



Stata自动化报告与 可重复研究

陈堰平



www.uone-tech.cn

友万科技

自我介绍

- 雪晴数据网创始人
- 北理工大数据创新学习中心业界导师
- 微软最有价值专家



大纲

- 什么是可重复研究
- Stata里的自动化报告功能
- 借助RStudio实现自动化报告

什么是可重复研究

多维晋升激励对地方政府举债行为的影响

贾俊雪, 张晓颖, 宁 静

2017年, 第7期:5-23页 [【下载全文和数据】](#) (759)

生产性服务业集聚与制造业升级

刘 奕, 夏杰长, 李 垚

2017年, 第7期:24-42页 [【下载全文和数据】](#) (719)

中国资本回报率嬗变之谜

许 捷, 柏培文

2017年, 第7期:43-61页 [【下载全文和数据】](#) (448)

最低工资标准的资源错配效应及其作用机制分析

刘贯春, 陈登科, 丰 超

2017年, 第7期:62-80页 [【下载全文和数据】](#) (499)

中国进口商品价格传导效应及其变动趋势：进口商品Armington替代弹性的视角

孙 飞, 吴崇宇, 陈福中



要解决的问题

- 复制 + 粘贴？
- 如果想输出成网页格式发布出来呢？
- 如果想插入更丰富的表现形式呢？
 - 动态图形
 - 视频

通过stata生成文档

- putpdf
- putdocx
- putexcel



用markdown生成文档

- dyndoc
- dyntext + pandoc
- 借助RStudio

关于Markdown

□ 语法说明

- <http://www.appinn.com/markdown>

□ 在线编辑器

- <https://pandao.github.io/editor.md>
- <https://www.zybuluo.com/mdeditor>

dyndoc举例



网页结构

```
<!DOCTYPE html>
<html>
<head>
<meta charset="UTF-8">
<title>页面标题</title>
</head>
<body>

<h1>我的第一个标题</h1>

<p>我的第一个段落。 </p>

</body>
</html>
```

我的第一个标题

我的第一个段落。



中文支持

```
<!DOCTYPE html>
<html lang="en-us">
<head>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="stmarkdown.css">

<script type="text/x-mathjax-config">
  MathJax.Hub.Config({tex2jax: {inlineMath: [['$', '$'], ['\(', '\)']]}});
</script>
<script type="text/javascript" async
  src="https://cdn.mathjax.org/mathjax/latest/MathJax.js?config=TeX-AMS_CHTML">
</script>
</head>
<body>
```



支持LaTeX公式

```
<!DOCTYPE html>
<html lang="en-us">
<head>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<link rel="stylesheet" type="text/css" href="stmarkdown.css">
<script type="text/x-mathjax-config">
  MathJax.Hub.Config({tex2jax: {inlineMath: [['$', '$'], ['\(', '\)']]}});
</script>
<script type="text/javascript" async
  src="https://cdn.mathjax.org/mathjax/latest/MathJax.js?config=TeX-AMS_CHTML">
</script>
</head>
<body>
```

动态标签

```
<<dd_do>>
```

```
<</dd_do>>
```

插入stata代码，并可以选择性地将结果显示在文档中

```
<<dd_display>>
```

在文档中显示某个stata表达式的值

```
<<dd_graph>>
```

插入图片

```
<<dd_include>>
```

将外部文件的内容插入到文档中

```
<<dd_ignore>>
```

```
<</dd_ignore>>
```

处理时忽略这两个标签内的代码

```
<<dd_skip_if>>
```

```
<<dd_skip_else>>
```

```
<<dd_skip_end>>
```

按某条件来显示文本

```
<<dd_skip_if: Stata expression>>
```

```
lines of text ...
```

```
<<dd_skip_else>>
```

```
lines of text ...
```

