

R与Office的整合

李舰



第七届中国R语言会议

北京，中国人民大学

2014年05月25日

目 录

1 自动化报告与可重复研究

- Office报告
- L^AT_EX报告
- 基于标记语言的报告
- 报告系统

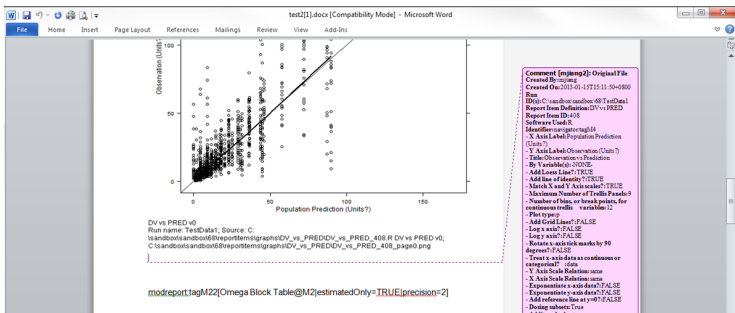
2 PPT和VBA

3 Rofficetool 简介

目 录

- 1 自动化报告与可重复研究
 - Office报告
 - L^AT_EX报告
 - 基于标记语言的报告
 - 报告系统
- 2 PPT和VBA
- 3 Rofficetool 简介

Word报告是所见即所得排版的典范



Excel也是方便的排版工具

D部分：受访者基本情况

为了能对问卷数据进行更好的分析，我们需要一些不涉及您隐私的基本信息，请回答以下问题：

- 您的性别？ ☐ 男 ☐ 女
- 您的年龄在那个范围？ ☐ 25岁及以下 ☐ 26~40岁 ☐ 41~60岁 ☐ 60岁以上
- 您的职业类型符合哪种情况？ ☐ 经常出差 ☐ 偶尔出差 ☐ 基本上不出差
- 您的月收入大约在哪个范围？ ☐ 2000元及以下 ☐ 2001~5000 ☐ 5000~10000 ☐ 10000以上
- 您这次旅程的主要目的是什么？ ☐ 出差 ☐ 旅游 ☐ 探亲访友 ☐ 其他
- 您将乘坐那种舱位？ ☐ 头等舱 ☐ 商务舱 ☐ 经济舱
- 您在过去12月内共有多少次航空旅程？ ☐ 1~2次 ☐ 3~5次 ☐ 6~10次 ☐ 11~20次 ☐ 20次以上

PPT适合图和表格的报告

带有 SmartArt 的标题和内容版式

明确问题

- 了解背景知识
- 大致了解数据

提出猜想

- 问题的假设
- 可能的解决方案

验证猜想

- 制定可行的研究方案
- 数据实现
- 修正方法和猜想

总结成文

- 写论文
- 写博客

Office报告与可重复研究技术

● 报告排版

- Word报告适合于文字比较多的文档
- Excel报告适合表格和问卷
- PPT报告适合图表混排的最终报告

● 可重复研究技术

- Office 报告中的任何对象都可以由VBA控制
- 通过设置模板，针对数据的变化进行更新，可以实现可重复研究

目 录

- 1 自动化报告与可重复研究
 - Office报告
 - L^AT_EX报告
 - 基于标记语言的报告
 - 报告系统
- 2 PPT和VBA
- 3 Rofficetool 简介

L^AT_EX简介

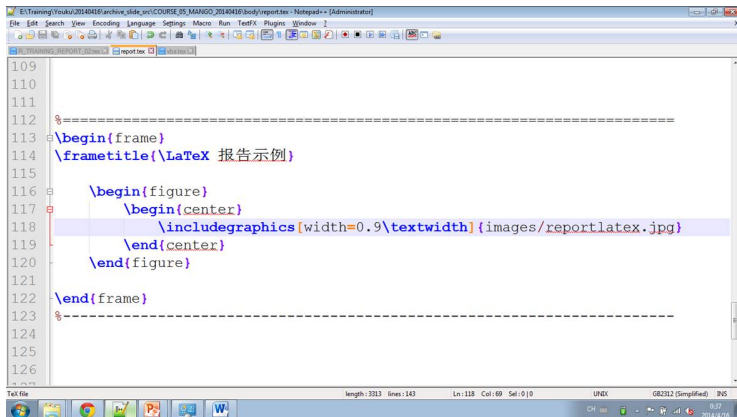
● 简介

- 注意书写规范和大小写
- 拉丁文发音，读作lay-tech
- 最伟大的科技文排版工具
- 符合印刷标准的专业排版系统

● 安装

- 下载CTeX: <http://www.ctex.org/CTeXDownload>
- Windows下默认安装; Linux下安装TexLive
- 使用Notepad++等文本编辑器进行编辑和编译

