan

Department of Statistics and Michael Smith Laboratories

any questions from first tutorial re: basic use of R via RStudio, RStudio projects?

you do know that R ≠ RStudio, right?

we use RStudio because it makes us happier in our work, but notice that **nothing we produce -- no code, no figures, nothing -- requires RStudio to be created, appreciated or reused**

you should master various ways of sending code from editor to Console so that saving scripts becomes your R Way of Life

today

show how to use RStudio's Compile Notebook ...
Publish to RPubs workflow

registration admin stuff

draw attention to info and "to do's" on course webpage

start next tutorial, to be finished by students on their own before next class

Basic care and feeding of data in R

weak links in the chain: process, packaging and presentation

I assume your toyline.R script looks similar to this.

```
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))

write(avgX, "results/avgX.txt")

pdf("figs/niftyPlot.pdf")
plot(x, y)
abline(a, b, col = "blue", lwd = 2)
dev.off()
```

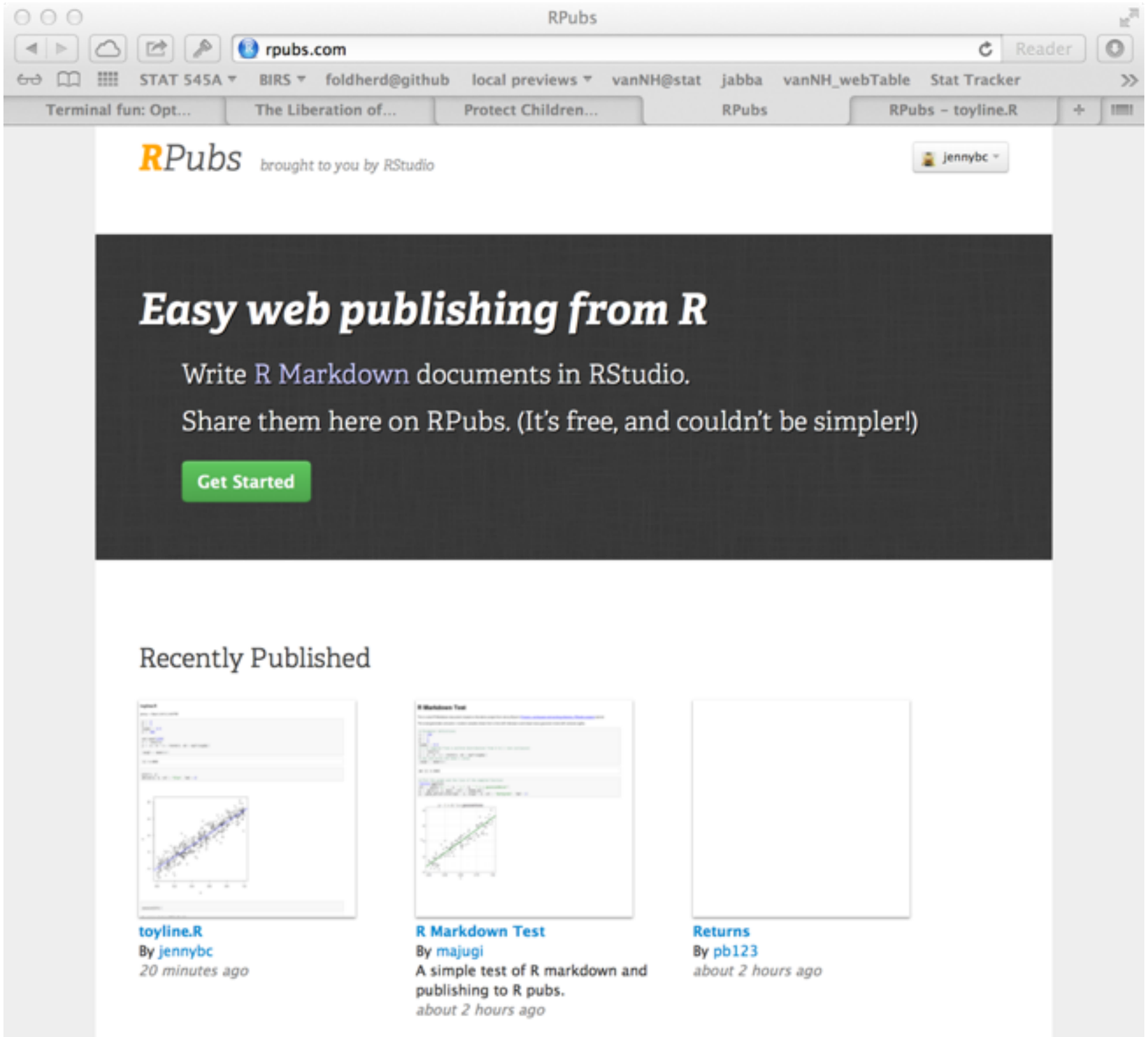
```
a <- 2
b <- 7
sigSq <- 0.5
n <- 400

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

write(avgX, "results/avgX.txt")

pdf("figs/niftyPlot.pdf")
plot(x, y)
abline(a, b, col = "blue", lwd = 2)
dev.off()
```

It's great we are saving important things to file with code -- versus letting them cruise by in the Console and/or saving via mouse clicks -- but we can do better.