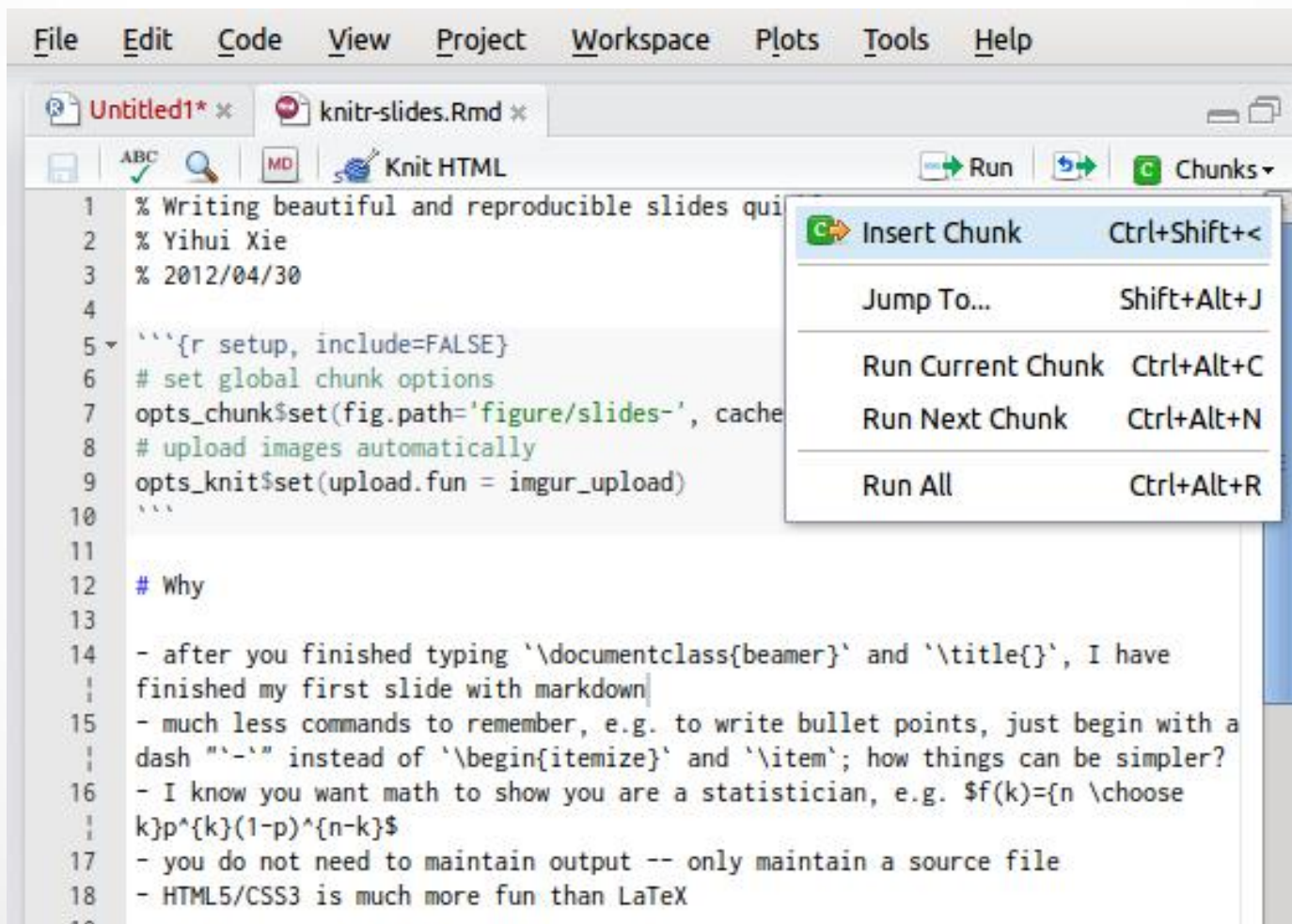
```
<<dd_skip_end>>
```



dyntext

- 先用stata里的dyntext处理markdown里的动态标签
- 再用markdown渲染工具来生成文档
 - pandoc
 - R(rmarkdown、knitr)
 - ruby(kramdown)

