



CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY

knitr 与动态报告生成系统

第五届 R 会议北京会场

魏太云

中国人民大学统计学院

2012 年 5 月

- 1 文学化编程与 knitr
- 2 knitr 简介
 - 语法概览和全局参数
 - 代码和文本输出
 - 图片和动画
 - 缓存
 - 代码引用和外部代码
 - 编辑器
 - 常见错误
- 3 应用案例与学习资源

提纲

1 文学化编程与 knitr

2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

3 应用案例与学习资源

什么是文学化编程？

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 humans what we want the computer to do.

– Donald E. Knuth, *Literate Programming*, 1984

简单的说就是：文章和代码混编，生成动态报告。

为什么需要文学化编程？

- ▶ 减少体力劳动
- ▶ 保证正确
- ▶ 提高可读性
- ▶ 便于可重复研究 (Reproducible Research)
- ▶ 便于维护

已有的工作

- ▶ 十年之痒 Sweave (S + weave)
- ▶ 其他相关 R 包: cacheSweave, pgfSweave, odfWeave, R2HTML, highlight, ...
- ▶ 其他统计软件: SASweave, Statweave, ...
- ▶ 其他语言: Org-mode (Emacs), Javadoc, noweb, cweb,

为何重复发明轮子？

- ▶ 爱之深，痛之切
- ▶ 改良之路走不通：
 - ▶ 700 行源代码，想添加 3 行？先复制 700 行...
 - ▶ 位图支持 (png, jpeg)
 - ▶ 需要更灵活的输出 (比如图片大小控制)
- ▶ $\text{knitr} = \text{Sweave} + \text{cacheSweave} + \text{pgfSweave} + \text{weaver} + \text{R2HTML}::\text{RweaveHTML} + \text{highlight}::\text{HighlightWeaveLatex} + 0.2 * \text{brew} + 0.1 * \text{SweaveListingUtils} + \text{more}$

如何用 R 进行可重复研究？

如何用 R 进行可重复研究？

- ▶ 有些青年：运行程序复制黏贴复制黏贴... 运行程序复制黏贴...

如何用 R 进行可重复研究？

- ▶ 有些青年：运行程序复制黏贴复制黏贴... 运行程序复制黏贴...
- ▶ 普通青年：Sweave

如何用 R 进行可重复研究？

- ▶ 有些青年：运行程序复制黏贴复制黏贴... 运行程序复制黏贴...
- ▶ 普通青年：Sweave
- ▶ 文艺青年：knitr

如何用 R 进行可重复研究？

- ▶ 有些青年：运行程序复制黏贴复制黏贴... 运行程序复制黏贴...
- ▶ 普通青年：Sweave
- ▶ 文艺青年：knitr

有何文艺之处？请听后续分解

提纲

1 文学化编程与 knitr

2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

3 应用案例与学习资源

名称

- ▶ $\text{knitr} = \text{knit} + \text{R}$ (类似 $\text{S} + \text{weave}$);
- ▶ 发音: neater?
- ▶ R 小写
- ▶ 唯一性, 不是一个单词 (失败案例: ConTeXt)



特性

- ▶ 代码整理和高亮
- ▶ 图片控制
- ▶ 结果缓存
- ▶ 支持 \LaTeX 、Markdown、HTML、rST 等多格式输入和 HTML、PDF 等格式的输出
- ▶ 代码外部化和引用
- ▶ 灵活的 API：定制方便

提纲

1 文学化编程与 knitr

2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

3 应用案例与学习资源

语法概览

▶ Rnw 文件:

段落: `<<name, echo=TRUE, tidy=TRUE>>=`
这里是段落代码
`@`

行内: `\Sexpr{这里是行内代码}`

▶ Rhtml 文件:

段落: `<!--begin.rcode my-label, fig.width=5, dev=png`
这里是段落代码
`end.rcode-->`

行内: `<!--rinline 这里是行内代码 -->`

▶ Rmd 文件:

段落: ```` {r test-r, engine='R', comment=NA, dev='CairoPNG'}`
这里是段落代码
`````

行内: ``r 这里是行内代码``

# 全局参数

## ▶ Rnw 文件:

