·调查报告与分析·

基于 R Markdown 文档的疾病监测报告自动化生成方法及应用：以蚊媒监测报告为例

陈创端[1]，肖树[2,3]，肖建鹏[3]，芦瑞鹏[4]，容祖华[3]，曾韦霖[3]，赵建国[3]，龚德鑫[3]

1. 广东省普宁市疾病预防控制中心，揭阳 515300；
2. 暨南大学基础医学与公共卫生学院，广州 510632；
3. 广东省疾病预防控制中心 广东省公共卫生研究院，广州 511430；
4. 广东省疾病预防控制中心消毒与病媒生物预防控制所，广州 511430
通信作者：龚德鑫，E-mail：dxgong@gdiph.org.cn

【摘　要】本文提出了一种基于 R Markdown 文档的疾病预防控制机构监测报告自动化生成方法，以蚊媒监测报告为例，展示了从数据收集、分析、可视化到报告生成的全流程。通过整合文本、代码及输出结果，R Markdown 文档显著减少了人工操作，提高了报告的准确性和可重复性。本研究首次将 R Markdown 文档应用于蚊媒监测报告的自动化生成，为公共卫生领域提供了一种高效、可复用的解决方案。研究结果表明，该方法能够显著提升数据处理效率、增强报告灵活性，并为决策支持提供实时、可靠的数据支持。
【关键词】 R Markdown 文档；自动化报告；疾病监测；蚊媒监测；数据处理；可视化

Automating report generation for disease surveillance data using R Markdown: a case study on mosquito surveillance data

CHEN Chuangduan[1], XIAO Shu[2,3], XIAO Jianpeng[3], LU Ruipeng[4], RONG Zuhua[3], ZENG Weilin[3], ZHAO Jianguo[3], GONG Dexin[3] (*1. Puning Municipal Center for Disease Control and Prevention*, *Jieyang* 515300, *China*; *2. School of Medicine*, *Jinan University*, *Guangzhou* 510632, *China*; *3. Guangdong Provincial Institute of Public Health*, *Guangdong Provincial Center for Disease Control and Prevention*, *Guangzhou* 511430, *China*; *4. Institute of Disinfection and Vector Control*, *Guangdong Provincial Center for Disease Control and Prevention*, *Guangzhou* 511430, *China*)
Corresponding author: GONG Dexin, E-mail: dxgong@gdiph.org.cn

【Abstract】 This study proposes a method for automated report generation in disease control agencies using R Markdown and demonstrates the entire workflow from data collection, analysis, and visualization to report generation through a case study on mosquito surveillance data. By integrating text, code, and output results, R Markdown significantly reduces manual operations and improves report accuracy and reproducibility. This study represents the first application of R Markdown in the automated report generation of mosquito surveillance, offering an efficient and reusable solution for public health. The findings highlight the method ability to enhance data processing efficiency, report flexibility, and real-time decision-making support.
【Keywords】 R Markdown; automated report generation; disease surveillance; mosquito surveillance; data processing; visualization

疾病监测是疾病预防控制机构的重要工作，传统的疾病预防控制机构监测数据处理流程通常涉及数据收集、清洗、分析、可视化和报告撰写等多个独立步骤。然而，这一过程一般在不同软件间切换，高度依赖于重复性的人工操作，效率较低且易出错。此外，静态报告难以实时更新数据，限制了应急响应速度。在多人协作撰写复杂或紧急报告时，版本差异和同步问题进一步影响了团队协作效率，可能导致信息滞后或不一致。近年来，随着信息技术和人工智能的快速发展，自动化和动态报告生成技术逐渐成为解决上述问题的有效手段。R Markdown 文档作为一种基于 *R* 语言的文本标记工具[1]，能够将数据处理、分析和报告撰写整合到一个统一的框架中，实现从数据导入、清洗、分析到可视化及报告生成的全流程自动化[2−3]。与传统的报告生成方法相比，R Markdown 文档通过无缝集成代码、文本和输出结果，显著提高了工作效率，并确保了分析过程的透明性和可重复性[4]。尽管开源的 R Markdown 在计算科学[2]、统计学和教育等领域已得到广泛应用[5−6]，