利用RStudio生成报告



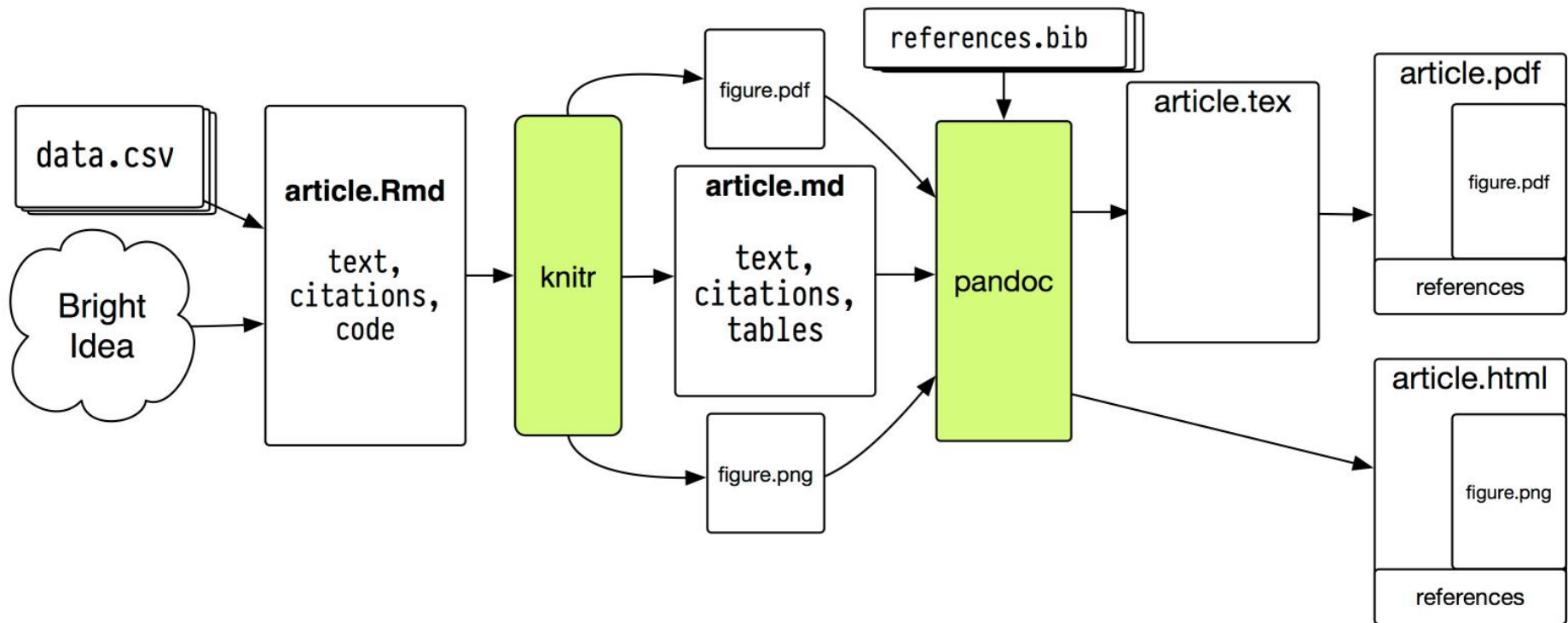
The screenshot shows the RStudio interface with a Knit HTML menu open. The menu options and their keyboard shortcuts are:

Option	Keyboard Shortcut
Insert Chunk	Ctrl+Shift+<
Jump To...	Shift+Alt+J
Run Current Chunk	Ctrl+Alt+C
Run Next Chunk	Ctrl+Alt+N
Run All	Ctrl+Alt+R

The background code in the editor is as follows:

```
1 % Writing beautiful and reproducible slides quickly
2 % Yihui Xie
3 % 2012/04/30
4
5 ```{r setup, include=FALSE}
6 # set global chunk options
7 opts_chunk$set(fig.path='figure/slides-', cache=TRUE)
8 # upload images automatically
9 opts_knit$set(upload.fun = imgur_upload)
10 ```
11
12 # Why
13
14 - after you finished typing `\\documentclass{beamer}` and `\\title{}`, I have
15   finished my first slide with markdown
16 - much less commands to remember, e.g. to write bullet points, just begin with a
17   dash "-" instead of `\\begin{itemize}` and `\\item`; how things can be simpler?
18 - I know you want math to show you are a statistician, e.g.  $f(k)=\\binom{n}{k}p^k(1-p)^{n-k}$ 
19 - you do not need to maintain output -- only maintain a source file
20 - HTML5/CSS3 is much more fun than LaTeX
```

knitr




[关于Stata与R Markdown](#)



