

文章编号: 1000-5641(2019)01-0124-12

R 语言在生物科学研究绘图中的应用

蓝 洋, 何 秀, 朱诚勛, 张玉娟

(重庆师范大学 生命科学院, 重庆 401331)

摘要: R 语言具有强大的数据分析处理和可视化绘图功能, 可以在 Window、Linux 以及 Mac 系统上使用, 并且根据其编写新代码或调整已有代码可轻松实现科研中数据呈现与图形绘制的要求. 然而其代码学习较为艰难、R-package 使用复杂, 所以未受到科研新手的青睐. 基于生物科学相关领域的背景, 整理汇集已出版的文献、公共数据库以及国家统计局中提供的数据, 使用 R 语言、R 编辑器 RStudio 并载入相关的 R-package 绘制出地图、热图、关联网络图、韦恩图和柱形图等高质量的图片, 并提供相应的脚本与说明, 以方便生物科研人员直接更改使用. 本研究对生物科研中合理、直观地表述研究结果提供了良好的范例, 并进行了详细讨论, 且与其他绘图软件作了比较, 以期 R 语言能够成为生物科学领域科研工作者入门学习、研究应用中绘制图片的首选工具.

关键词: 地图; 热图; 关联网络图; 韦恩图; 柱形图; R 语言

中图分类号: Q811.4 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.2019.01.014

Application of R language graphics in biological research

LAN Yang, HE Xiu, ZHU Cheng-xu, ZHANG Yu-juan

(College of Life Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: The R programming language offers powerful statistical analysis, data processing, and visualization capabilities, which can run on Windows, Linux, and Mac operating systems; the software allows users to show scientific research data through graphics by writing new code or adjusting existing code. However, it is difficult to learn R's programming code and use R packages, so R is not commonly favored by novices in the scientific research community. Based on the context of bioscience fields and data from published papers, public databases, and the National Bureau of Statistics, this study demonstrates how to plot high quality maps, heat maps, functional networks, Venn diagrams, and column charts using R, R Studio, and R-packages. To facilitate the use of R, we offer detailed scripts for biological researchers. After comparing R with other graphing software, this study provides good examples and discussions of how to reasonably

收稿日期: 2017-12-06

基金项目: 国家自然科学基金(31871274); 重庆市教育委员会科学技术研究项目(KJ1600304);
重庆市科委基础研究与前沿探索项目(cstc2018jcyjA2487)

第一作者: 蓝 洋, 女, 硕士研究生, 研究方向为生物信息. E-mail: 908805060@qq.com.

通信作者: 张玉娟, 女, 教授, 硕士生导师, 研究方向为生物信息、化学计量组学、比较基因组学.
E-mail: zhangyj@cqnu.edu.cn.

and intuitively present results in biological research. We hope R will become the preferred plotting tool for elementary biological researchers in their learning and research activities.

Keywords: map; heat map; functional network; Venn diagram; column chart; R language

0 引 言

为了更加清晰且直观地反映生物科学研究中的结果, 论文中通常会将数据合理量化处理后绘制成图像. R 语言作为一种功能性编程语言兼具统计分析、数据处理和可视化绘图三大功能^[1]. R 语言与其他数据处理、分析与绘图工具不同, 其在具体操作时可以轻松地编写新代码或调整其他用户的代码, 适应特定的科研问题以实现用户的绘制要求^[2], 凸显图像绘制的科学性与新颖性. 因此合理掌握并使用 R 语言在生物科学研究中进行绘图, 不仅是论文中科学规律揭示的可视化保证, 更可以促进生物科学研究中的学术交流. 图形绘制一直是人类展示数据、解释规律的重要手段. 随着近年来科学技术的迅猛发展, 图形绘制软件被不断开发出来, 例如软件 OriginPro^[3]、SigmaPlot^[4]、Microsoft Excel 和 SPSS^[5]等均可绘制较为直观的矢量图. 然而这些绘图软件大多被应用于柱形图、折线图和散点图等简单图像的绘制, 在热图、地图等特异性、复杂性较高的图像绘制上显得捉襟见肘. 自 1992 年 Ross Ihaka 和 Robert Gentleman 开发 R 语言以来^[6], 应用 R 语言绘图已经出现在农业^[7]、生态学^[8]、植物学^[9]、癌症研究^[10-12]、医疗^[13,14]、生理学^[15]和种群遗传^[16]等众多领域. 在实际的 R 语言学习应用时, 虽然已有部分系统介绍 R 语言绘图的书籍, 如 An Introduction to R^[1]、R (programming language)^[17]和 An Overview of the R Language^[18]等, 但内容繁杂、艰深, 不能够快速专一地完成生物科学领域的绘图. 另一方面, 在 R 语言应用生物科学研究绘图的论文中, 大多针对某一类图像, 如箱形图^[19]、柱形图^[20]等; 部分集中在某一研究方面, 如藓类^[21]、橡胶林动态检测^[22]、差异表达基因检测^[23]和水文预测分析^[24]等, 还没有系统地介绍 R 语言应用于生物科技论文绘图方面的文章. R 语言之所以功能强大是因为具有数量众多的外源 R-package(R 扩展包), 截至 2017 年 3 月 CRAN(Comprehensive R Archive Network)包含 R-package 已超过 10 320 个^[25]. 不同功能的 R-package 有很多, 即使是相同功能的 R-package 也需要不同的参数进行调用, 这成为生物科学研究者绘制图片的困难之处. 本研究为 R 语言在生物科学论文中的绘图提供了源代码, 并对源代码进行了相应解释, 以方便生物领域科技工作者直接更改使用. 本论文对生物科研中合理、直观地表述数据结果展示了良好范例, 具有一定的借鉴意义.