You did install `knitr` and its dependencies, as instructed in the set-up tutorial, right?

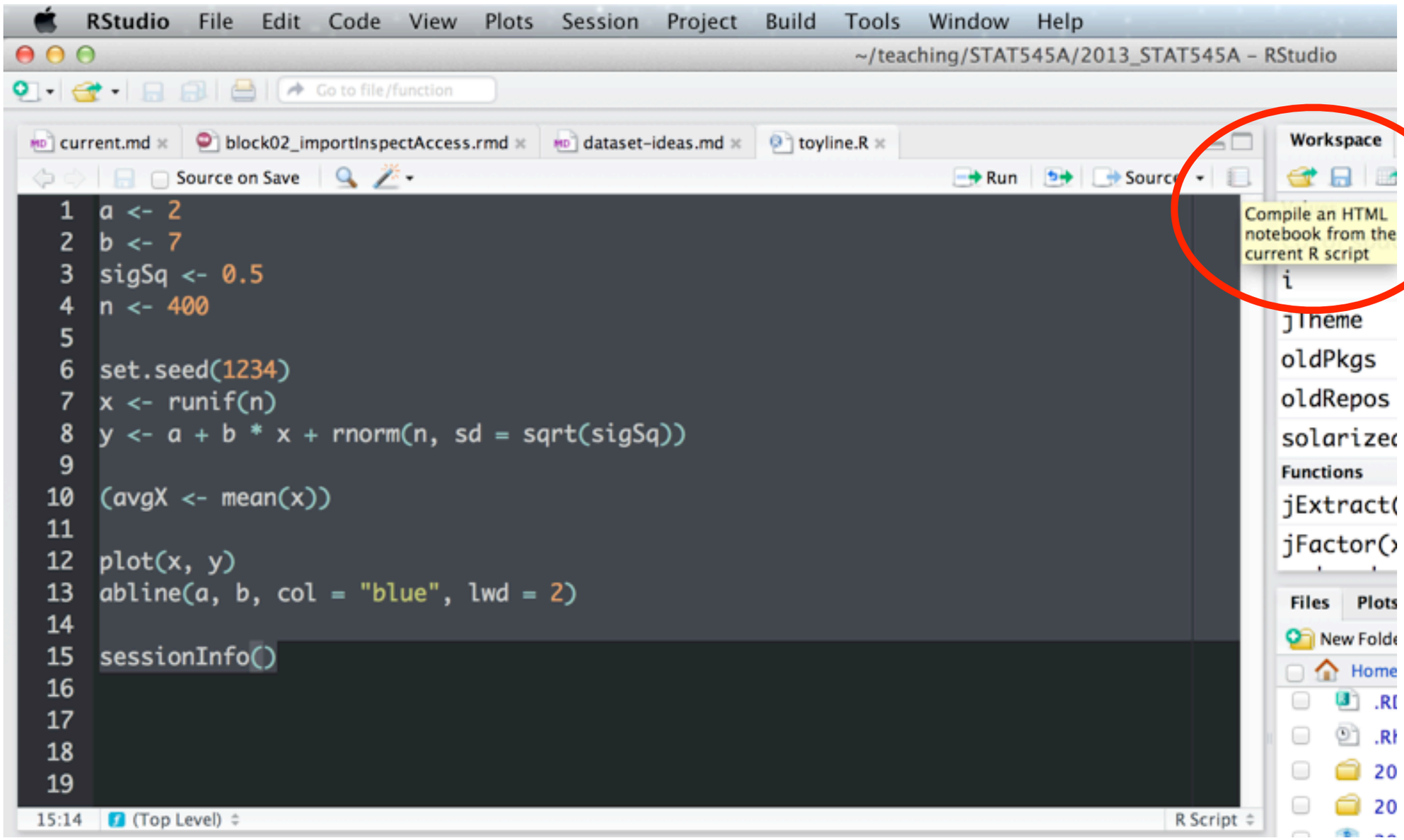
```
install.packages("knitr", dependencies = TRUE)
```

You did get an RPubs account, as requested, right?