开放获取：CC BY-NC-ND 4.0　DOI：10.11847/zgggws1147047
基金项目：广东省重点领域研发计划项目（2022B1111020006）；广东省自然科学基金面上项目（2023A1515011921）
第一作者：陈创端（1989—），主管医师，学士，研究方向：传染病预防控制。
收稿日期：2025-02-07　修回日期：2025-05-13　录用日期：2025-09-04
利益冲突：不存在　伦理审查：免除审查　出版授权：全体作者已与编辑部签署作者贡献声明及版权转让协议

但其在公共卫生领域的应用仍处于起步阶段，相关研究较为有限。本研究将 R Markdown 文档应用于疾病预防控制机构监测报告的自动化生成，以蚊媒监测报告为例，详细展示了从数据收集、分析、可视化到报告生成的全流程。

1 R Markdown 文档简介

R Markdown 文档是 *R* 语言生态系统中的一个重要工具，旨在创建动态文档，将文本、代码和输出结果有机整合[7]。与传统的报告生成方法相比，R Markdown 文档的核心优势在于其可重复性和自动化能力。通过将统计结果、静态图表、交互式地图等要素嵌入到叙述文本，R Markdown 文档极大地提高数据分析效率、保证数据准确性、增强报告实用性和可重复性[3,8–9]。

1.1 文档结构（图 1） R Markdown 文档的基本结构包括 YAML（标记语言“YAML Ain't Markup Language”的缩写）头部和 R Markdown 文档正文两部分。YAML 头部位于文档的最开始，用于定义文档的元数据（如标题、作者、输出格式等）[1]。R Markdown 文档正文部分则允许用户嵌入 R 代码块和 Markdown 文档文本（见 1.2）。新建的 R Markdown 程序文档结构和说明如图 1 所示。



图 1 R Markdown 程序文档基本结构截图及说明

Fig. 1 Basic structure and corresponding explanation of an R Markdown document

1.2 Markdown 文本（表 1） R Markdown 文档使用简洁易懂的 Markdown 文档标记语言语法，便于编写结构清晰、易于维护的文档。更多详细的 Markdown 语法细节可参考网站（https://markdown.com.cn）。表 1 总结了 R Markdown 文档的一些关键语法及其应用实例。

表 1 R markdown 文档的一些关键语法及其渲染效果

Table 1 Some key syntax of R Markdown and its renderings

标记类型	描述	R Markdown 文档输入内容	渲染效果
普通文本	两个空格结束一行以开始新段落	这是第一行；这是第二行	这是第一行 这是第二行
斜体	单个星号或下划线包围文本	*斜体*或_斜体_	斜体或斜体
粗体	双星号或双下划线包围文本	**粗体**或__粗体__	粗体或粗体
代码	反引号包围文本	`print(“Hello, world!”)`	print(“Hello, world!”)
下标/上标	波浪线或插入符号^包围文本	H~2~O和E=mc^2^	H_2O和$E=mc^2$
删除线	双波浪线包围文本	~~删除线~~	删除线

续表 1
Table 1 Continued

标记类型	描述	R Markdown 文档输入内容	渲染效果
转义字符	反斜杠转义特殊字符	*_\\	*_\
破折号	连字符创建破折号	–短破折号；—长破折号	–短破折号；—长破折号
方程	括在美元符号内	$A=r^{2}$	$A = r^2$
标题	井号表示标题	#标题1##标题2	标题1 标题2
链接	方括号加圆括号创建链接	[R Studio](https://www.rstudio.com)	R Studio
图片	感叹号加方括号和圆括号(图的路径)插入图片	![GitHub Logo](D:/github-mark/github-mark.png)	
列表	星号、加号或减号创建无序列表	*项目1 +项目2	•项目1 •项目2
代码块	3个反引号包围多行代码	```print(“Hello”) print(“world!”)```	print(“Hello”) print(“world!”)
编号列表	数字加点号创建有序列表	1. 第一项 2. 第二项	1. 第一项 2. 第二项
表格	竖线和破折号创建表格	列1|列2 –|– 内容|内容	列1　列2 内容　内容