1 材料与方法

1.1 R 软件的准备

R 语言安装程序包是从 <https://www.r-project.org/> 下载得到的, 版本为 R ×64 3.0.1(R 版本需要与 R-package 相匹配以方便操作). RStudio 安装程序包是从 <https://www.rstudio.com/> 下载得到. RStudio 是一种 R 语言编辑器, 可方便快捷地编写 R 脚本.

1.2 设置和查看路径

设置路径使 R 可以方便导入数据文件, 同时 R 语言绘制形成的图像也会保存在此文件夹下, 因此设置的路径要指向数据所在文件夹. 设置路径和查看当前路径的代码如下.

```
setwd("D:/R") #设置路径(在 R 中, 脚本中#后的内容为注释信息, R 不会执行注释信息)
```

```
getwd() #查看当前路径
```

1.3 R-package 的下载和使用

R 语言绘图功能的实现很大程度上依赖于不同的 R-package(具体如表 1 所示). 任何人都可以自由编写 R-package, 然后由 R 核心小组审核通过 CRAN 网站(<https://cran.r-project.org/>)发布^[26]. R-package 中包含了很多用户可以直接调用的函数^[27], 只要稍作修改, 就可以展示出不同的分析与绘图结果^[28].

表 1 10 种常用 R-package 信息介绍

Tab. 1 Introduction to common R-packages

序号	R-package	作用特点	开发时间	参考文献
1	ggplot2	使用图形语法创建数据可视化	2013 年 10 月	[29]
2	lattice	复合图像组合图形	2008 年	[30]
3	VennDiagram	绘制 Venn 图	2011 年 1 月	[31]
4	Metacoder	群体数据分类并以树形式可视化显示	2017 年 2 月	[32]
5	treeman	高效和直观地构建系统发育树	2017 年 1 月	[33]
6	DGCA	差异基因分析	2016 年 11 月	[34]
7	vcfr	操作和可视化 VCF 数据	2017 年 1 月	[35]
8	sfinx	消除亲和纯化-质谱数据集中的假阳性	2017 年 2 月	[36]
9	snpGeneSets	基因组学研究注释	2016 年 12 月	[37]
10	Gogadget	GO 富集结果的分析和可视化	2016 年 12 月	[38]

虽然 R-package 可以单独使用绘制简单图像, 但更多时候为了满足数据分析与复杂图形绘制的要求都会将 R-package 组合使用, 例如下文将介绍的热图、地图和关系图等. 需要注意的是, R-package 只需安装一次, 但是在每次使用之前, 必须将其重新载入. 下载和载入 R-package 的代码如下.

```
install.packages("ggplot2") #下载 ggplot2
```

```
library(ggplot2) #载入 ggplot2
```