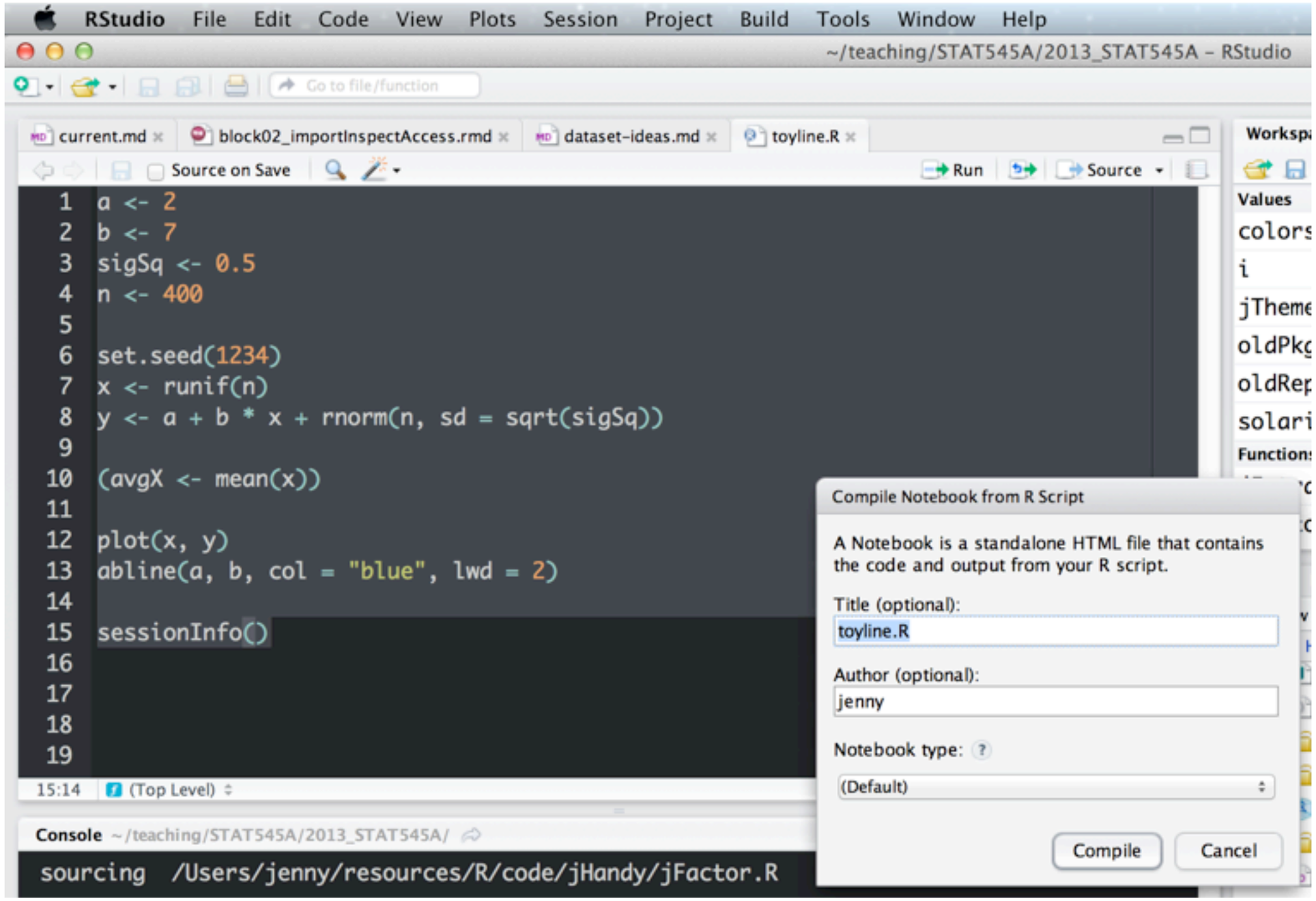
```
a <- 2
b <- 7
sigSq <- 0.5
n <- 400

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()
```

Edit the script -- more like it was during development, when we were watching results and figures appear on the screen.