1.3　R Markdown 文档核心功能(表 2)　包括参数化设计、多格式输出和交互式内容生成。通过参数化报告，只需调整时间、地区等变量，即可快速生成定制化报告，显著提升效率；支持 HTML、PDF、Word 等格式[1,10–11]，满足不同场景需求；结合 Shiny 和 Leaflet 等工具[1]，可嵌入交互式图表和地图。此外，R Markdown 文档依托 *R* 语言生态系统，无缝集成数据处理、分析和可视化工具，为报告生成提供集成解决方案。

1.4　与其他工具的比较　与 Jupyter Notebook、Microsoft Word/WPS 和 Tableau/Power BI 等工具相比，R Markdown 文档在一体化工作流、可重复性和输出灵活性方面具有显著优势，特别适合需要高效、重复和灵活报告生成的场景[2]。表 2 总结了 R Markdown 文档与其他工具的比较结果。

表 2　R Markdown 与其他工具比较
Table 2　Comparison of R Markdown with other tools

工具	一体化工作流	关键人工人操作步骤数[a]	学习曲线	适用场景	使用成本
R Markdown文档	高(集成代码、分析和报告)	1(参数录入)	中等偏高(需*R*基础)	可重复性报告，学术研究	免费(开源)
Jupyter Notebook	中(集成代码和文档，弱报告)	4(手动运行单元→清理输出→导出→排版)	中等偏高(需编程基础)	交互式分析，探索性研究	免费(开源)
WPS/微软Word	低(需VBA或插件扩展)	3(Excel分析→粘贴数据→调格式)	低(基础功能)或中(用VBA)	简单文档，基础报告	商业版需购买
Tableau/Power BI	低(专注可视化)	4(设置筛选器→导出数据→粘贴数据→排版)	低(可视化)或中(用函数)	商业可视化，实时仪表板	商业版需购买

注：a 人工操作步骤定义为从数据就绪至排版的非自动化操作(不含前期数据/代码准备)，测试基于疾控典型场景，测试用例为含2张图和5个表的8页蚊媒监测报告，核心数据为布雷图指数与诱蚊诱卵器指数(见2.2.1.1)。

2 蚊媒监测报告自动化生成实例（图 2）

疾病预防控制机构常需要基于监测数据定期生成固定报告，如日报、周报、月报和应急报告等。通过预先定义的报告模板与自动化数据处理脚本，可显著减少工作量。本文基于广东省普宁市蚊媒密度监测数据，展示如何利用 R Markdown 文档实现监测数据的自动化处理和报告生成全流程。

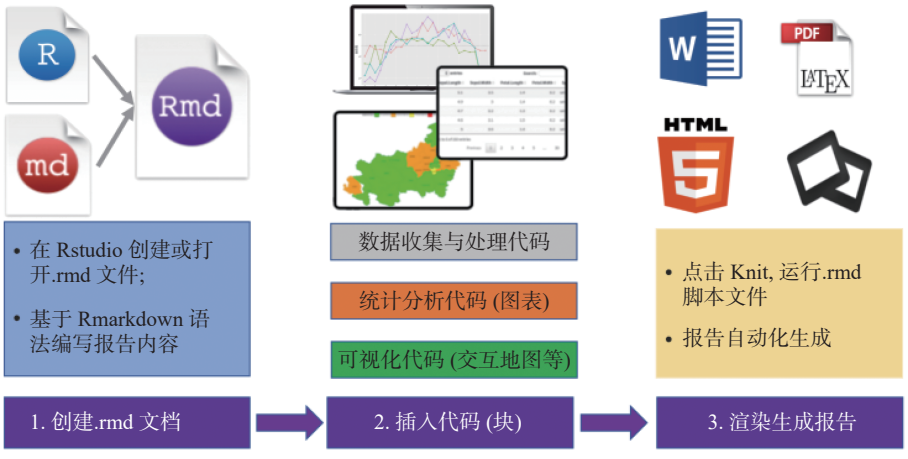


图 2 基于 R Markdown 文档的蚊媒密度监测报告自动化生成流程图

Fig. 2 Example flowchart of automated report generation of mosquito surveillance

2.1 创建 rmd 文档 Rstudio 中点击新建 R Markdown 文档即可创建 rmd 文档，预设目标输出文档类型和元数据，然后即可参考普宁市蚊媒密度监测报告的模板，基于 R Markdown 文档语法编写内容。