2 结 果

2.1 地图绘制应用农业生态环境

地图是依据一定法则制图并体现在载体上, 并借以反映区域内的自然要素和社会经济现象的普通特征的图像. 因地图具有统一的数学基础; 按国家统一测量和编绘规定完成; 几何精度高且内容详细这三大特征^[39], 从而通常将其用于描述研究对象在地理位置的分布状况、时间或条件引起的在不同地区的改变等问题. 在生物科学研究中的生态方面, 常常需要反映某一地区和作物相关宏观生态状况, 如某一地区的蔬菜种植面积、果树分布状况、外来入侵植物分布状况、农业害虫分布情况、灾情分布及改善状况和麦区冬春气象干旱趋势^[40]等问题. 本研究则以近 10 年西瓜单位面积平均产量的数据为基础, 绘制出反映西瓜在各省份的平均产量地图. 数据来自国家统计局^[41]. 需要注意的是, 绘制地图时除了绘图数据, 还要下载一个地图的 shp 文件并将文件放在绘图路径文件夹下.

```
library(maps)
```

```
library(ggplot2)
```

```
library(maptools)
```

```
library(directlabels)
```

```
library(mapproj)
library(plyr)
library(reshape2)
library(ggsubplot)
setwd("D:/R/map")
mymap <- readShapePoly("bou2_4p.shp") #读取地图
mymapd <- fortify(mymap) #打散地图为数据框, 方便 ggplot 读取
temp <- mymap@data #提取地图省份
xs <- data.frame(temp, id = seq(0:923)-1)
china_mapdata <- join(mymapd, xs, type = "full")
healthdata <- read.csv("healthmap.csv", header=T, sep=",", stringsAsFactors = F) #读取数据信息
NAME <- healthdata$NAME #提取省份名称
hdoct <- healthdata$产量 #提取各省西瓜种植面积
mydata1 <- data.frame(NAME, hdoct)
mydata1 <- join(mydata1, xs, type = "full")
mydata1 <- data.frame(NAME = mydata1$NAME, 产量 = mydata1$hdoct, id = mydata1$id)
myepidat <- data.frame(id = unique(sort(mymapd$id)))
myepidat <- join(myepidat, mydata1, type = "full")
temp <- data.frame(NAME = healthdata$NAME, lat = healthdata$lat, long = healthdata$long)
theme_opts <- list ( theme (panel.grid.minor = element_blank(), #设置网格线为空
panel.grid.major = element_blank(),
panel.background = element_rect(fill=rgb(red = 255, green = 250, blue = 250, max = 255)), #设置图版背景色
plot.background = element_rect(fill=rgb(red = 255, green = 250, blue = 250, max = 255)), #设置绘图区背景色
panel.border = element_blank(),
legend.background = element_rect(fill=rgb(red = 255, green = 250, blue = 250, max = 255)),
axis.line = element_blank(),
axis.text.x = element_blank(), #设置 x、y 轴相关信息
axis.text.y = element_blank(),
axis.ticks = element_blank(),
axis.title.x = element_blank(),
axis.title.y = element_blank(),
plot.title = element_text(size=10)))
ggplot(myepidat)
+ geom_map(aes(map_id = id, fill = 产量), color = "white", map = mymapd)
+ geom_point (data = temp, aes (x = long, y = lat, fill = NULL), colour = rgb(red = 165, green = 165, blue = 165, max = 255))
```

```

+ geom_dl (data = temp, aes (x = long, y = lat, label = NAME), list('last.points', cex =
0.7, hjust = 1))
+ scale_fill_gradient (name="", high = "#006600", low = "#ffffcc")
+ expand_limits (x = c (73, 136), y = c (6, 54))
+ coord_map()
+ theme_opts
+ theme(legend.text=element_text(size=10))

```



图1 近10年33个省份西瓜平均产量

Fig.1 The average yield of watermelon crops across 33 provinces in the last 10 years

2.2 热图绘制应用遗传育种

热图(heatmap)是通过使用颜色渐变来显示数据矩阵的图像,其可以直观的显示出矩阵中数值的差异^[37].热图适用于研究实验数据的质量控制和差异数据,还可以对数据和样品进行聚类,观测样品质量^[42].目前生物科学研究中的遗传领域已经深入到分子水平,其大多基于基因组、转录组和蛋白质组等组学数据,绘制热图以初步体现不同样品中数据的关联性与差异性,从而便于下一步研究操作.基于基因组和蛋白质组数据,绘制热图分析其聚类的关联度与差异情况,可以找到显著性差异的基因或蛋白质;基于转录组的数据,用RNA-seq的数据绘制热图以分析表达量的情况,可以找到在某一时期或是某一条件下高表达、低表达或不表达的基因.本文绘图数据来自应激反应基因表达时间的自然变化预测拟南芥的种内杂种的杂种优势^[38],将原始数据log10处理后绘图.

```

library(mixOmics)
library(MASS)
library(lattice)
setwd("D:/R/heatmap")
mat<-read.csv("heatmap.csv", header = T) #读取数据信息