L^AT_EX报告示例



```
109
110
111
112 %=====
113 \begin{frame}
114 \frametitle{\LaTeX 报告示例}
115
116 \begin{figure}
117 \begin{center}
118 \includegraphics[width=0.9\textwidth]{images/reportlatex.jpg}
119 \end{center}
120 \end{figure}
121
122 \end{frame}
123 %-----
124
125
126
```

TeX file length: 3313 lines: 143 Ln: 118 Col: 69 Sel: 0 | 0 UNIX GB2312 (Simplified) BNS CH 0.37 2014/4/26

L^AT_EX报告与可重复研究

● L^AT_EX排版

- L^AT_EX可以认为是一种语言
- L^AT_EX是典型的非所见即所得排版方式
- 人只需要考虑内容，版式交由L^AT_EX解决
- 其基于语言的特性使得自动化报告非常便利

● 可重复研究技术

- 结合R 中内置的Sweave 技术可以实现基于L^AT_EX的自动排版
- 使用knitr 包可以实现灵活美观的自动化报告

目 录

- 1 自动化报告与可重复研究
 - Office报告
 - L^AT_EX报告
 - 基于标记语言的报告
 - 报告系统
- 2 PPT和VBA
- 3 Rofficetool 简介

轻量级的标记语言——Markdown

```
# knitr

[![Build Status](https://travis-ci.org/yihui/knitr.png)](ht

The R package knitr is a general-purpose literate progr
with lightweight API's designed to give users full control
without heavy coding work. It combines many features into c
slight tweaks motivated from my everyday use of Sweave. See
[homepage](http://yihui.name/knitr) for details and example
[FAQ's](https://github.com/yihui/knitr/blob/master/FAQ.md)
frequently asked questions (including where to ask question

## Installation

You can install the stable version on
[CRAN](http://cran.r-project.org/package=knitr):

...$
install.packages('knitr', dependencies = TRUE)
...

You can also install the development version from
[RForge](http://rforge.net/knitr/), which provides daily bu

...$
# update all existing packages first
update.packages(ask = FALSE, repos = 'http://cran.rstudio.c
install.packages('knitr', repos = c('http://rforge.net', 'h
type = 'source')
...

If you know GIT and 'R CMD build', here is another way:

...$
git clone https://github.com/yihui/knitr.git
R CMD build knitr
R CMD INSTALL knitr_*.tar.gz
...
```

knitr

build passing

The R package **knitr** is a general-purpose literate programming er the output without heavy coding work. It combines many features ii use of Sweave. See the package [homepage](#) for details and examp where to ask questions).

Installation

You can install the stable version on [CRAN](#):

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

You can also install the development version from [RForge](#), which i

```
# update all existing packages first
update.packages(ask = FALSE, repos = 'http://cran.rstudio.or
install.packages('knitr', repos = c('http://rforge.net', 'ht
type = 'source')
```

Markdown语法简介

● 标题

- # 表示一级标题
- 有几个# 表示几级标题

● 列表

- * 开头表示无序列表，换行加Tab 键表示下一级项目
- 类似1. 开头的项目表示有序列表

● 列表

- * 开头表示无序列表，换行加Tab 键表示下一级项目
- 类似1. 开头的项目表示有序列表

● 代码

- 以Tab 符或者至少四个空格开始的行表示程序代码

● 图形

- ![Foo](http://myhomepage.com/pic.png)

最潮的动态报告——html5



动态的html5 报告

● 动态效果

- html5 报告最大的特点就是动态可交互
- 报告的显示依赖浏览器，可以实现跨平台
- 结合js 技术可以开发出任意效果的可视化展现

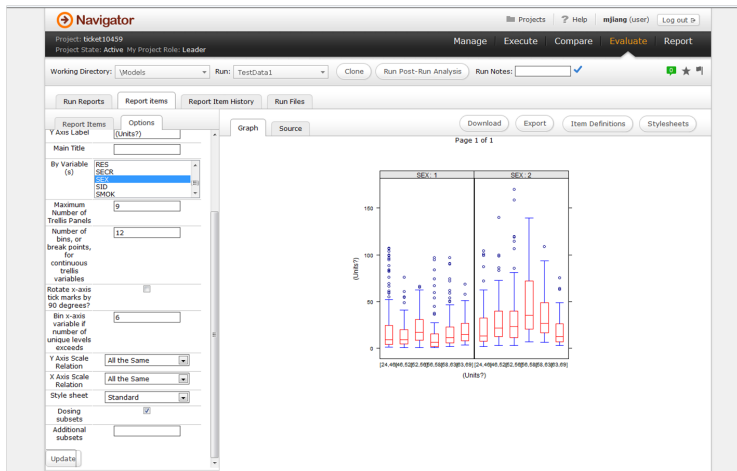
● 可重复研究

- 标记型语言非常方便集成自动化报告的功能
- knitr 提供了自动生成动态报告的功能

目 录

- 1 自动化报告与可重复研究
 - Office报告
 - L^AT_EX报告
 - 基于标记语言的报告
 - 报告系统
- 2 PPT和VBA
- 3 Rofficetool 简介