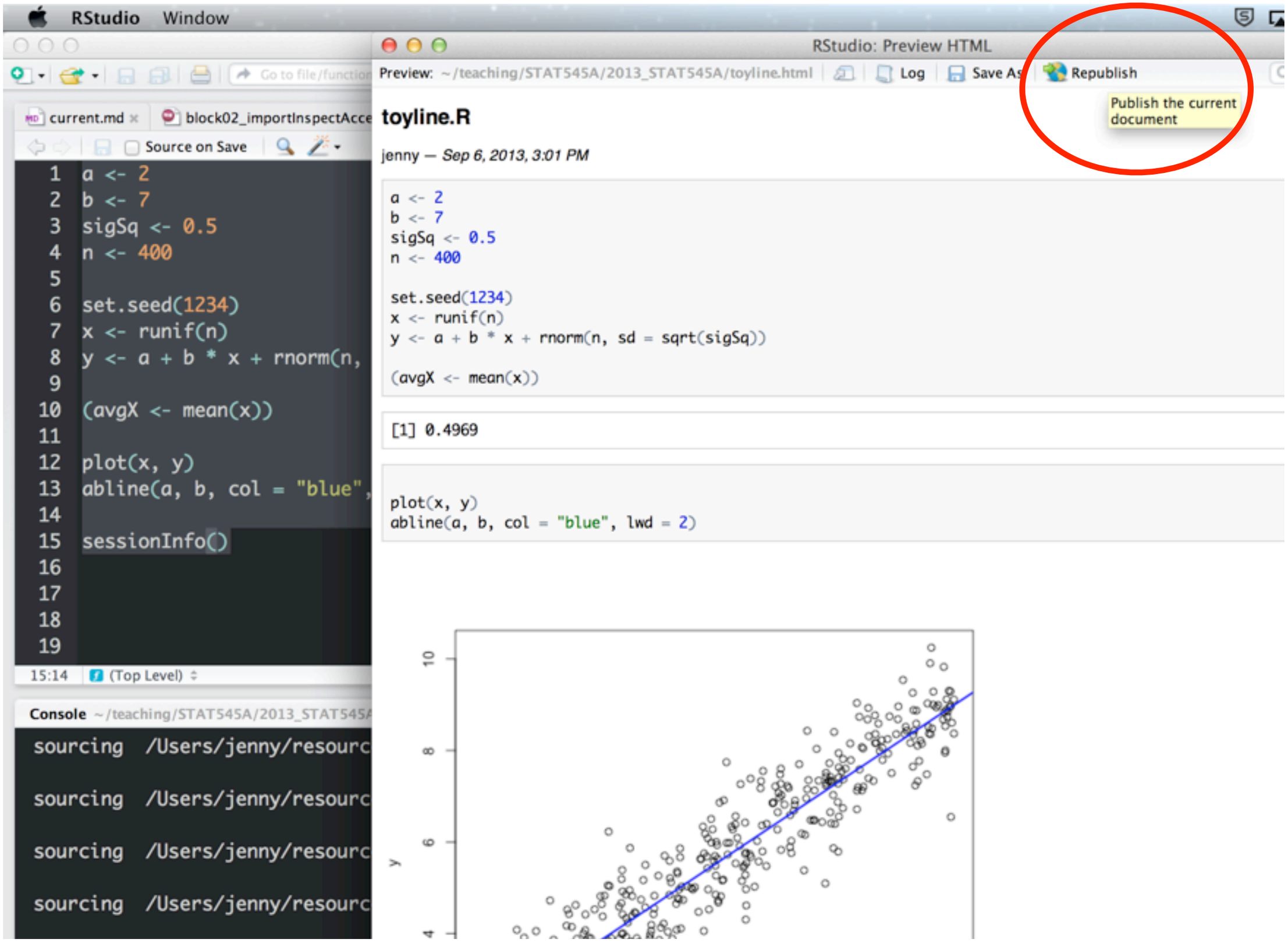


Compile an HTML notebook.

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



I just accept all these defaults.



This is where you'll need that RPubs account.