```
\SweaveOpts{fig.align='center', cache=TRUE, tidy=FALSE, comment=NA}
```

或:

```
<<setup, include=FALSE, cache=FALSE>>=
opts_chunk$set(fig.align='center', cache=TRUE, tidy=FALSE, comment=NA)
@
```

## ▶ Rhtml 文件:

```
<!--begin.rcode setup,echo=FALSE,results=hide,message=FALSE,cache=FALSE
 opts_chunk$set(fig.width=5, fig.height=5, ,fig.align='center')
end.rcode-->
```

## ▶ Rmd 文件:

```
`r opts_chunk$set(fig.width=6, fig.height=6, fig.path='')`
```

# 所有内建模式

## ▶ 更多细节:

```
> require(knitr)
> str(all_patterns)
```

# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

## 3 应用案例与学习资源

## 段落代码和文本输出参数

- ▶ `eval`: (`TRUE`; 逻辑): 是否执行代码
- ▶ `tidy`: (`TRUE`; 逻辑): 是否整理代码
- ▶ `prompt`: (`FALSE`; 逻辑): 是否添加引导符 `'>'`
- ▶ `highlight`: (`TRUE`; 字符): 是否高亮代码
- ▶ `size`: (`'normalsize'`; 字符): 大小 (`暂时无效, 开发版已经生效`)
- ▶ `background`: (`'#F7F7F7'`; 字符或数值): 背景颜色
- ▶ `comment`: (`'##'`; 字符): 结果输出前缀符号
- ▶ `echo`: (`TRUE`; 逻辑或数值): 是否输出代码或输出哪些行
- ▶ `results`: (`'markup'`; 字符): 装裱输出 (`'markup'`)、原样输出 (`'asis'`)、隐藏 (`'hide'`)
- ▶ `warning`, `error`, `message`: (`TRUE`; 逻辑): 是否显示相应信息
- ▶ `split`: (`FALSE`; 逻辑): 是否剥离代码和文本到外部文件
- ▶ `include`: (`TRUE`; 逻辑): 是否保留代码或结果到最终文档

# 代码

```
<<example-sexpr, prompt=TRUE>>=
```

```
x = date()
```

```
y = 1:10
```

```
matrix(1:6, nrow=2)
```

```
@
```

本幻灯片最后修改时间是\Sexpr{x}, y值是\Sexpr{y}。

```
> x = date()
```

```
> y = 1:10
```

```
> matrix(1:6, nrow=2)
```

```
 [,1] [,2] [,3]
[1,] 1 3 5
[2,] 2 4 6
```

本幻灯片最后修改时间是 Thu May 24 02:52:40 2012, y 值是 1, 2, 3, 4, 5, 6, 7, 8, 9, 10。

# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- **图片和动画**
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

## 3 应用案例与学习资源

# 主要参数

- ▶ `fig.path`: ('figure/'; 字符): 图片路径, 支持前缀模式 ('figure/prefix-')
- ▶ `fig.keep`: ('high'; 字符): 保存图形类型, 高级图形 ('high')、不保存 ('none')、所有图形 ('all')、第一张 ('first')、最后一张 ('last')
- ▶ `fig.show`: ('asis'; 字符): 展示方式, 紧随代码输出 ('asis')、最后统一输出 ('hold')、动画输出 ('animate')
- ▶ `dev`: (LaTeX 为 'pdf', HTML/markdown 为 'png'; 字符): 输出设备, knitr 支持很多种设备
- ▶ `fig.width`, `fig.height`: (7; 数值): 图片文件的宽、高 (英寸 2.54cm 为单位)
- ▶ `out.width`, `out.height`: (NULL; 字符): 图片在输出文档中的宽、高
- ▶ `fig.align`: ('default'; 字符): 对齐方式, 不做调节 ('default')、左 ('left')、右 ('right')、居中 ('center')
- ▶ `interval`: (1; 数值): 动画参数, 切换画面时间, 单位为秒



# 图片和动画

## ▶ 图片

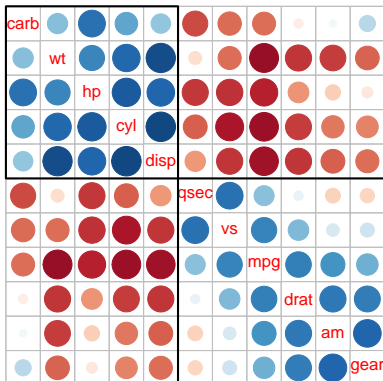
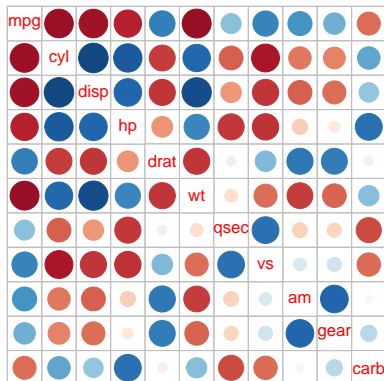
```
<<fig.width=4, fig.height=4, out.width='0.46\\linewidth'>>=
require(corrplot)
corrplot(cor(mtcars),addtextlabel="d", addcolorlabel="no")
corrplot(cor(mtcars),order="h",addtextlabel="d", addcolorlabel="no",addrect=2)
@
```

## ▶ 动画

