Applications of R packages in Clinical Epidemiology (RiCE)

Week 1 - Introduction

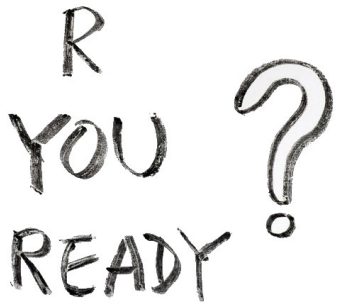
唐迅

tangxun@bjmu.edu.cn

2020-05-19

教学大纲

- 课程代码：GD191008；
- 课程名称：医学统计学专题 - R 程序包在临床流行病学中的应用；
- 英文名称：Applications of R packages in Clinical Epidemiology (RiCE)；
- 开课单位：北京大学公共卫生学院流行病与卫生统计学系；
- 教学内容：
  1. 绪论（2学时）；
  2. 数据管理进阶（4学时）；
  3. 病因学研究（4学时）；
  4. 诊断学研究（4学时）；
  5. 预后研究（4学时）。
- 相关网站：
  1. https://www.r-project.org/
  2. https://www.rstudio.com/



前期先修课程和知识

1. 流行病学研究方法 I
2. 临床研究方法学（基础 + 高级）
3. 医学统计学专题 - R 程序包应用基础

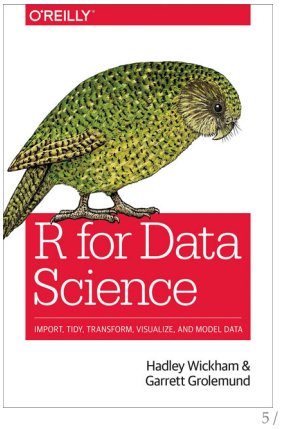
开始本课程学习前，需要先熟悉R的基本知识，包括程序包的安装等，建议：

1. 提前完成上述先修课程的学习；
2. 安装 swirl 程序包，并自行进行交互式学习和练习；

```
install.packages("swirl")  # 安装swirl程序包
library(swirl)  # 载入swirl程序包
swirl()  # 运行swirl程序包
```


课程推荐阅读

- *Hadley Wickham, Garrett Grolemund*. **R for Data Science: Import, Tidy, Transform, Visualize, and Model Data**. O'Reilly Media; 1st edition. 2017；
- 免费在线阅读 R for Data Science https://r4ds.had.co.nz/



American Journal of Epidemiology
Copyright © 2006 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved; printed in U.S.A.
Vol. 163, No. 9
DOI: 10.1093/aje/kwj093
Advance Access publication March 1, 2006

Commentary

Reproducible Epidemiologic Research

Roger D. Peng, Francesca Dominici, and Scott L. Zeger

From the Biostatistics Department, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Received for publication November 4, 2005; accepted for publication January 10, 2006.

The replication of important findings by multiple independent investigators is fundamental to the accumulation of scientific evidence. Researchers in the biologic and physical sciences expect results to be replicated by independent data, analytical methods, laboratories, and instruments. Epidemiologic studies are commonly used to quantify small health effects of important, but subtle, risk factors, and replication is of critical importance where results can inform substantial policy decisions. However, because of the time, expense, and opportunism of many current epidemiologic studies, it is often impossible to fully replicate their findings. An attainable minimum standard is "reproducibility," which calls for data sets and software to be made available for verifying published findings and conducting alternative analyses. The authors outline a standard for reproducibility and evaluate the reproducibility of current epidemiologic research. They also propose methods for reproducible research and implement them by use of a case study in air pollution and health.

Reproducible research

What is reproducible research?

> "... the calculation of quantitative scientific results by independent scientists using the original datasets and methods"

What do we need?

- Reproducibility = Tools + Workflow
- We need to link data, code, results, and interpretation
- use `knitr` to integrate R and markdown with RStudio

```
install.packages("knitr")  # 安装knitr程序包
```

- write text (in markdown) separated by code chunks (written in R)

Example

What dataset are they from?

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```


Example

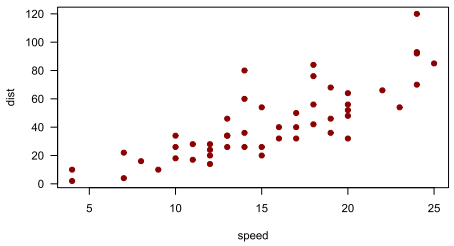
What dataset are they from?

