

Literate Statistical Programming with knitr and R Markdown

Biostatistics 140.776

What is knitr?

- An R package written by Yihui Xie (while he was a grad student at Iowa State)
 - Available on CRAN
- Supports RMarkdown, LaTeX, and HTML as documentation languages
- Can export to PDF, HTML
- Built right into RStudio for your convenience

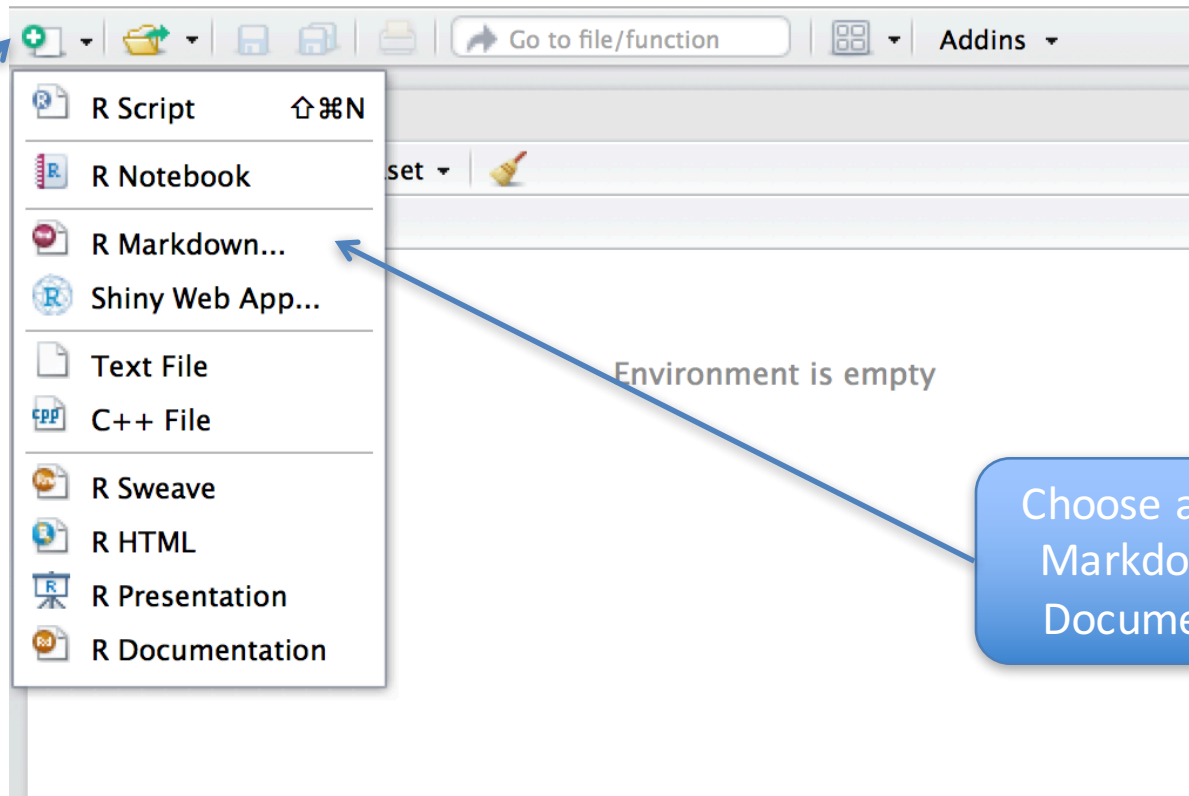
What is knitr Good For?

- Manuals
- Short/medium-length technical documents
- Tutorials
- Reports (esp. if generated periodically)
- Data preprocessing documents/summaries

What is knitr NOT Good For?