```
<<fig.width=3, fig.height=3, out.width='.35\\linewidth', fig.show='animate'>>=
par(mar = rep(2.3, 4))
for (i in seq(pi/2, -4/3 * pi, length = 12)) {
 plot(0, 0, pch = 20, ann = FALSE, axes = FALSE)
 arrows(0, 0, cos(i), sin(i))
 axis(1, 0, "6"); axis(2, 0, "9")
 axis(3, 0, "12"); axis(4, 0, "3"); box()
}
@
```

# 图片输出

```
require(corrplot)
corrplot(cor(mtcars), addtextlabel="d", addcolorlabel="no")
corrplot(cor(mtcars), order="h", addtextlabel="d", addcolorlabel="no", addrect=2)
```



# 动画输出

```
par(mar = rep(2.3, 4))
for (i in seq(pi/2, -4/3 * pi, length = 12)) {
 plot(0, 0, pch = 20, ann = FALSE, axes = FALSE)
 arrows(0, 0, cos(i), sin(i))
 axis(1, 0, "6"); axis(2, 0, "9")
 axis(3, 0, "12"); axis(4, 0, "3"); box()
}
```

# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- **缓存**
- 代码引用和外部代码
- 编辑器
- 常见错误

## 3 应用案例与学习资源

# 缓存

- ▶ 为什么需要缓存?
- ▶ 主要参数:
  - ▶ `cache`: (`FALSE`; 逻辑): 是否开启缓存
  - ▶ `cache.path`: (`'cache/'`; 字符): 缓存路径

# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

## 3 应用案例与学习资源

# 代码引用 (对 chunk1 |)

## ① <<>> 格式 (可以多级别引用):

```
<<chunk2>>=
<<chunk1>>
@
```

## ② 相同标签模式 (后一个代码必须置空):

```
<<chunk1, echo=FALSE, results='markup'>>=
@
```

## ③ 使用 ref.label 参数 (这样可以分离代码和结果):

```
<<chunk2, ref.label='chunk1', echo=FALSE, results='markup'>>=
@
```

## ④ 使用 run\_chunk() 函数 (支持嵌套结构)

```
<<a>>=
x = 1; x
run_chunk('b'); x
@
<>=
x = 2
@
<<c>>=
run_chunk('a'); x
@
```

# 外部代码

## ▶ 外部代码out.R:

```
@knitr Q1
rnorm(10)
@knitr Q2
sample(10)
```

## ▶ 引用外部代码

```
<<set-options, echo=FALSE, cache=FALSE>>=
read_chunk('out.R')
@
<<Q1>>=
@
<<Q2>>=
@
```



# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- **编辑器**
- 常见错误

## 3 应用案例与学习资源

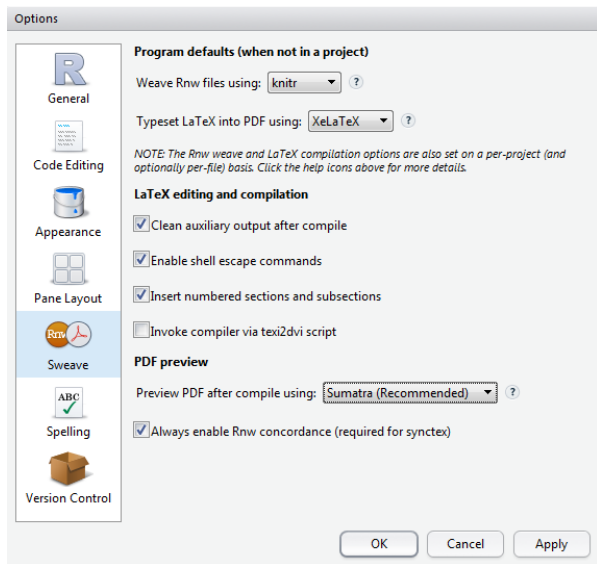
# 编辑器的配置

- ▶ **RStudio**
- ▶ **LyX**
- ▶ Emacs Org-mode
- ▶ Eclipse
- ▶ TeXStudio
- ▶ WinEdt

参见：

- ▶ <http://yihui.name/knitr/demo/rstudio/>
- ▶ <http://yihui.name/knitr/demo/lyx/>
- ▶ <http://yihui.name/knitr/demo/editors/>

## RStudio(最新版 0.96.225)



# NYC Meetup: What's Next for R Markdown

RStudio Blog (May 24, 2012 in News | by jjallaire):

There's been lots of excitement about the new R Markdown feature introduced as part of knitr 0.5 and RStudio 0.96. People see R Markdown as both a simpler way to do reproducible research and as a great way to publish to the web from R. Jeromy Anglim has a nice write up on getting started with R Markdown and Marcus Gesmann describes how to embed Google Visualizations using his googleVis package.

We are just as excited about R Markdown and think there is lots more that can be done with it. We'll be talking about this along with **Yihui Xie (author of knitr)** and **Jeff Horner (author of R/Apache and Rook)** on Tuesday June 5th in New York:

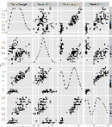
<http://www.meetup.com/nyhackr/events/64279002/>

At the meetup we'll be showing the latest versions of knitr and RStudio and will be announcing some new R Markdown stuff—if you are in New York we'd love to see you there!

# NYC Meetup: What's Next for R Markdown





## New York Open Statistical Programming Meetup

Home
Members
Sponsors
Photos
Pages
Discussions ▾
More ▾
Join us!




**New York, NY**  
Founded Mar 12, 2009

Read more about us...


- Data hackers 1,745
- Group reviews 44
- Upcoming Meetups 1
- Past Meetups 28
- Our calendar 
- Follow us   

We're about:  
[Internet & Technology](#) ·  
[Math](#) · [Data Management](#) ·  
[Software Developers](#) · [R Users Group](#)

Organizer:  
**Drew Conway** 

## Dynamic Reports with RStudio, knitr, and Markdown


---



**Tuesday, June 5, 2012**  
6:15 PM

SELECTED BY: DREW CONWAY


---



**AOL HQ**  
770 Broadway 6th Floor  
B'way & 9th St. Use entrance on 9th St.  
New York, NY ([map](#))

SELECTED BY: DREW CONWAY

---



**Price: \$5.00/per person**

[Refund policy](#)

---

Yihui Xie ([knitr](#)), JJ Allaire ([RStudio](#)), and Jeff Horner ([R/Apache and Rook](#)) will unveil a new suite of tools they've collaborated on that make dynamic reports and reproducible research with R easier and more flexible. They'll cover new features of knitr and RStudio for creating publication-quality PDF reports. They'll also describe a new authoring format, R Markdown, that makes it easy to create web pages with embedded R computations and graphics.

---

You must be a member to post a comment. [Join](#) or [login](#).

Will you attend?

Join us!

**81 attending**

119 spots available

RSVPs close: Jun 05, 12:00 PM

---

EVENT

[Drew Conway](#)

ORGANIZER

---

[Parag](#)

---

[Vincent Louie](#)

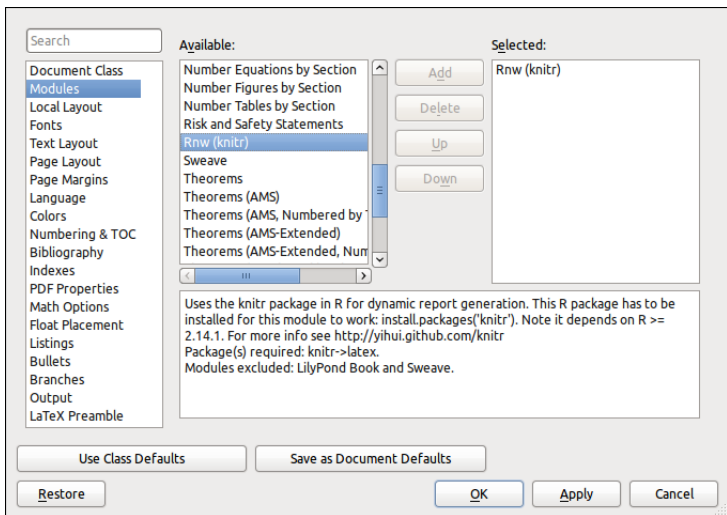
---

[Randy Reynolds](#)

---

[Joscha Legewie](#)

◀ ▶ 📄 ◀ ▶ ☰ 🔍 ↺



# 提纲

## 1 文学化编程与 knitr

## 2 knitr 简介

- 语法概览和全局参数
- 代码和文本输出
- 图片和动画
- 缓存
- 代码引用和外部代码
- 编辑器
- 常见错误

## 3 应用案例与学习资源

# 常见错误

## ▶ 参数问题