jenny — *Sep 6, 2013, 3:13 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", l
```

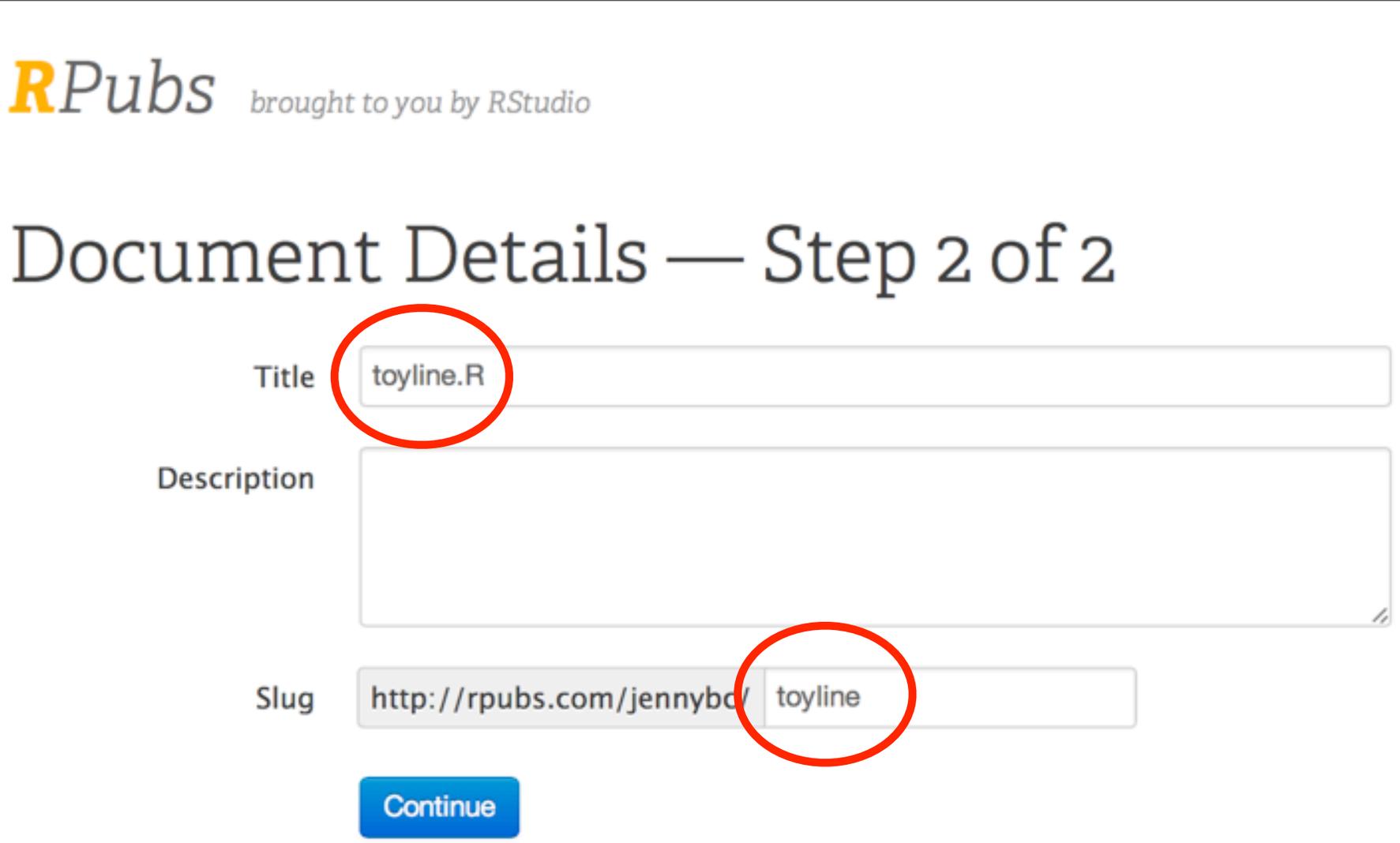
Publish to RPubs

RPubs

RPubs is a free service from RStudio for sharing R Markdown documents on the web. Click Publish to get started.

IMPORTANT: All documents published to RPubs are publicly visible. You should only publish documents that you wish to share publicly.

Publish Cancel



Seems like a good idea to keep script name and slug same, at least as default.

Expect me to give you a naming convention for future STAT 545A coursework.

RPubs – toyline.R

rpubs.com/jennybc/toyline

STAT 545A | BIRS | foldherd@github | local previews | vanNH@stat | jabba | vanNH_webTable | Stat Tracker | Save to Mendeley | jbR | NYT

Terminal fun: Op... | The Liberation o... | Protect Children... | RPubs | RPubs – toyline.R | RPubs – toyline.R

RPubs *brought to you by RStudio* jennybc

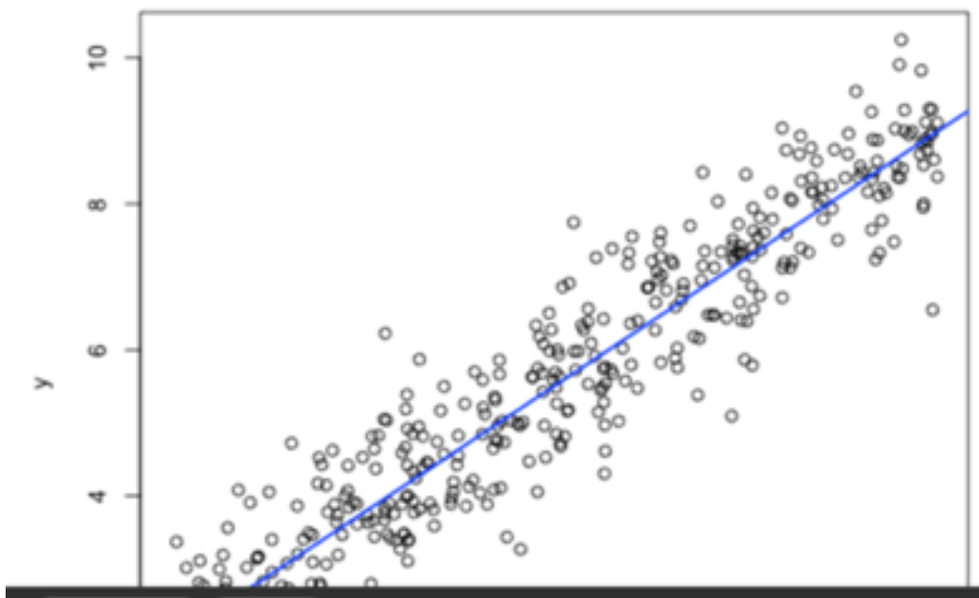
toyline.R

jenny — *Sep 6, 2013, 3:01 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)
```