环境设置

C:\Program Files (x86)\Stata15\stata.bat - Notepad++

文件(E) 编辑(E) 搜索(S) 视图(V) 编码(N) 语言(L) 设置(I) 工具(O) 宏

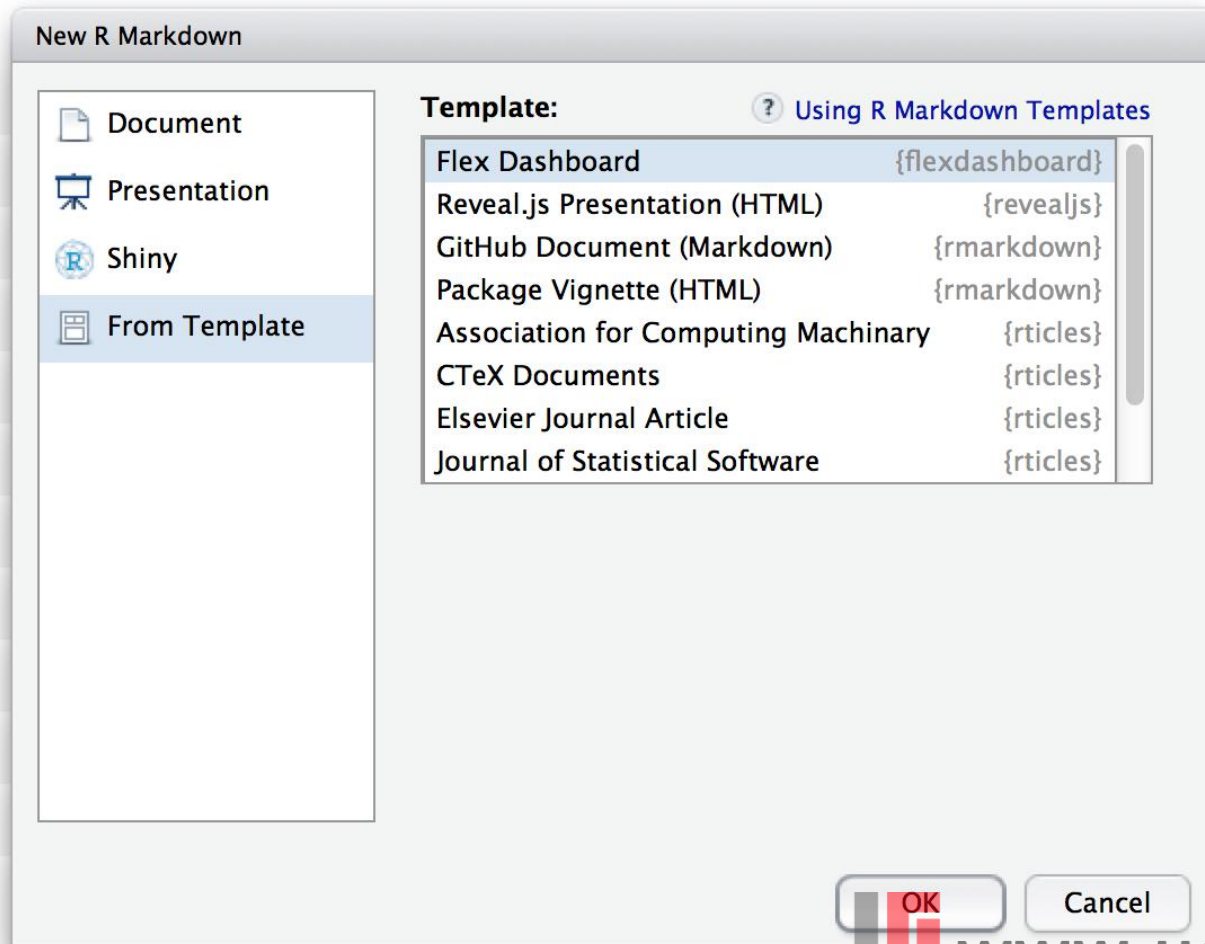


stata.bat

```
1 @echo off
2 call StataMP-64.exe %*
```

flexdashboard仪表盘

□ 安装 `install.packages("flexdashboard", type = "source")`



布局：单列

```
1 |---  
2 title: "Single Column (Fill)"  
3 output:  
4   flexdashboard::flex_dashboard:  
5     vertical_layout: fill  
6 |---  
7  
8 ### Chart 1  
9  
10 ```{r}  
11 ```  
12  
13  
14 ### Chart 2  
15  
16 ```{r}  
17 ```  
18  
19  
20  
21  
22  
23  
24  
25  
26
```