- Very long research articles
- Complex time-consuming computations
- Documents that require precise formatting

My First knitr Document



Create a new document

Choose an R Markdown Document

My First knitr Document

```
1 My First knitr Document
```

```
2
```

```
3
```

```
4 This is some text (i.e. a "text chunk").
```

```
5
```

```
6 Here is a code chunk
```

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

```
8 set.seed(1)
```

```
9 x <- rnorm(100)
```

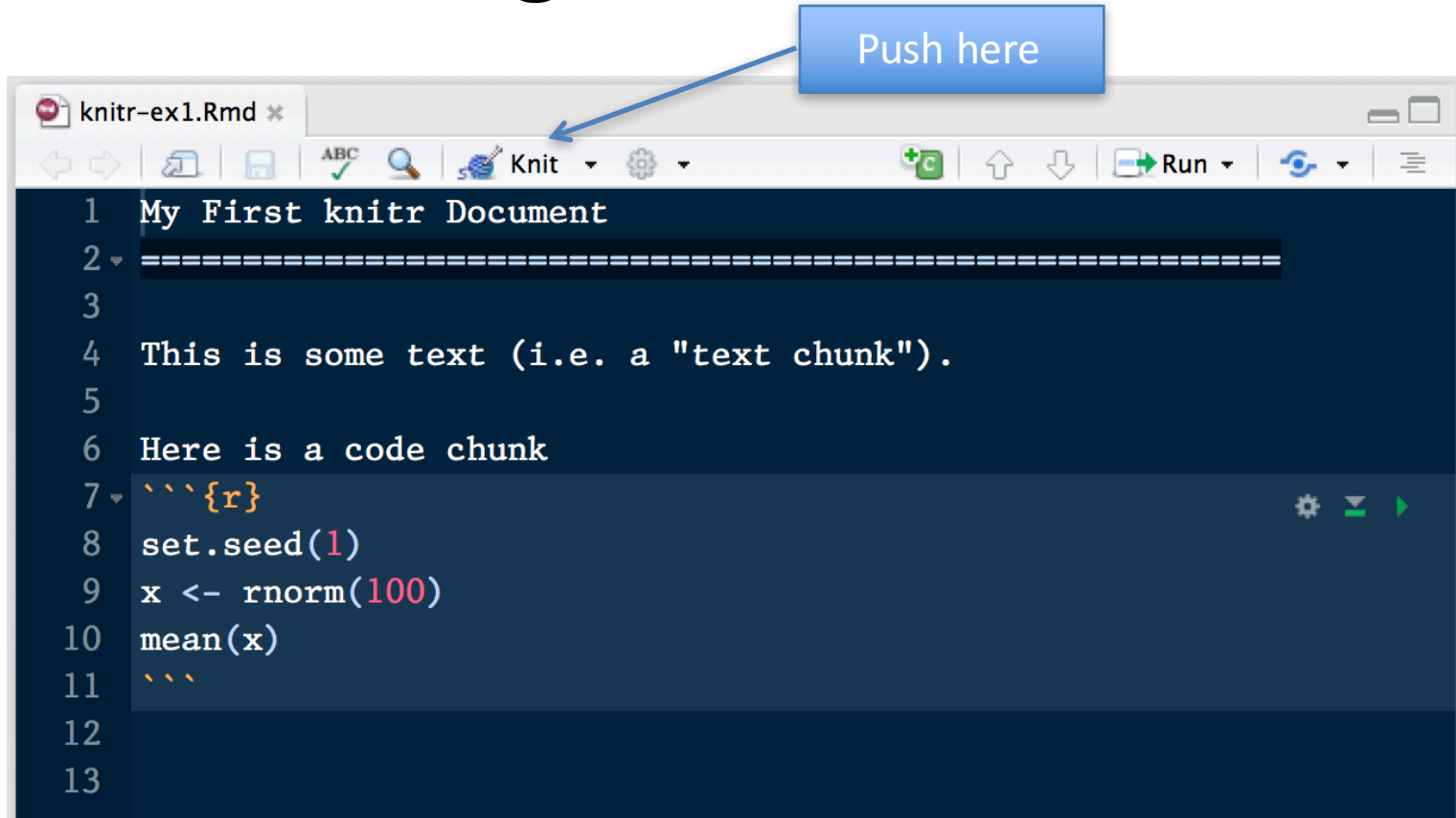
```
10 mean(x)
```

```
11 ```
```

Start of code chunk

End of code chunk

Processing a knitr Document



More Complicated Way

```
library(rmarkdown)
setwd(<working directory>)
render("document.Rmd")
browseURL("document.html")
```


HTML Output

My First knitr Document

This is some text (i.e. a “text chunk”).