```

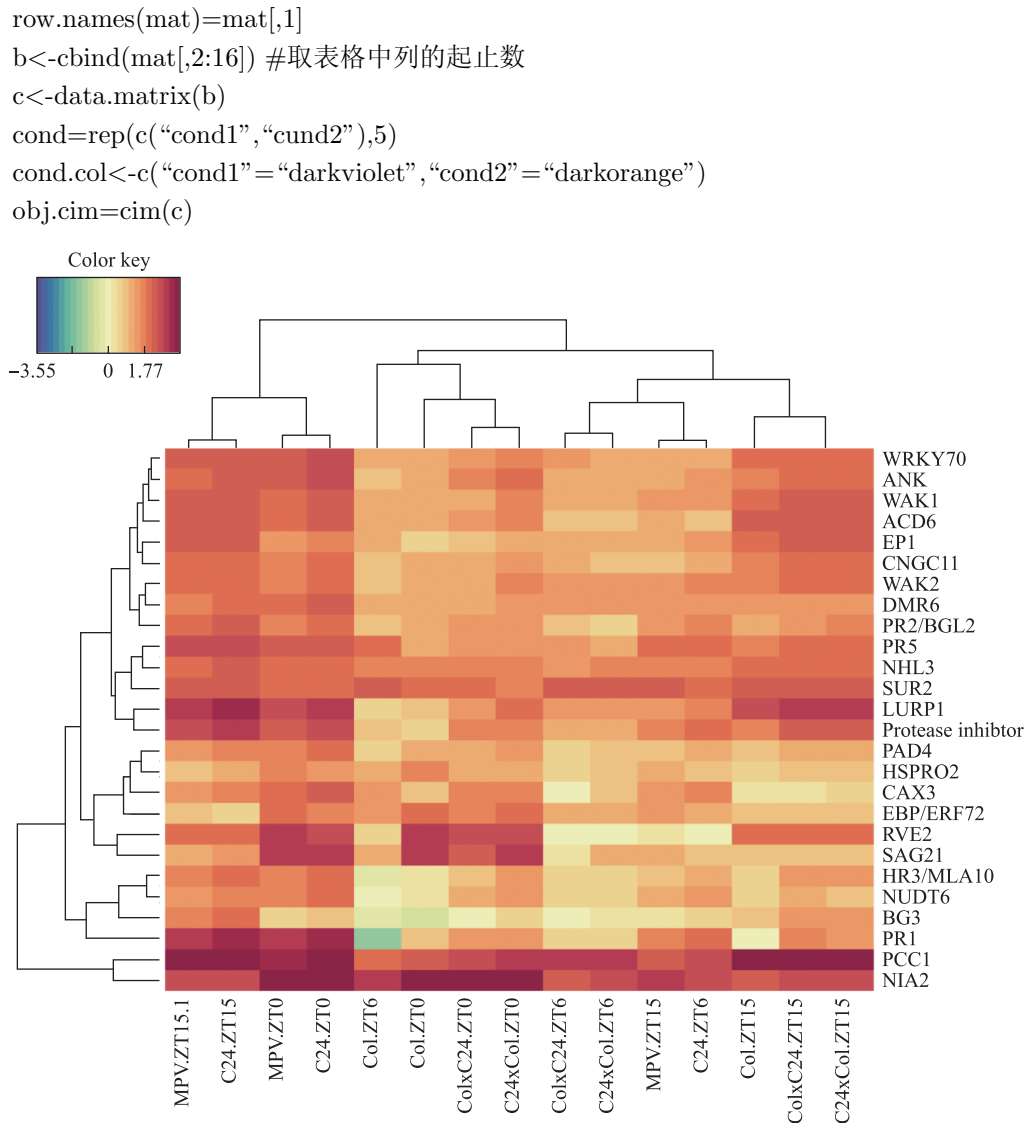


图2 生物胁迫 RNA-seq 在一天中 3 个时刻的表达量的变化

Fig. 2 RNA-seq data showing subsets of genes for biotic stress at three instances per-day

2.3 关联图绘制应用植物保护

关联图是把数据中各个对象根据一定的相关关系进行串联绘制出的图形. 关联图可以清晰的反映个体与群体间, 群体中个体和个体间存在的相关联系, 其适用于研究含有复杂关系的对象, 用以反映出各个对象之间的关系. 在生物科学研究中, 常常需要研究物质之间的关系或是物质与环境之间的关系, 如某些基因和蛋白质与抗性相关、某些基因或蛋白质对疾病发生至关重要、某些环境与植物生长状况的关系等问题. 在分析基因和蛋白质方面的数据时, 会查看与该基因或蛋白质有关的基因或蛋白质, 进而初步判断该蛋白质的功能等. 热激蛋白质(Heat stress proteins, Hsp)是生物为了应对环境的突然变化, 合成的一种特定的蛋白质, 可使生物有效应对外来胁迫^[43]. 本文选用玉米(*zea mays*) 的 *HSP18* 基因, 在 STRING(<http://string-db.org>)上查找玉米中与 *HSP18* 相互联系的蛋白质, 并导出数据, 绘制出与 *HSP18* 存在关系的蛋白质关联图.

```
library(igraph)
```

```

setwd("D:/R/cy")
dolphin <- read.csv('cy.csv', head=T, fileEncoding = 'UTF-8', stringsAsFactors=F) #读取数据信息
g <- graph.data.frame(dolphin)
source("http://michael.hahsler.net/SMU/ScientificCompR/code/map.R")
tiff(filename='dolphins_curve1.tiff', width=800, height=800, units='px') #设置输出图片大小以及格式
layout=layout.fruchterman.reingold
plot(g, layout=layout, vertex.size=map(degree(g), c(20, 40)), vertex.color="#99CC66",
vertex.frame.color="#FFFF99", edge.color="#996633", edge.arrow.size=0) #设置图形颜色
dev.off() #绘图结束后图片保存在设置路径的文件夹下