voilà!

Edit Details | Delete | toyline.R | by Jennifer Bryan | Last updated less than a minute ago | Comments (–) | Share | Hide Toolbars

http://rpubs.com/jennybc/toyline

Let's pilot this workflow for your submission of coursework.

Later, we will take it to the next level of sophistication: authoring in R markdown.

Feel free to start learning about that on your own.

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

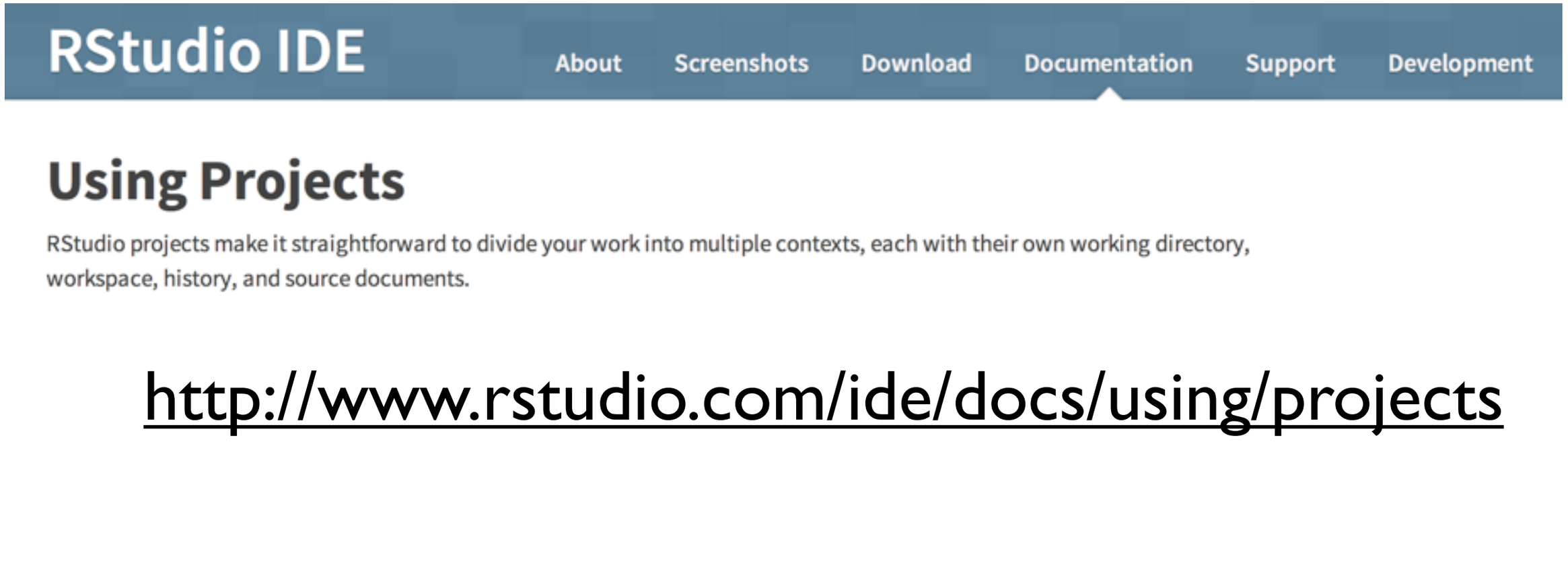
http://www.rstudio.com/ide/docs/authoring/using_markdown

http://yihui.name/knitr/

deep thoughts and pretty pictures to shape your data analytical mindset ....