2.2 插入代码（块） 在创建的 rmd 文档中，可以直接嵌入数据处理和分析的全部代码，但在实际报告撰写中，推荐先独立编写数据处理和分析的 R 脚本，然后通过 source() 函数在 rmd 文件中调用 R 脚本及其结果。这种方法可保持代码模块化和可维护性，还确保报告的清晰与高效。

2.2.1 数据收集与预处理

2.2.1.1 数据来源 本文使用的普宁市蚊媒监测数据（来源于普宁市疾病预防控制中心），监测指标包括布雷图指数（breteau index，*BI*）和诱蚊诱卵器指数（mosquito ovitrap index，*MOI*），时间涵盖 2021 年 1 月—2024 年 11 月。首先用 R 分别读取 *BI* 和 *MOI* 文件：

```
files_bi<-list.files(folder, pattern="20\\d{2}年 BI 防控风险", full.names=TRUE)
files_moi<-list.files(folder, pattern="20\\d{2}年 MOI 防控风险")
bi_all<-do.call(rbind, lapply(files_bi, read_excel)) #合并 BI 数据
moi_all<-do.call(rbind, lapply(files_moi, read_excel)) #合并 MOI 数据
```

2.2.1.2 数据清洗和预处理 本研究基于病媒生物监测系统导出的结构化数据集进行分析，该数据已完成标准化清洗流程，故无需额外数据净化步骤。在数据转换阶段，采用 *str_extract* 函数提取时间与地区相关的变量，包括镇街名称、年份、月份以及区分上半月与下半月的标识。在此过程中，主要使用 dplyr 包实施数据操作，代码如下：

```
moi_all1<-moi_all%>%
mutate(乡镇=str_extract(地区, "^(.*?(街道'镇'乡'农场))")%>% gsub("广东省揭阳市普宁市", "",.),
year=str_extract(监测日期, "\\d{4}") %>% as.integer(),
month=str_extract(监测日期, "\\d{1,2}(?=月)") %>% as.integer(),
half_month=ifelse(str_detect(监测日期, "上半月"), "上", "下"))
```

2.2.2 统计分析 使用 *R* 语言编写 R 统计分析脚本，并将其嵌入 R Markdown 文档，以实现监测数据结果的自动化更新。分析的内容包括镇街的蚊媒监测指标（如 *BI*、*MOI* 等）均值、指标排序及增减变化、蚊媒密度等级构成比等，并生成相应的图表和地图。

2.2.2.1 描述性分析 以 2024 年 11 月下半月的报告为例，描述性统计分析 *BI* 和 *MOI* 数据，评估各监测点的蚊媒密度分布，计算各镇街的平均值。根据中国疾病预防控制中心发布的《登革热防治技术指南（2014 版）》[12]，监测项目包括幼虫密度（布雷图指数法 *BI*）和成蚊密度（诱蚊诱卵器法 *MOI*），以高密度（*BI* 或 *MOI*>20）、中密度（10<*BI*≤20 或 10<*MOI*≤20）、低密度（5<*BI*≤10 或 5<*MOI*≤10）、符合防控要求（*BI*≤5，*MOI*≤5）作

为风险评估依据，对各镇街的风险等级分类。

2.2.2.2 趋势分析（图 3A） 通过绘制当年及既往 3 年的月折线，直观展示普宁市伊蚊密度的变化趋势，并与往年同期比较，以识别潜在的风险上升期。图 3A 显示，2024 年 *BI* 指数呈现典型单峰分布，于 7 月接近峰值 8 后呈下降趋势，较 2021—2023 年相比，峰值略高、高峰持续时间稍长。主要绘图代码如下：