基于Java 和R 可以开发复杂的报告系统



目 录

1 自动化报告与可重复研究

2 PPT和VBA

- R2PPT
- ReporteRs

3 Rofficetool 简介

目 录

- 1 自动化报告与可重复研究
- 2 PPT和VBA
 - R2PPT
 - ReporteRs
- 3 Rofficetool 简介

R2PPT包简介

● 简介

- 基于COM方式操作PPT的R包
- 可以新建或写入幻灯片

● 常用功能

- 新建或打开PPT文件：PPT.Init / PPT.Open
- 添加幻灯片：PPT.AddBlankSlide / PPT.AddTitleSlide / PPT.AddTextSlide
- 在幻灯片中加入图片：PPT.AddGraphicstoSlide
- 在幻灯片中加入表格：PPT.AddDataFrame
- 应用模板文件：PPT.ApplyTemplate

R2PPT包示例

```
library(R2PPT)

# 新建文档
myPres <- PPT.Init(method="RDCOMClient")

# 添加封面
myPres <- PPT.AddTitleSlide(myPres, title = "PPT报告",
                             subtitle = "基于R2PPT")

# 文字
myPres <- PPT.AddTextSlide(myPres, title = "本报告包括",
                             text="文字区域 \r图形 \r表格")

# 表格
myPres <- PPT.AddTitleOnlySlide(myPres, title = "表格")
myPres <- PPT.AddDataFrame(myPres, df = head(iris),
                             row.names=FALSE, size=c(55, 150, 600, 300))
```

R2PPT包示例（续）

```
# 图形
jpeg(file = "D:\\testRplot1.jpeg")
hist(rnorm(100))
dev.off()
myPres <- PPT.AddTitleOnlySlide(myPres, title = "图形")
myPres <- PPT.AddGraphicstoSlide(myPres,
  file = "D:\\testRplot1.jpeg",
  size= c(55, 150, 600, 300))

# 关闭进程
myPres <- PPT.Close(myPres)
```

目 录

1 自动化报告与可重复研究

2 PPT和VBA

- R2PPT

- ReporteRs

3 Rofficetool 简介

ReporteRs包简介

● 简介

- <http://davidgohel.github.io/ReporteRs/index.html>
- 基于JAVA 方式操作PPT的R包
- 通过一个调用COM的JAVA库来实现，功能比较丰富
- 可以综合处理Word、PPT及html的报告

● 常用功能

- 新建或打开PPT文件：pptx
- 添加幻灯片：addSlide函数和slide.layout参数
- 在幻灯片中添加标题：addTitle / addSubtitle
- 在幻灯片中添加文本：addParagraph
- 在幻灯片中加入图片：addPlot
- 在幻灯片中加入表格：addTable

ReporteRs包示例

```
library(ReporteRs)
# 设置全文字体
options("ReporteRs-e" = 28)
options("ReporteRs-default-font"="Arial")
# 创建新的文档
pptx.file <- "document_example.pptx"
doc = pptx( title = "title" )
# 添加封面
doc = addSlide( doc, slide.layout = "Title Slide" )
doc = addTitle( doc, "PPT报告" )
doc = addSubtitle( doc , "基于ReporteRs")
# 添加文字
doc = addSlide( doc, slide.layout = "Title and Content" )
doc = addTitle( doc, "Texts demo" )
doc = addParagraph( doc, value = "haha")
```

ReporteRs包示例（续）

```
# 添加表格
doc = addSlide( doc,
  slide.layout = "Title and Content" )
doc = addTitle( doc, "Table example" )
doc = addTable( doc, data = iris[25:33, ] )

# 添加图形
doc = addSlide( doc,
  slide.layout = "Title and Content" )
doc = addTitle( doc, "Plot example" )
doc = addPlot( doc, function( ) hist(rnorm(100)) )

# 写入文件
writeDoc( doc, pptx.file)
```

目 录

1 自动化报告与可重复研究

2 PPT和VBA

3 Rofficetool 简介

- 解析PPT
- Rofficetool

目 录

- ① 自动化报告与可重复研究
- ② PPT和VBA
- ③ Rofficetool 简介
 - 解析PPT
 - Rofficetool

PPT对象的层次结构

- PPT应用 (PowerPoint.Application)
 - 这是一个抽象的对象，代表PPT的应用程序
- 幻灯片演示文件 (Presentations)
 - 通常也会简称为“幻灯片”
 - 每个对象对应一个具体的PPT文件
- 幻灯片页 (Slides)
 - 具体某一页的幻灯片，通常也简称为“幻灯片”
 - 每个对象对应一页幻灯片
- Shapes
 - 没有通俗的名称，但在PPT中，每一个可插入的元素都对应一个具体的Shape
 - Shapes 是PPT中最基础的操作对象