```
label参数重复
echo参数越界
图片大小设置不当，比如混淆fig.width和out.width
```

## ▶ 中文问题：UTF-8 编码，代码不能有中文变量；最好 X<sub>Y</sub>LaTeX 编译

```
##Rprofile.site 文件配置
options(stringsAsFactors=FALSE, help_type="html", encoding="UTF-8")
Sys.setlocale("LC_CTYPE", "chs")
```

## ▶ Beamer 中的问题

```
[fragile] 参数问题：
\begin{frame}[fragile]{代码例子} 含代码的片子中勿忘fragile参数
\AtBeginSection[] 等配置环境中，{frame}后不能加fragile参数
```



# 提纲

- 1 文学化编程与 knitr
- 2 knitr 简介
  - 语法概览和全局参数
  - 代码和文本输出
  - 图片和动画
  - 缓存
  - 代码引用和外部代码
  - 编辑器
  - 常见错误
- 3 应用案例与学习资源

# 应用案例

- ▶ 大家手中的会议手册
- ▶ 本次 R 会议邱怡轩、魏太云、陈丽云、黄金山等人的幻灯片
- ▶ ggplot2 0.9.0 guide 文档
- ▶ ggbio 文档: <http://tengfei.github.com/ggbio/>
- ▶ 更多 (包括书籍、网站等):  
<http://yihui.name/knitr/demo/showcase/>

# 学习资源

- ▶ 主页: <http://yihui.name/knitr/>
- ▶ 演示文件: <http://yihui.name/knitr/demo/>
- ▶ 开发页面: <https://github.com/yihui/knitr/>
- ▶ 邮件列表: <http://groups.google.com/group/knitr>
- ▶ 文档下载: <https://github.com/yihui/knitr/downloads>

# 谢谢大家！



邮箱: [taiyun.wei@cos.name](mailto:taiyun.wei@cos.name)

博客: <http://taiyun.cos.name>

微博: <http://weibo.com/taiyun>

github: <https://github.com/taiyun>

## Q/A?