```

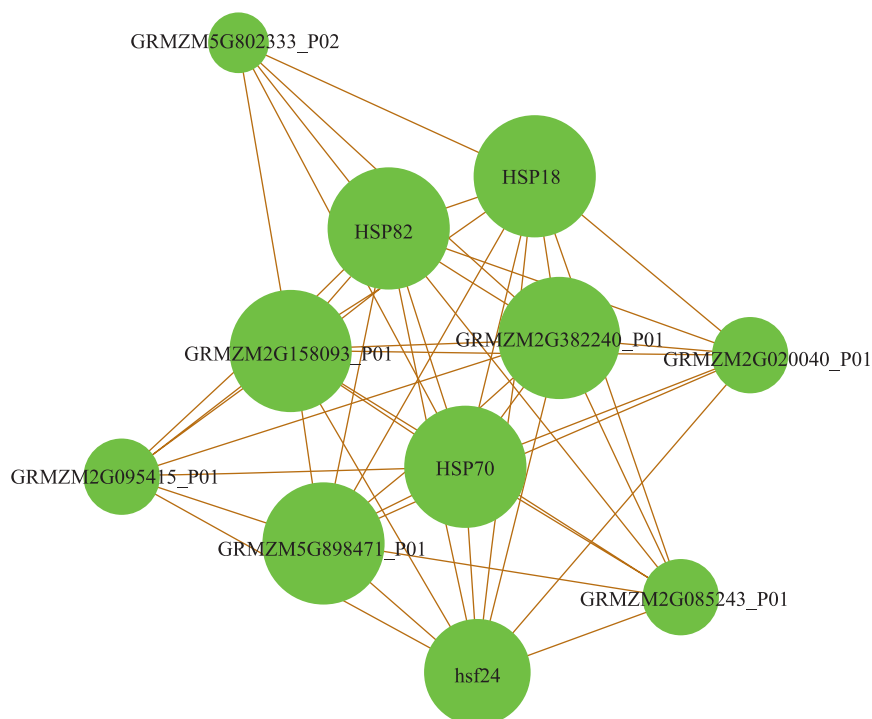


图3 HSP18 相互作用蛋白质

Fig. 3 Proteins interacting with HSP18

2.4 韦恩图绘制应用资源昆虫

韦恩图是以图形表示集合的重要方式^[44], 其通过表示各集合间的交集和并集情况, 可以清晰地反映数据集间的关系, 以体现其共性基数或个性基数的情况^[45]. 在生物科学研究中, 无论是通过宏观数据体现不同地区昆虫的种类分布特点, 还是在微观视角下通过宏基因组研究人类的肠道微生物, 抑或基于转录组水平分析家畜在不同处理下的生长繁殖情况, 这些均需要以韦恩图作为可视化对象反映数据之间的交集等情况. 如资源昆虫对不同环境的抗逆性不同, 其表达的基因也有所差异. 利用维恩图描述该种昆虫在不同时间或抗性环境中表达转录数据、差异表达的蛋白质数据信息, 即可筛选出重要功能的基因或蛋白质. 本文以黑腹果蝇幼虫的低温存活和贮存的生理基础^[46]的数据为基础绘制韦恩图.