使用RDCOMClient包进行幻灯片结构的解析

```
library(RDCOMClient)

# 创建PowerPoint应用
pptapp <- COMCreate("PowerPoint.Application")

# 打开PPT文件
pptfile <- normalizePath("template.pptx",
  winslash = "/", mustWork = TRUE)
pptpre <- pptapp[["Presentations"]][1]$Open(pptfile)

# 查看PPT文件的信息
pptpre[["FullName"]]
pptpre[["PageSetup"]][["SlideHeight"]]

# 定位到幻灯片页面
pptslide <- pptpre[["Slides"]][1]$Item(1)

# 定位到Shapes
pptslide[["Shapes"]][["Count"]] # Shapes对象个数
shape1 <- pptslide[["Shapes"]][1]$Item(1)
shape1$Type() # 查看Shapes对象类型
```

Shapes对象简介

● 文字区域

- Placeholder对象，编号是“14”，内置的文本区域
- TextBox 对象，编号是“17”，插入的文本框
- AutoShape 对象，编号是“1”，插入的形状、艺术字等

● 图表类

- Chart对象，编号是“3”，Office作的图表
- Picture对象，编号是“13”，插入的图片
- Table对象，编号是“19”，表格对象
- Diagram对象，编号是“24”，插入的组织结构等对象
- Line对象，编号是“9”，插入的线
- Media对象，编号是“16”，插入的媒体文件

● 组

- Group对象，编号是“6”，组对象，最特殊的对象，可以包含其他的基础对象

Shapes对象简介

● 其他对象

- Callout 对象，编号是“2”
- Comment 对象，编号是“4”
- Freeform 对象，编号是“5”
- EmbeddedOLEObject 对象，编号是“7”
- FormControl 对象，编号是“8”
- LinkedOLEObject 对象，编号是“10”
- LinkedPicture 对象，编号是“11”
- OLEControlObject 对象，编号是“12”
- TextEffect 对象，编号是“15”
- ScriptAnchor 对象，编号是“18”
- Canvas 对象，编号是“20”

常用对象的操作

● 操作方式简介

- 基于COM 与VBA 交互
- 使用RDCOMClient 包
- 所有的对象都是名为COMIDispatch 的引用对象
- 使用 “[” 及 “\$” 获取方法和属性

● 文本类对象

- 通过 “Text” 属性对内容进行读取和修改

● 表格对象

- 通过 “Cell” 属性操作每一个单元格
- 可以自行开发函数和R中的Matrix 进行对应

● 图形对象

- Chart对象是Office中灵活的图表对象
- Picture 对象是外部的静态图
- Chart 的灵活性在R的环境下意义不大，而且还会有数据源方面的困扰，不建议在自动化PPT报告中使用该对象，最好是使用静态的Picture对象来代替

对象属性的修改

● 定位到具体的属性

- 利用VBA 中对象的层次结构定位到关注的属性
- 可以通过Office环境下VBE的提示框来寻找
- 更方便的方式是通过互联网搜索感兴趣的属性及方法

● 属性修改

- RDCOMClient 重写了“<-”运算符，可以使用R中常用的方式进行赋值
- 需要注意的是，修改属性时需要将该属性上层的对象赋值给一个变量，然后通过变量赋值的方式修改属性，注意不要直接从某个上层对象定位到该属性之后再直接赋值，否则会报错
- 确定需要修改的对象，编写函数将具体数值写入指定位置

目 录

- ① 自动化报告与可重复研究
- ② PPT和VBA
- ③ Rofficetool 简介
 - 解析PPT
 - Rofficetool

项目介绍

- 源代码

- GitHub: <https://github.com/lijian13/Rofficetool>

- 项目功能

- Office自动化报告的工具包
- 通过在PPT模板的基础上进行修改的方式生成复杂报告
- 对复杂的PPT对象进行解析

- 项目状态

- 初步的版本发布在GitHub上,
- 目前可以解析PPT的对象, 并能将Excel中的表格数据写入到PPT中的表格对象。
- 未来会增加表格和图片的转换、Excel图形的自动更新以及VBA的深入连接。

示例代码

```
library("Rofficetool")
pptfile <- "C:/pptfile.pptx"
list1 <- parsePPT(pptfile)
dF <- iris
excelfile <- "C:/excelfile.xlsx"
shtindex <- 1
rangelefttop <- c(5, 2)
writeDfToRange(dF, excelfile, shtindex, rangelefttop)
```

Thank you!

Homepage: <http://www.mango-solutions.com/>