A place for everything and everything in its place

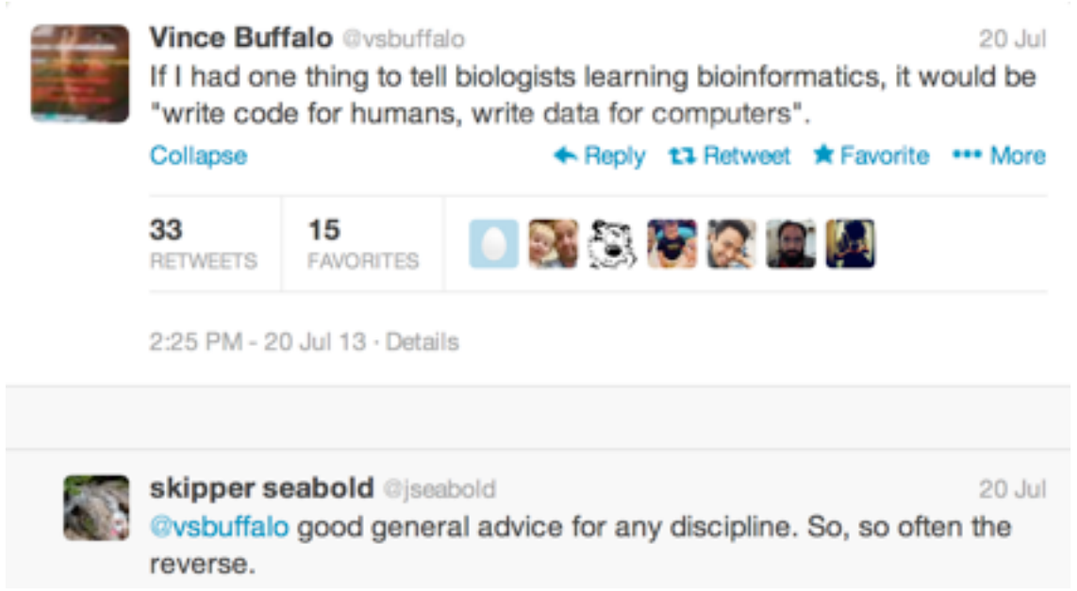
«



http://www.rstudio.com/ide/docs/using/projects

Use RStudio Projects!

write code for humans, write data for computers



> Let us change our traditional attitude to the construction of programs. Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.
>
> **Donald Knuth**

Nine simple ways to make it easier to (re)use your data

Ethan P. White, Elita Baldridge, Zachary T. Brym, Kenneth J. Locey, Daniel J. McGlinn, and Sarah R. Supp

Our nine recommendations are:

this section is especially good reading

1. Share your data
2. Provide metadata
3. Provide an unprocessed form of the data
4. Use standard data formats (including file formats, table structures, and cell contents)
5. Use good null values
6. Make it easy to combine your data with other datasets
7. Perform basic quality control
8. Use an established repository
9. Use an established and liberal license

White et al. (2013) Nine simple ways to make it easier to (re)use your data. *PeerJ PrePrints* 1:e7v2 http://dx.doi.org/10.7287/peerj.preprints.7v2

reshape your data

as in real life, it has a tendency to get short and fat, when you'd really prefer tall and skinny

source is real

PERFECT MATCH: TRIFLE WITH MOSCATO

AT A GLANCE

SERVES 20 PEOPLE

1 HR PREPARATION
50 MIN COOKING (PLUS COOLING, SETTING)

You'll need

1.5 kg	blackberries or mulberries, plus extra to serve (see note)
300 gm	caster sugar
2	vanilla beans, split and seeds scraped
10	gelatine leaves (titanium strength), softened in cold water for 5 minutes
300 ml	pink moscato
1	lemon, juice only
330 ml	crème de mûre (see note)
1.25 kg	crème fraiche
150 ml	milk, or enough to thin
2	lemons, finely grated rind only
40 gm	(¼ cup) pure icing sugar, sifted
	Sponge
8	eggs, at room temperature
250 gm	raw caster sugar
250 gm	plain flour, sieved
50 gm	butter, melted and cooled

Method

1. For sponge, preheat oven to 175C. Whisk eggs and sugar in an electric mixer until tripled in volume (7 minutes). Fold through flour in batches, fold in butter, pour into a 28cm-square cake tin lined with baking paper. Bake until golden and centre springs back when pressed (20-25 minutes). Cool in tin, turn out, halve sponge horizontally, trim each half to fit a 6 litre-capacity glass bowl, then remove from bowl and set aside, reserving trimmings.

2. Meanwhile, combine 1kg berries, sugar, 1 vanilla bean and seeds and 1.1 litres water in a large saucepan, simmer over low heat until infused (50 minutes). Strain through a fine sieve (discard solids), transfer 1 litre hot liquid to a bowl (reserve remainder). Squeeze excess water from gelatine, add to bowl, stir to dissolve. Add moscato, lemon juice and 80ml crème de mûre. Strain half into trifle bowl, scatter over 250gm berries and refrigerate until set (2-2½ hours). Chill remaining berry jelly, removing from refrigerator if it starts to set.