Chart 1

Chart 2



布局：单列可滚动

```
1 |---  
2 title: "Single Column (Scrolling)"  
3 output:  
4   flexdashboard::flex_dashboard:  
5     vertical_layout: scroll  
6 ---  
7  
8 ### Chart 1  
9  
10 ```{r}  
11 ```  
12 ```  
13  
14 ### Chart 2  
15  
16 ```{r}  
17 ```  
18 ```  
19  
20 ### Chart 3  
21  
22 ```{r}  
23 ```  
24 ```  
25  
26  
27  
28  
29
```

Chart 1

Chart 2

Chart 3



布局：多列

```
1 |---
2 title: "Multiple Columns"
3 output: flexdashboard::flex_dashboard
4 ---
5
6 Column {data-width=600}
7 -----
8
9 ### Chart 1
10 ```{r}
11 ```
12
13
14
15 Column {data-width=400}
16 -----
17
18 ### Chart 2
19 ```{r}
20 ```
21
22
23
24 ### Chart 3
25 ```{r}
26 ```
27
28
29
```



布局：按行布局

```
1 |---  
2 title: "Row Orientation"  
3 output:  
4   flexdashboard::flex_dashboard:  
5     orientation: rows  
6 ---  
7  
8 Row  
9 -----  
10  
11 ### Chart 1  
12  
13 ```{r}  
14 ```  
15  
16 Row  
17 -----  
18  
19 ### Chart 2  
20  
21 ```{r}  
22 ```  
23  
24  
25  
26 ### Chart 3  
27  
28 ```{r}  
29 ```  
30  
31
```

Chart 1

Chart 2

Chart 3



关于LaTeX

- 下载

<http://www.ctex.org/CTeXDownload>

- beamer 幻灯片

- markdown -> LaTeX -> pdf

插播广告



□ 活动宗旨

- 提升数据科学素养
- 交叉学科通用知识
- 拓宽学校与业界零对接教育渠道
- 探索多元化的创新教育模式

□ 学习模式

- 实战中学、互助中学、导师指导学
- 学中用、用中思、教学相长、共同进步
- 学术界专家与业界专家合作指导
- 兴趣激发、传帮接代
- 模块分割式学习



学习形式

实战学习

- ✓ 业界专家提供学习实践的小项目或小设计，由业界研发主力指导完成
- ✓ 业界提供研发核心岗位

竞赛活动

国际知名学者远程指导与交流

理论学习

学界专家列出学习计划（如参阅书籍、论文、线上线下资源等）
定期讨论交流



集中学习



小组学习



主题模块

计算机领域知识

Python、R 语言
数据存储、Linux
等

统计与数理模型

大数据前沿理论
数学理论模型

知识工程

知识图谱
自然语言处理



导师团队

高校导师

- 北京理工大学、北京大学、清华大学、
- 中国人民大学、东南大学、复旦大学等

国际导师

- 美国北卡罗莱纳大学教堂山分校
- 美国德克萨斯A&M大学
- 牛津大学

业界导师

- 上海海翼知信息科技、深圳狗尾草科技
- 猎聘网、融360、微软、人人车、滴滴出行等

Python实战课题

- 在处理淘宝数据实战中熟悉python机器学习工作流程
- 基于python的人脸识别简单实现
- 链家网房价数据分析
- 基于Python的信用卡评分
- 推荐系统python实践

学习汇报会



与部分参与汇报的学生合影



知识图谱实战（进行中）

- 基因信息的知识图谱
- 农业信息的知识图谱
- 旅游行业的知识图谱

联系我们

186-0002-1946
yanping.chen@xueqingtv.com