Here is a code chunk

```
set.seed(1)  
x <- rnorm(100)  
mean(x)
```

Code input

```
## [1] 0.1089
```

Numerical output

What knitr Produces: Markdown

RMarkdown Document

```
1 My First knitr Document
2 
3 
4 This is some text (i.e. a "text chunk").
5 
6 Here is a code chunk
7 ```{r}
8 set.seed(1)
9 x <- rnorm(100)
10 mean(x)
11 ```
```

Code is
echoed

Markdown Document (generated)

```
1 My First knitr Document
2 
3 
4 This is some text (i.e. a "text chunk").
5 
6 Here is a code chunk
7 
8 ```r
9 set.seed(1)
10 x <- rnorm(100)
11 mean(x)
12 ```
13 
14 ```
15 ## [1] 0.1089
16 ```
```

Result of
evaluating R
code

A Few Notes

- knitr will fill a new document with filler text; delete it
- Code chunks begin with ```` `{r}` and end with ``````
- All R code goes in between these markers
- Code chunks can have **names**, which is useful when we start making graphics

```
````{r firstchunk}
R code goes here
````
```
- By default, code in a code chunk is echoed, as will the results of the computation (if there are results to print)

Processing of knitr Documents (what happens under the hood)

- You write the RMarkdown document (.Rmd)
- knitr produces a Markdown document (.md)
- knitr converts the Markdown document into HTML (by default)
- .Rmd → .md → .html
- You should NOT edit (or save) the .md or .html documents until you are finished

Another Example

```
# My First knitr Document  
Roger D. Peng
```

Level 1 heading

```
## Introduction
```

Level 2 heading

This is some text (i.e. a "text chunk"). Here is a code chunk.

```
``{r simulation,echo=FALSE}  
set.seed(1)  
x <- rnorm(100)  
mean(x)  
``
```

Do not echo code

Output

My First knitr Document

Roger D. Peng

Introduction

This is some text (i.e. a “text chunk”). Here is a code chunk.

```
## [1] 0.1089
```

Hiding Results

```
# My First knitr Document  
Roger D. Peng
```

```
## Introduction
```

This is some text (i.e. a "text chunk"). Here is a code chunk but it doesn't print anything!

```
``{r simulation,echo=FALSE,results="hide"}  
set.seed(1)  
x <- rnorm(100)  
mean(x)  
``
```

Output

My First knitr Document

Roger D. Peng

Introduction

This is some text (i.e. a “text chunk”). Here is a code chunk but it doesn't print anything!

Inline Text Computations

```
# My First knitr Document
```

```
## Introduction
```

```
```${r computetime,echo=FALSE}  
time <- format(Sys.time(), "%a %b %d %X %Y")
rand <- rnorm(1)
```
```

The current time is ``r time``. My favorite random number is ``r rand``.|

Inline Text Computations

My First knitr Document

Introduction

The current time is Wed Sep 04 16:42:09 2013. My favorite random number is 1.1829.

Incorporating Graphics

```
## Introduction
```

Let's first simulate some data.

```
` `{r simulatedata,echo=TRUE}
```

```
x <- rnorm(100); y <- x + rnorm(100, sd = 0.5)
```

```
` `
```

Here is a scatterplot of the data.

```
` `{r scatterplot,fig.height=4}
```

```
library(ggplot2)
```

```
qplot(x, y, main = "My Simulated Data")
```

```
` `
```