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```


Example

```
plot(cars, pch = 19, col = 'darkred', las = 1)
```



Markdown

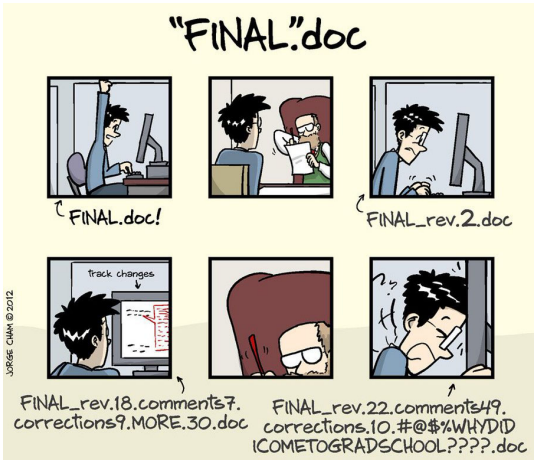
> *Markdown*是一种轻量级的**标记语言**：

1. 标题 Heading
   - `# 一级标题`
   - `## 二级标题`
   - `### 三级标题`
2. 列表 List
   - `-`
     - `(tab)-`
3. 链接 Link
   - `[website name](url)`
4. 图片 Image
   - `![alt text](http://example.com/logo.png)`

使用Markdown的优点

- [*]语法简单，写作时可以专注于文字内容，而不是排版格式；
- 纯文本内容，可以兼容所有的文本编辑器与字处理软件；
- 轻松导出到HTML、PDF和本身的`.md`后缀名的文件；
- 随时修改文件版本，不必担心字处理软件修改版本后的混乱；
- 可读、直观，学习成本低。

[*] 关于 Markdown 的语法可参考 GitHub 网站的 Mastering Markdown，https://guides.github.com/features/mastering-markdown/, 不同软件或网站支持的 Markdown 语法略有不同，但简单的语法可能也足够了。



Rmarkdown

- 包含**代码**与**分析结果**的分析报告，用于展示和交流；
- 可重现分析结果的文件（分析/研究的再现性高）；
- 支持各种输出格式：Slides, PDF, HTML，Word，……

Rmarkdown

- http://rmarkdown.rstudio.com/
- R + Markdown (基于Markdown文件)
- 从菜单 `File -> New File -> R Markdown` 创建一个新文档；
- **R Code Chunks**的R程序代码都可以执行，并输出结果；
- 打开扩展名为.Rmd的文件，点击 `Knit` 按钮编译文档，可输出各种格式；
- 输出选项：
  - 文件：HTML, PDF, Word
  - Slides：HTML, PDF
  - Shiny
  - 模板：放在GitHub使用的.md扩展名的文件

R Code Chunk的参数

- eval：是否**执行并显示**在输出的Markdown文件内
  - T：执行+显示
  - F：只显示，不执行
- echo：是否**显示代码**（不显示代码仍会执行，但只显示执行结果）
  - T：执行+显示
  - F：不显示代码，只执行，并显示结果
- warning, error, message：是否显示错误信息
  - T：显示
  - F：不显示
- cache：是否要缓存结果

Table的输出

若需要输出表格，输出方式如下：

```
iris[1:10, 1:5]
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1           5.1         3.5          1.4         0.2  setosa
## 2           4.9         3.0          1.4         0.2  setosa
## 3           4.7         3.2          1.3         0.2  setosa
## 4           4.6         3.1          1.5         0.2  setosa
## 5           5.0         3.6          1.4         0.2  setosa
## 6           5.4         3.9          1.7         0.4  setosa
## 7           4.6         3.4          1.4         0.3  setosa
## 8           5.0         3.4          1.5         0.2  setosa
## 9           4.4         2.9          1.4         0.2  setosa
## 10          4.9         3.1          1.5         0.1  setosa
```


Table的输出（knitr::kable）

如果想要更好展示，建议使用**knitr::kable**

```
knitr::kable(
  iris[1:10, 1:5], caption = "A knitr kable."
)
```

Table: A knitr kable.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa

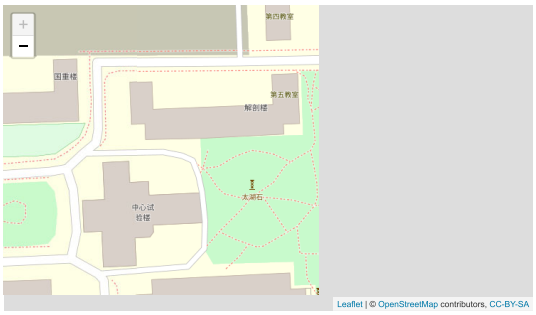
```
DT::datatable(
  head(iris, 10),
  fillContainer = FALSE, options = list(pageLength = 8)
)
```

Show 8 entries Search:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
4	4.6	3.1	1.5	0.2	setosa
8	5	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa
1	5.1	3.5	1.4	0.2	setosa
5	5	3.6	1.4	0.2	setosa

Showing 1 to 8 of 10 entries Previous 1 2 Next


```
library(leaflet) # 安装地图程序包
leaflet() %>% addTiles() %>% setView(116.35222, 39.98286, zoom = 17)
```



总结和建议

- 本节课程主要涉及的程序包主要有：
  - `knitr`（R markdown）
  - `swirl`（Learn R, in R）
  - `rticles`（LaTeX, PDF中文支持）
- 英文版及默认设置：特别避免使用中文路径和文件夹名称；
- 避免中文乱码：使用**UTF-8编码**；
- 养成编码**注释**（#）的习惯，多用快捷键；
- 学会使用Help和网络搜索引擎，并**不断试错**；
- 形成工作流（Workflow）；
- 练习，再练习，反复练习！

课后阅读

R for Data Science - V. Communicate
https://r4ds.had.co.nz/communicate-intro.html/

- 27 R Markdown
- 29 R Markdown formats
- 30 R Markdown workflow

谢谢！