

# Rmarkdown 生成 pdf 文档

周世祥

2020 年 5 月 16 日

## 关于 RMarkdown 生成 pdf 文档

为追求完美，费了好长时间，对新版本的 R(R4.0) 实验成功。

### 先看 rmarkdown 包创始人谢大大的文档

<https://yihui.org/tinytex/cn/>

rmarkdown 包从版本 1.9 开始，编译 R Markdown 为 PDF 时会调用 tinytex，这样一来，R Markdown 用户的 LaTeX 世界应该就安静了：因为对 R Markdown 用户来说，编译 PDF 的头号出错可能就是缺失 LaTeX 包（Pandoc 翻译 Markdown 为 LaTeX 代码时，几乎不太可能产生 LaTeX 语法错误），而可怜的 R 用户哪能明白那错误消息是什么意思。

TinyTeX 是一个瘦身版的 TeX Live。TeX Live 的庞大体型问题困扰我多年，在 2018 年之前我终于抽出一周时间来解决问题，其实方案很简单：把对普通用户毫无用处的源代码和文档去掉即可。具体技术细节参见常见问题，总体而言就是利用了 TeX Live 的自动化安装方式，配置文件中设置不要安装源文件和文档。

### LaTeX 编辑器的安装

TinyTeX 的安装和维护对 R 用户来说最简单，两行 R 代码加上两到六分钟的等待时间：

```
devtools::install_github('yihui/tinytex') #  
# 这个命令的作用是从github上安装R扩展包  
tinytex::install_tinytex() # 这个命令的作用是安装LaTeX编辑系统，据说很小
```

安装 TinyTeX 之前建议卸载系统中已有的 LaTeX 套件，如 TeX Live、MiKTeX、MacTeX 等。一个系统中最好不要有两个 LaTeX 套件同时存在，否则可能会产生冲突。

对其它用户，请参阅首页上的脚本安装方式，通常就是打开命令行窗口，运行一行命令即可。该命令会自动下载 TeX Live 的安装脚本并自动安装，由于是从网络下载安装 TeX Live，所以具体等待时间取决于网速。Unix 系统应该不会超过两分钟，Windows 系统可能需要五分钟左右（其间可能会弹出两次错误对话框，点确定即可；如果杀毒软件弹出警告，请允许修改）

<https://github.com/yihui/tinytex>

在 R 中运行命令 `devtools::install_github('yihui/tinytex')`，提示需要安装 Rtools 工具，这个可以从官网下载：<https://mirrors.tongji.edu.cn/CRAN/> 100 多兆，需要注意的是选国内镜像下载，我一般选同济大学或中科大，根据你的网速情况。

这里注意一个问题，我开头是 R3.6.3 版下了一个 Rtools4.0 结果不匹配，又重新升级 R 到 R4.0。

`tinytex::install_tinytex()`，这个命令我运行也不成功，开始下载安装了一部分就死了。注意到里面有个链接，是 `texlive2020` 安装文件压缩包，直接下载下来 20 多兆。解压后，里面有个批处理的安装文件，直接安装，结果安装失败，提示 C 盘空间不足，原来这是一个完全安装，删除安装文件夹，重新安装到 E 盘，发现占用了 7 个多 G 的空间，安装界面中也要选国内的镜像，安装了 3900 多个 TeX 包，花了一个多钟头。

看环境变量，`texlive` 确实安装成功了。但编译 pdf 仍然不成功。

### 编写 yaml 文件这是元数据，格式有讲究的，注意前面的空格。下面是网上的说法：

```
title: "PDF 测试"  
author: "douzi"
```

```

date: "2018/4/14"
CJKmainfont: Microsoft YaHei
output:
  pdf_document:
    includes:
      in_header: header.tex
    keep_tex: yes
    latex_engine: xelatex

```

原文链接：<https://blog.csdn.net/u012111465/java/article/details/79945372>

这是第二种说法：

```

header.tex
注意：

```

header.tex文件一定要是UTF-8 格式  
header.tex一定要与rmd放在同一路径下  
header.tex文件内容

```

%\documentclass{article}
\usepackage[BoldFont,SlantFont,CJKsetspace,CJKchecksingle]{xeCJK}
\setCJKmainfont[BoldFont=SimSun]{Microsoft YaHei} % 设置为雅黑字体
\setCJKmonofont{SimSun} % 设置缺省中文字体
\parindent 2em % 设置段首缩进