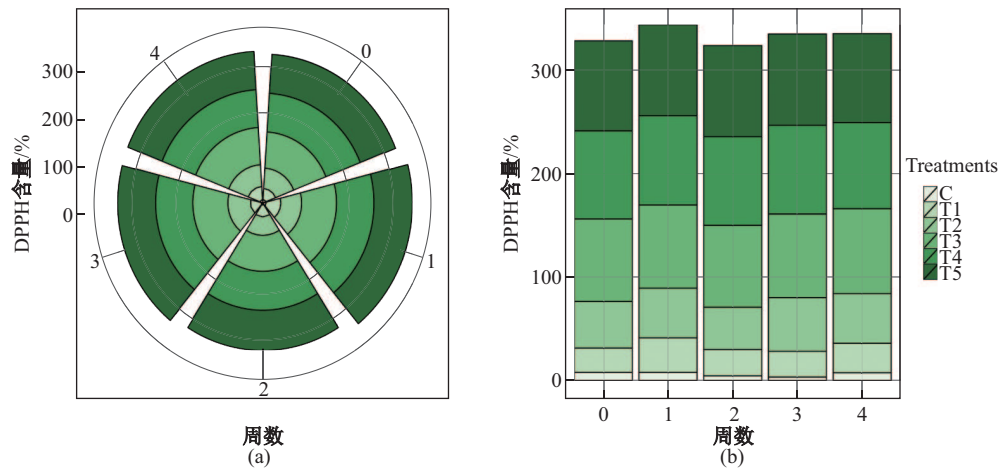
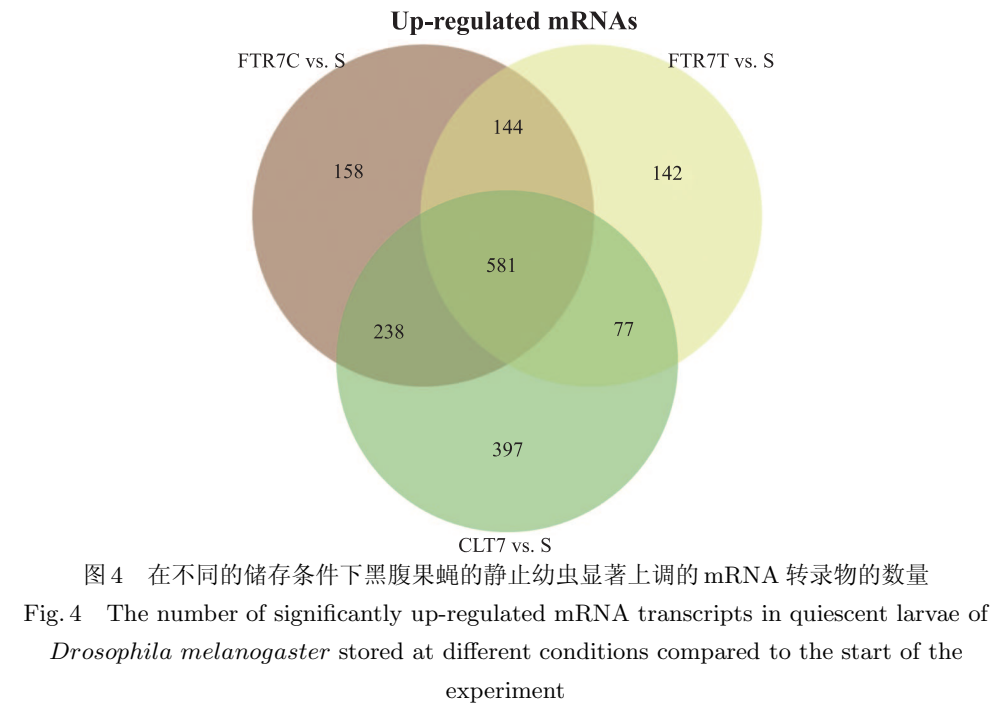
```
library(grid)
```

```
library(VennDiagram)
setwd("D:/R/VE")
venn<-read.csv("venn.csv", header = T) #读取数据信息
a<-venn$FTR.7C.vs..S #选择绘制图形所需列
b<-venn$FTR.7T.vs..S
c<-venn$CLT.7.vs..S
venn.plot<-venn.diagram(
list("FTR7C vs. S"=a,"FTR7T vs. S"=b,"CLT7 vs. S"=c),
col = "transparent",
fill = c("#996633", "#FFFF99", "#99CC66"), # venn图颜色
alpha =0.5,
scaled = TRUE,
ext.text = TRUE,
ext.line.lwd = 2,
ext.dist = -0.15,
ext.length = 0.9,
ext.pos = -4,
inverted = TRUE,
cex = 1.7, #图中数字大小
cat.cex = 1.5, #图中 a、b、c 字体大小
filename = "1.tiff", #输出文件名以及格式
main = "Up-regulated mRNAs", #图标题, 正中间
main.cex = 2, #图标题字体大小
sub.cex = 1) #绘图结束后图片保存在设置路径的文件夹下
```

2.5 柱形图的绘制应用加工保鲜

柱形图是通过柱子的高低直接反映不同样品数据差异的图形, 其在科研中是最为常见、直观反映数据的一类重要图像. 在生物科学研究中, 常常用柱形图来反映不同食品处理方式对食品感官、营养和生化指标等方面的差异状况. 在食品加工与保鲜方面, 常会探究不同包装方式和贮藏温度对食品中基本指标造成的变化, 如自由基含量、pH 值、酸价和亚硝酸盐等生化指标, 形成的微生物数量和种类变化等生物指标. DPPH(2,2-二苯基-1-苦肟基自由基)是一种很稳定的以氮为中心的自由基, 可以测定抗氧化物质含量高低^[47-48], 在食品的加工、保鲜方面有广泛的应用. 本文绘制累积直方图的数据来自苏木(*Caesalpinia sappan* L.)提取物对冷藏期间猪肉肠的理化性质的影响^[49]. 提供两种堆积图形, 一种传统的堆积柱形图(见图 5(a)), 清晰明了地反映数据情况; 一种“玫瑰花环”柱形图(见图 5(b)), 更为新颖、美观地反映出数据情况, 吸引读者阅读.

其中, 图形的横坐标表示周数, 纵坐标表示 DPPH 百分含量, 图例表示不同的处理方式.