Adjust figure height

What knitr Produces in HTML

```
<body>
```

```
<h2>Introduction</h2>
```

```
<p>Let's first simulate some data.</p>
```

```
<pre><code class="r">x <- rnorm(100)
y <- x + rnorm(100, sd = 0.5)
</code></pre>
```

```
<p>Here is a scatterplot of the data.</p>
```

```
<pre><code class="r">par(mar = c(5, 4, 1, 1), las = 1)
plot(x, y, main = "My Simulated Data")
</code></pre>
```

```
<p>Estimate</b> | <b>Std. Error</b> | <b>t value</b> | <b>Pr(&gt;  t )</b> |
|-------------|-----------------|-------------------|----------------|---------------------|
| (Intercept) | -64.3421        | 23.0547           | -2.79          | 0.0062              |
| Wind        | -3.3336         | 0.6544            | -5.09          | 0.0000              |
| Temp        | 1.6521          | 0.2535            | 6.52           | 0.0000              |
| Solar.R     | 0.0598          | 0.0232            | 2.58           | 0.0112              |

# Setting Global Options

- Sometimes we want to set options for **every** code chunk that are different from the defaults
- For example, we may want to suppress all code echoing and results output
- We have to write some code to set these global options



# Setting Global Options

```
Introduction
```

```
``{r setoptions,echo=FALSE}
knitr::opts_chunk$set(echo = FALSE, results = "hide")
``
```

Set default to NOT  
echo code

First simulate data

```
``{r simulatedata,echo=TRUE}
x <- rnorm(100); y <- x + rnorm(100, sd = 0.5)
``
```

Override default

Here is a scatterplot of the data.

```
``{r scatterplot,fig.height=4}
library(ggplot2)
qplot(x, y, main = "My Simulated Data")
``
```

Don't echo code here

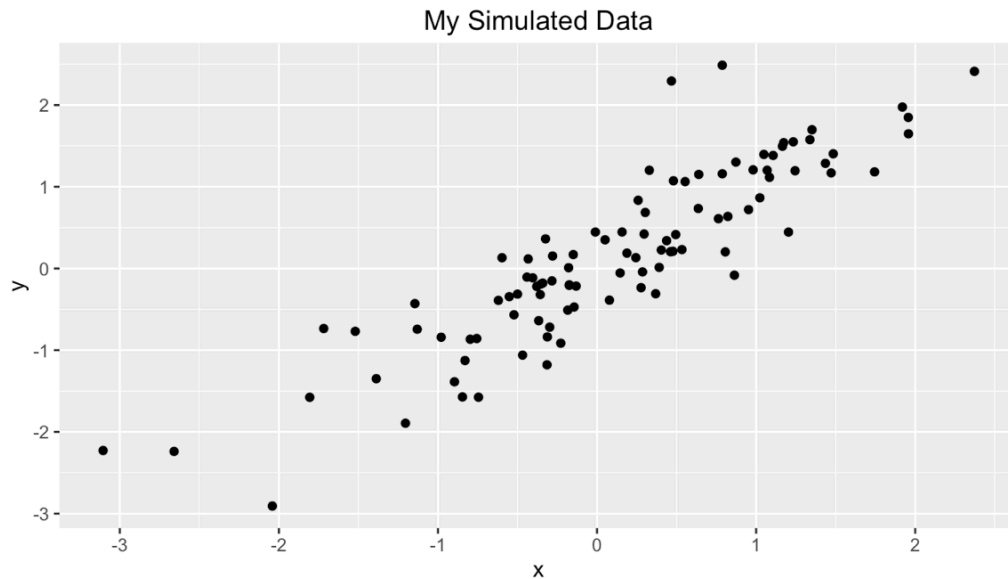
# Setting Global Options

## Introduction

First simulate data

```
x <- rnorm(100); y <- x + rnorm(100, sd = 0.5)
```

Here is a scatterplot of the data.



# Some Common Options

- Output
  - results: “asis”, “hide”
  - echo: TRUE, FALSE
- Figures
  - fig.height: numeric
  - fig.width: numeric

# Caching Computations

- What if one chunk takes a long time to run?
- All chunks have to be re-computed every time you re-knit the file
- The `cache=TRUE` option can be set on a chunk-by-chunk basis to store results of computation
- After the first run, results are loaded from cache

# Caching Caveats

- If the data or code (or anything external) changes, you need to re-run the cached code chunks
- Dependencies are not checked explicitly
- Chunks with significant *side effects* may not be cacheable

# Summary

- Literate statistical programming can be a useful way to put text, code, data, output all in one document
- knitr is a powerful tool for integrating code and text in a simple document format