```

在代码中加入fig.showtext = TRUE, fig.align='center', 同时, 需要library(showtext)、showtext::showtext.be  
pdf中的图确实是中文给我整哪里去了

这是第三种解释：

```

title: "中文"
CJKmainfont: Microsoft YaHei
output:
  pdf_document:
    includes:
      header-includes:
        - \usepackage{xeCJK}
      keep_tex: yes
    latex_engine: xelatex

```

```
install.packages("rticles")
```

最后我的设置；

```

---
title: 'Rmarkdown生成pdf文档 '
author: "周世祥"
date: "2020年5月16日"
CJKmainfont: Microsoft YaHei
output:
  pdf_document:
    includes:
      in_header: header.tex
    keep_tex: yes
    latex_engine: xelatex
  html_document: default

```

---

另外要做的工作是再在当前工作目录中建一个 header.tex 文件，内容如下：

```
%\documentclass{article}
\usepackage{xCJK}
\setCJKmainfont{Microsoft YaHei}
%\usepackage[BoldFont,SlantFont,CJKsetspace,CJKchecksingle]{xeCJK}
%\setCJKmainfont[BoldFont=SimSun]{Microsoft YaHei} % 设置为雅黑字体
%\setCJKmonofont{SimSun} % 设置缺省中文字体
%\parindent 2em % 设置段首缩进
```

还要注意保存的编码格式为 utf8 格式。

虽然运行后有一些警告，最后是生成了中文正常的文档。

## 安装包

```
install.packages("rticles")
```

这个包安装后，在 Rstudio 菜单中有生成中文 LaTeX 的模板可用。

下面是范例文档。

## 假设检验

上海 1875-1955 年的 81 年间，根据其中 63 年的观察记录到的一年中 (5-9) 月下暴雨的次数整理资料：

一年中暴雨次数 $i$	0	1	2	3	4	5	6	7	8	$\geq 9$
实际年数 $n_i$	4	8	14	19	10	4	2	1	1	0

试检验一年中下暴雨的次数  $X$  是否服从泊松分布 (取显著性水平  $\alpha = 0.05$ )

解：现在的问题是显著性水平  $\alpha = 0.05$  下要检验

$H_0 : X$  服从  $P(\lambda)$

依题意，

$$\sum_{i=0}^9 f_i = n = 63$$

先计算出在假定  $H_0$  为真时，参数  $\lambda$  的极大似然估计值  $\lambda = \bar{x} = \frac{1}{63} * (0 * 4 + 1 * 8 + \cdots + 8 * 1 + 9 * 0) = \frac{180}{63} = 2.8571$

按  $X$  服从泊松分布，由  $p_i = \frac{2.8571^i}{i!} e^{-2.8571}, i = 0, 1, \dots, 6$

把暴雨次数大于 6 的频数和为一组算得  $p_0 = 0.0574, p_1 = 0.1641, p_2 = 0.2344, p_3 = 0.2233, p_4 = 0.1595, p_5 = 0.0911,$   
 $p_6 = \sum_{i=6}^{\infty} \frac{2.8571^i}{i!} e^{-2.8571} = 0.0702$  检验统计量的观测值为： $X^2 = \sum_{i=0}^6 \frac{(f_i - 63p_i)^2}{63p_i} = 2.9064$

查表，可得  $X^2(k - r - 1) = X^2_{0.05}(7 - 1 - 1) = 11.07$

所以接受原假设，认为  $X$  服从泊松分布。

```
x<- c(rep(0,4),rep(1,8),rep(2,14),rep(3,19),rep(4,10),rep(5,4),rep(6,2),7,8)
xbar<- mean(x)
p<- c(dpois(0:5,xbar),1-ppois(5,xbar))
v<-c(4,8,14,19,10,4,4)
sum((v-63*p)^2/(63*p))
```

```
## [1] 2.908962
```

```
qchisq(1-0.05,7-1-1)
```

```
## [1] 11.0705
```