2.5.1 传统柱形图

```
library(reshape2)
library(ggplot2)
setwd("D:/R/3Dplot")
rose<-read.csv("plot.csv") #读取数据信息
rose_long<-melt(rose,id.vars="Treatments",variable.name="weeks",
value.name="DPPH(%)")
```

```
p<-ggplot(rose_long,aes(x=weeks, y="DPPH(%)", fill=Treatments))+geom_bar(stat="identity", color="black") #边框颜色
```

```
p+ scale_fill_brewer(palette="Greens") + theme_bw() #“Greens”图形颜色
```

2.5.2 “玫瑰花环”柱形图

```
library(reshape2)
```

```
setwd("D:/R/3Dplot")
```

```
rose<-read.csv("plot.csv") #读取数据信息
```

```
rose_long<-melt(rose,id.vars="Treatments",variable.name="weeks",
value.name="DPPH(%)")
```

```
p<-ggplot(rose_long,aes(x=weeks,y="DPPH(%)",fill=Treatments))+
geom_bar(stat="identity", color="black") #边框颜色
```

```
p+coord_polar()+ scale_fill_brewer(palette="Greens") + theme_bw()
```

3 讨 论

生物统计分析、绘图软件众多,难以选择.虽然有些软件在某些方面较为出众,但不够广泛,科研工作者难以花费大量精力学习大量软件.这就需要一款有强大技术支撑的开源软件,可以普遍地适用于各类数据,绘制出高质量的图片.在众多统计绘图软件之中,R已经成为了绘图软件的首选,其使用面广,更适用于生物科研领域的各类数据绘制图像.

3.1 R 语言绘图的功能实现

由上文使用 R 语言实现生物科研绘图的过程来看,其清楚、便捷,能够明确说明相关问题并实现功能.通过其实现的地图绘制既是对生物科研基础数据的准确、直观描述,又能够通过整体与部分的情况充分展示统一性与差异性,以方便生物科研人员与统计决策者更好地进行分析研究、制定政策.而绘制的热图既可以通过数据聚类找到数据之间的相关性,而且可以通过对差异性的清晰反映,以方便生物科研工作者迅速筛出重要基因进行进一步研究.关联图的绘制既可以反映了复杂的网络关系,又展现了该网络关系中各组分之间的重要程度,方便生物科研人员更好地分析数据,做出正确判断.此外,维恩图绘制方便了资源昆虫中重要靶标基因的筛选,以完成基础生物学研究.R实现的柱形图绘制更是可以将食品的各类指标综合处理,清晰展示.因而探究发现,R语言可以贴合生物科研的数据有效实现其绘图功能.

3.2 科研绘图软件比较

现代生物科研论文绘图软件主要有 Excel、SPSS、OriginPro、SigmaPlot、Python 等.Excel 可谓是最常见的统计绘图软件,其优势就是操作简单、使用方便,并且科研工作者对 Excel 非常熟悉.但是 Excel 绘图实现的图形类型十分有限,无法绘制复杂图形,且图片的灵活性很低,仅可对颜色等进行略微调整,无法任意的改变.SPSS 作为一款统计软件,其统计功能非常强大,但是用于绘图不尽人意.与 Excel 一样,SPSS 操作简单,但出图速度缓慢,图片不够清晰、美观.而 R 出图速度快,回车之后立即见图,图片的美观程度是绘图者决定的,图片质量高,可以输出 tiff、png、jpeg 和 pdf 等 8 种图片格式.

SigmaPlot 与 OriginPro 相似,二者均是为友好的用户界面,可以轻松地从 Excel 中提取数据,操作简单且绘图功能强大.但是,两者均为付费软件,需要支付昂贵的使用费.Python 和 R 一样也是一门编程语言,但是 Python 与 R 最大的区别在于 Python 的数据统计分析是通过第三方 package 来实现的,且 Python 是并不是专化的数据统计与绘图分析的语言,因此在数据处理、统计分析、绘图方面略逊一筹.R 的统计函数包罗万象,无论是经典还是前沿的方法,都可以直接使用相应的 package 调用,相比 Python 在这方面贫乏不少.综上所述,R 应当成为生物科研中的首选.

3.3 R 绘图的优缺点

R 语言为开源性软件^[1], 其开源有两层意义: 首先 R 可以免费获得, 其在所有硬件和操作系统上安装均没有限制^[50], 适合各个领域各种专业背景的人使用; 其次, 任何人均可自由检查或修改源代码^[50], 以匹配各行各业的数据处理与图像绘制需求. 以上两点致使 R 语言在科研工作使用率已经越来越高, 并逐渐成为科研人员必备技能之一. R 为编程语言, 其统计绘图的灵活性也体现在此. 科研工作者可以根据自己需求编写脚本、R-package. 即便是没有任何编程功底, 仅仅想使用也可以在 CRAN 上找到相对应的 R-package. R 的灵活性还体现在图片中的任何属性都可以任意修改, 不仅仅是颜色、形状, 还可以对将数据分组, 也就是说, R 绘图取决于科研工作者对图形理解.