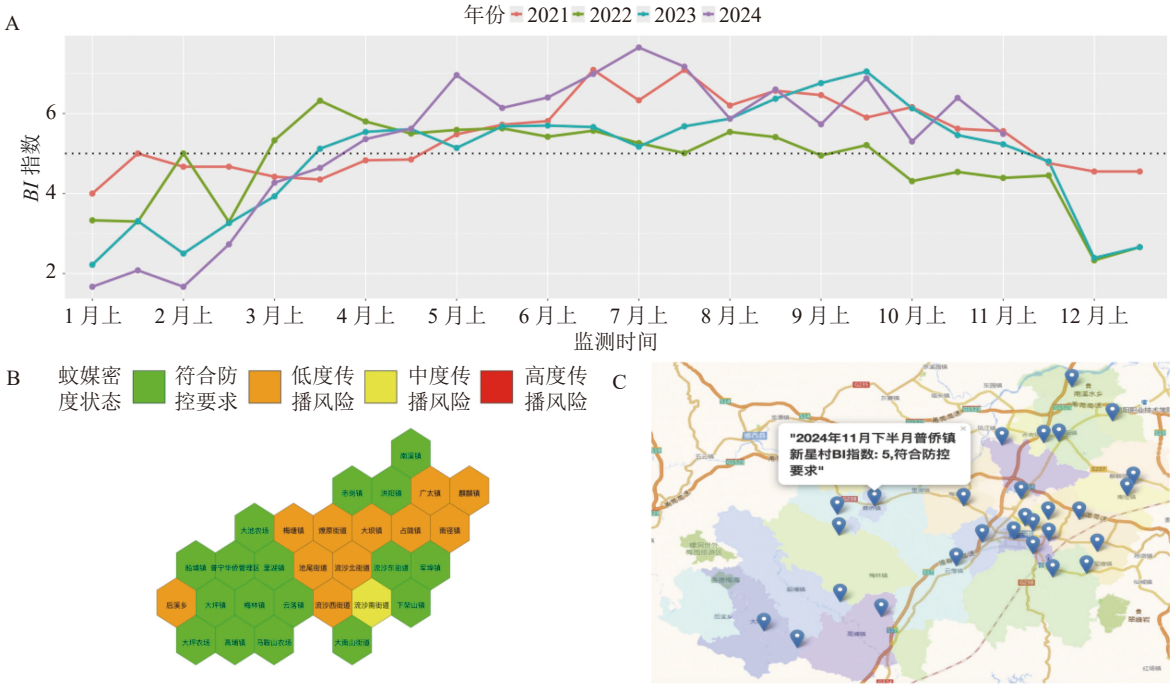
```
plot_bi_half1<-ggplot(Bi_half_months_all, aes(x=时间_半月, y=Bi, color=factor(year), group=factor(year)))+geom_line()+geom_point()+labs(x="监测时间", y="BI 指数", color="年份")
```

2.2.3 空间数据可视化（图 3B） 展示蚊媒监测数据的空间分布，用 *ggplot2* 包绘制了 *BI* 指数分布地图。地图直观显示出池尾、流沙北等街道为中度传播风险，提示有暴发风险，需密切关注。核心代码如下：

```
bi_map<-ggplot()+
geom_sf(data=guangdong, aes(fill=Bi 分类))+
scale_fill_manual(name="蚊媒密度状态",
values=bi_colors,
labels=c("符合防控要求""低密度传播风险""中密度传播风险""高密度传播风险"))+geom_sf_text(data=guangdong, aes(label=乡镇))
```

2.2.4 Rleaflet 交互可视化（图 3C） 本研究采用 Rleaflet 包，基于高德地图为底图生成交互式地图，可允许用户点击地图上的特定区域获取监测点位置、监测时间、数值等信息。相关代码如下：

```
MOI_puning<-leaflet(data=puning)%>%#调用空白地图为底图
amap(group="高德地图")%>%#调用高德地图
addPolygons(fillColor=topo.colors(12, alpha = NULL), stroke=FALSE)%>%
addMarkers(case_11 月下半月$lon, case_11 月下半月$lat,
popup=paste0("<b style='font-size:16px;'>", "\"", case_11 月下半月$监测日期, case_11 月下半月$地区, "BI 指数：", case_11 月下半月$BI, case_11 月下半月$蚊媒密度状态, "\"", "</b>")) #监测点时间、地区、BI 指数、风险等级显示。
```



注：图中展示非真实边界地图，为简化的演示图形。

图 3 *BI* 指数对比分析和空间可视化的呈现图

Fig. 3 Visualization for comparative analysis and spatial distribution of the *BI* index