3. Reduce 250ml remaining liquid (discard excess) over high heat to 50ml or until syrupy (10-15 minutes), refrigerate until required.

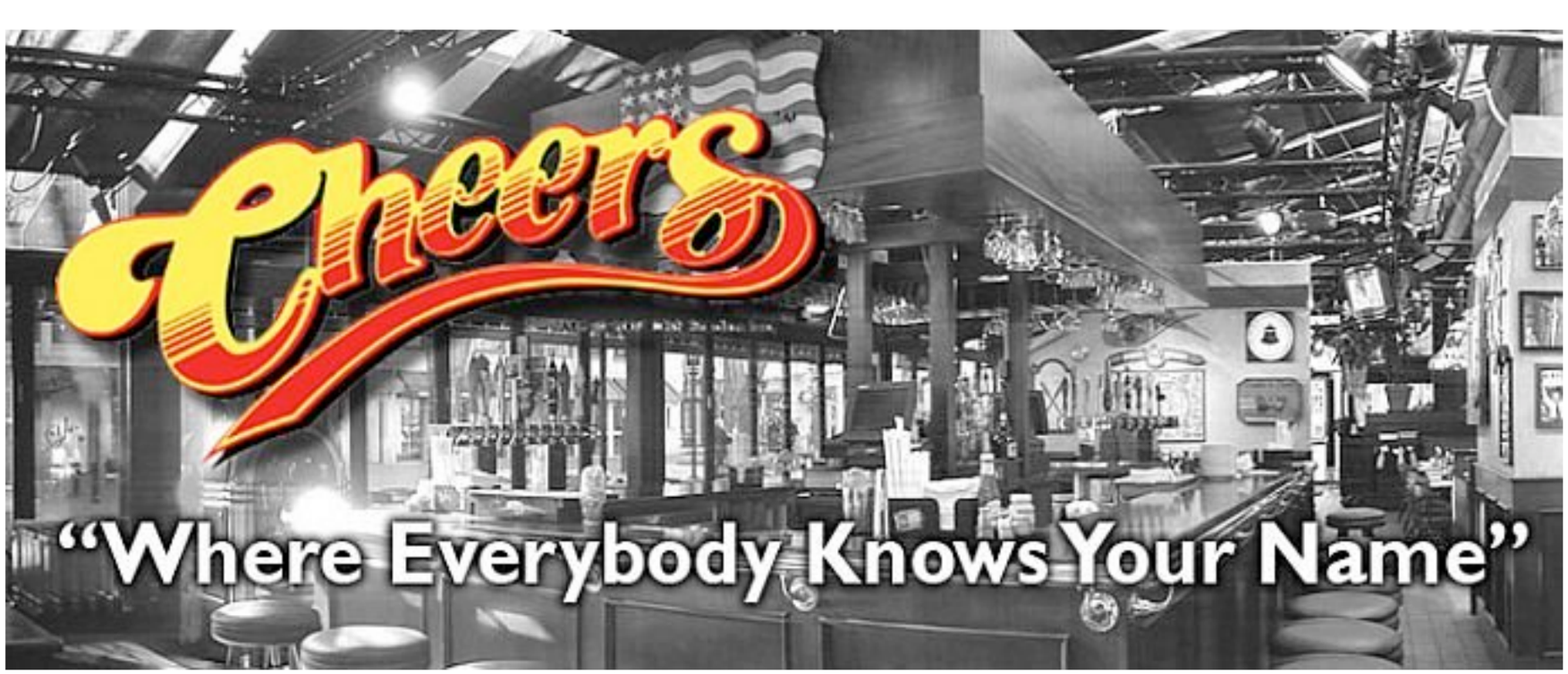
4. Meanwhile, combine crème fraiche, milk, rind, icing sugar and remaining vanilla seeds in a bowl, adding extra milk if necessary until spreadable. Spread one-third over set jelly, top with a sponge round, fill any gaps with trimmings, drizzle with 125ml crème de mûre. Scatter over remaining berries, pour over remaining jelly (mixture should be starting to set). Refrigerate until set (2-2½ hours). Top with half the remaining crème fraiche mixture, then remaining sponge. Drizzle with remaining crème de mûre, top with remaining crème fraiche mixture. Cover, refrigerate overnight. Serve scattered with extra berries and drizzled with blackberry syrup.

source is real

"**The source code is real**. The objects are realizations of the source code. Source for EVERY user modified object is placed in a particular directory or directories, for later editing and retrieval."

-- from the Emacs Speaks Statistics (ESS) manual

Names matter



ain't nothing like the real thing

minimize the creation of excerpts and copies of your data ... it will just confuse you later

here's we switched to guided hands-on computing