2.3 渲染生成报告（图 4） 完成 R Markdown 文档编写后，点击 RStudio 中的“Knit”按钮即可导出为 HTML、PDF 或 Word 文档。文档支持参数化配置，便于快速生成符合需求的报告。在本案例中，引入全局参数 window_report 控制分析时间窗口，实现报告的自动化更新与批量生成，对应代码如下：

```
```{message=FALSE, warning=FALSE,include=FALSE} #隐形的代码执行
window_report<-"2024 年 11 月下半月" #分析时间为（默认既往 3 年）至 2024 年 11 月
source("code/data_analysis.R")
```
```



图 4 rmd 文档渲染生成 word 报告的效果展示

Fig. 4 Demonstration of R Markdown document rendering for word report generation

3 讨 论

本研究以蚊媒监测报告撰写为例，验证 R Markdown 文档在疾控机构报告自动化生成中的应用价值。经本团队采用标准办公环境（Intel i5-12500H/16GB RAM）对 15 份历史报告进行实证测算，传统人工流程（数据处理→分析→制图表→文本更新）耗时中位数为 1.5 个工作日（IQR：1.0～2.5 d），而 R Markdown 文档自动化流程在代码框架就绪后，同等质量报告生成耗时中位数仅 1.5 min（IQR：1.2～2.0 min），效率提升>99%。结果表明，R Markdown 文档通过整合数据处理、分析和报告撰写，在降低人工错误风险及保障分析透明度和可重复性的同时[2,13]，实现了分钟级报告产出，为突发公共卫生事件快速响应提供了技术基础。

与同类别工具相比，R Markdown 文档在工作流程整合度与人工操作效率方面均具有显著优势，与医院场景实证结论相互印证[14]。Jupyter Notebook 虽然支持交互式编程，但其报告生成功能较弱，且缺乏动态文档特性[15]；Microsoft Word/WPS 通过 VBA 脚本可以实现部分自动化，但数据—文档分离架构导致灵活性与可重复性不足；Tableau 和 Power BI 长于数据可视化，但在数据处理和报告生成方面功能有限[16–17]。相比之下，R Markdown 文档的一体化工作流程和动态文档特性，将人工操作精简至单一步骤（参数录入），兼顾自动化深度、灵活性、可重复性与格式一致性。

R Markdown 文档应用不仅限于病媒监测，其模板化思路还可适配传染病月报、疫情分析、病原学监测、环境卫生监测、医院统计分析等工作报告[14]。在数据兼容性上，R Markdown 文档支持结构化（如传染病报告病例数、环境监测指标）与非结构化数据（如舆情文本、实验室检测报告），例如在代谢组学领域[18]，R Markdown 文档通过整合报告指南清单和动态文档特性，提升了数据分析报告的标准化和可重复性。对于时空数据，如本案例中的 *BI* 和 *MOI* 指数，可结合 sf 包与 leaflet 生成交互式地图。技术扩展性方面，参数化设计（如动态调整分析时间和地区）支持环境监测机构实时关联气象数据与空气污染数据，自动生成超标预警报告。此外，通过预置模板库与领域专用脚本，R Markdown 文档能够显著降低迁移成本，成为跨学科监测的通用工具。

R Markdown 文档在自动化报告生成中虽具优势，但也存在挑战。比如，与易上手的 Tableau 和 Power BI 等工具相比，R Markdown 文档对用户的 *R* 语言编程技能有一定要求，可能成为初学者的门槛。在文档设计及格式化方面，尤其是中文处理，R Markdown 文档不如专业排版软件或深度定制的 Word。为此，可安装支持中文字符渲染 rticles 包，在创建 R Markdown 文档时选择专门支持中文的 CTeX Document 模板，在 YAML 中调节参数予以排版即可。

展望未来，R Markdown 文档在公共卫生领域的应用可以从以下 3 个方面进一步拓展：首先，在蚊媒监测方面，可结合机器学习技术开发智能分析模型，自动识别蚊媒密度异常变化，为疾病预防控制机构提供早期预警[19]。其次，利用 R Shiny

开发交互式报告模板，增强用户体验和报告可交互性。最后，基于疾病预防控制机构的具体需求，如传染病监测、慢性病管理、环境健康风险评估等，开发标准化的报告模板，并通过培训和提供技术支持推广 R Markdown 文档的应用，降低技术门槛。

数据可用性声明　读者可自行访问，用于方法验证或二次开发。本研究所使用的原始监测数据已由数据提供方（普宁市疾病预防控制中心）进行匿名化处理，其使用已获该中心授权，并严格遵守了国家及地方数据安全与隐私保护法规。相关数据处理方案已通过普宁市疾病预防控制中心伦理审查委员会的豁免。研究过程中开发的自动化分析代码、报告模板及模拟数据已公开存储在 Gitee 开源平台（https://gitee.com/dexin1215/mosquito-density-surveillance）。

参考文献

[1] Xie Y H, Allaire J J, Grolemund G. R Markdown: the definitive guide[M]. New York: Chapman and Hall/CRC, 2018.

[2] Peikert A, Brandmaier A M. A reproducible data analysis workflow with R Markdown, git, make, and docker[J]. Quantitative and Computational Methods in Behavioral Sciences, 2021, 1: e3763.

[3] Dvorak T, Halliday S D, O'hara M, et al. Efficient empiricism: streamlining teaching, research, and learning in empirical courses[J]. The Journal of Economic Education, 2019, 50(3): 242-257.

[4] Xie Y H. Dynamic documents with R and knitr[M]. 2nd ed. New York: Chapman and Hall/CRC, 2015.

[5] Baumer B, Cetinkaya-Rundel M, Bray A, et al. R Markdown: integrating A reproducible analysis tool into introductory statistics[J]. Technology Innovations in Statistics Education, 2014, 8(1): 1-29.

[6] Grayson K L, Hilliker A K, Wares J R. R Markdown as a dynamic interface for teaching: modules from math and biology classrooms[J]. Mathematical Biosciences, 2022, 349: 108844.

[7] Xie Y H. Bookdown: authoring books and technical documents with R Markdown[M]. New York: Chapman and Hall/CRC, 2016.

[8] Bauer P C, Landesvatter C. Writing a reproducible paper with R Markdown and Pagedown[EB/OL]. (2023-04-05)[2025-01-25]. https://osf.io/k8jhx.

[9] Rimal Y. Reproducible academic writing and interactive data visualization using R Markdown (*R* programming flex-dashboard: flex_dashboard packages)[M]//Rathore V S, Dey N, Piuri V, et al. Rising Threats in Expert Applications and Solutions: Proceedings of FICR-TEAS 2020. Singapore: Springer, 2021: 603-615.

[10] Auerbach C, Zeitlin W. SSD for R: an R package for analyzing single-subject data[M]. 2nd ed. Oxford: Oxford University Press, 2021: 155-166.

[11] Xie Y H, Thomas A, Hill A P. Blogdown: creating websites with R Markdown[J]. The American Statistician, 2019, 73(1): 94-104.

[12] 中国疾病预防控制中心. 登革热防治技术指南 [EB/OL]. (2014-09-29)[2025-01-25]. https://www.chinacdc.cn/jkyj/crb2/yl/dgr/jswj_dgr/201409/t20140929_298605.html.

[13] Mati S, Civcir I, Abba S I. Eviews R: an R package for dynamic and reproducible research using EViews, R, R Markdown and quarto[J]. The R Journal, 2023, 15(2): 169-205.

[14] 韩栋, 叶韵韶. R Markdown 在重复性医院统计分析报告中的应用 [J]. 中国医院统计, 2020, 27(6): 554-557.

[15] De Santana T L, Neto P A D M S, De Almeida E S, et al. Bug analysis in Jupyter Notebook projects: an empirical study[J]. ACM Transactions on Software Engineering and Methodology, 2024, 33(4): 1-34.

[16] Town P, Thabtah F. Data analytics tools: a user perspective[J]. Journal of Information and Knowledge Management, 2019, 18(1): 1950002.

[17] Carlisle S. Software: tableau and microsoft power BI[J]. Technology| Architecture + Design, 2018, 2(2): 256-259.

[18] Considine E C, Salek R M. A tool to encourage minimum reporting guideline uptake for data analysis in metabolomics[J]. Metabolites, 2019, 9(3): 43.

[19] 杨蕊, 龙佳敏, 王霞, 等. 我国登革热的流行特征及监测预警研究进展 [J]. 中华流行病学杂志, 2024, 45(2): 305-312.