R 功能强大, 但是也有一些图片难以绘制. R 绘图是以统计分析数据为基础的, 对于绘制结构式(如蛋白质结构图)、示意图(如磷脂双分子示意图)、图片数据分析(如测量拟南芥照片中角果长度)、照片美化(如电泳图添加文字或裁剪拼接)、质粒图等这些不是基于数据的统计和分析绘制出的图片, R 也无能为力. 只能运用 photoshop、Illustrator、corelDraw 和 3dsMax 等相关软件进行绘制美化.

4 结 论

R 作为一门现代统计绘图的工具, 不仅在统计方面有很强大的功能, 而且绘图方面具有极强的专业性. 本文以生物科研的相关领域为切入点, 使用 R 实现了不同图像的绘制, 既证实了 R 作为免费开源的软件可完整实现生物科研的多样化图像绘制, 又为图像的绘制与 R-package 的调用提供了代码范例.

[参 考 文 献]

- [1] DANIELSSON J. An Introduction to R[M]. System Dynamics Modeling with R. New York: Springer International Publishing, 2016.
- [2] CARSON M A, BASILIKO N. Approaches to R education in Canadian universities[J]. F1000research, 2016(5): 1-18.
- [3] SEIFERT E. OriginPro 9.1: Scientific data analysis and graphing software-Software review[J]. Journal of Chemical Information & Modeling, 2014, 54(5): 1552.
- [4] WASS J A. SigmaPlot 11: Now with total sigmaStat integration[J]. Scientific Computing, 2009, 26(1): 21-25.
- [5] MCCORMICK K, SALCEDO J. SPSS statistics for data analysis and visualization[J]. Drug Testing & Analysis, 2017, 1(6): 250-266.
- [6] GENTLEMAN R, HUBER W, CAREY V J. R Language[M]. New York: Springer Berlin Heidelberg, 2011.
- [7] GANCHEV D H. Using R language for statistical computing for pesticide application calculations[J]. MAYFEB Journal of Agricultural Science, 2016(1): 10-26.
- [8] FELD C K, SEGURADO P, GUTIERREZ-CANOVAS C. Analysing the impact of multiple stressors in aquatic biomonitoring data: A 'cookbook' with applications in R[J]. Science of the Total Environment, 2016, 573: 1320-1339.
- [9] GUENZI D, FRATIANNI S, BORASO R, et al. CondMerg: An open source implementation in R language of conditional merging for weather radars and rain gauges observations[J]. Earth Science Informatics, 2016, 10(1): 1-9.
- [10] LI J, WANG J, CHEN Y, et al. A prognostic 4-gene expression signature for squamous cell lung carcinoma[J]. Journal of Cellular Physiology, 2017, 232(12): 3702-3713.
- [11] ZHAI J, HSU C H, DAYE Z J. Riddle for sparse regression with mandatory covariates with application to the genetic assessment of histologic grades of breast cancer[J]. BMC Medical Research Methodology, 2017, 17(1): 12-24.
- [12] WANG Z, ZHANG C, LIU X, et al. Molecular and clinical characterization of PD-L1 expression at transcriptional level via 976 samples of brain glioma. Oncoimmunology[J], 2016, 5(11): e1196310.
- [13] OH D H, KIM I B, KIM S H, et al. Predicting autism spectrum disorder using blood-based gene expression signatures and machine learning[J]. Clin Psychopharmacol Neurosci, 2017, 15(1): 47-52.

- [14] WOJCIECHOWSKI J, HOPKINS A M, UPTON R N. Interactive pharmacometric applications using R and the shiny package[J]. *CPT Pharmacometrics Syst Pharmacol*, 2015, 4(3): e00021.
- [15] LARRIBA Y, RUEDA C, FERNANDEZ M A, et al. Order restricted inference for oscillatory systems for detecting rhythmic signals[J]. *Nucleic Acids Research*, 2016, 44(22): e163.
- [16] PARADIS E, GOSSELIN T, GRUNWALD N J, et al. Towards an integrated ecosystem of R packages for the analysis of population genetic data[J]. *Molecular Ecology Resources*, 2017, 17(1): 1-4.
- [17] MILLER F P, VANDOME A F, MCBREWSTER J, et al. R (programming language)[J]. *Betascript Publishing*, 2010, 6(2): 36-40.
- [18] CHAPMAN C, FEIT E M D. An Overview of the R Language[M]// CHAPMAN C, FEIT E M D. *R for Marketing Research and Analytics*. New York: Springer International Publishing, 2015.
- [19] 王怀亮. 箱须图在识别统计数据异常值中的作用及 R 语言实现[J]. *商业经济*, 2011(5): 64-65.
- [20] 王怀亮. 基于 R 语言的统计数据柱形图的实现[J]. *电子技术*, 2013(8): 78-80.
- [21] 石蕾. R 语言在藓类形态与遗传变异研究中的应用[D]. 上海: 上海师范大学, 2015.
- [22] 兰国玉, 陈伟, 王继坤, 等. R 语言在橡胶林动态监测地形图绘制方面的应用[J]. *热带农业科学*, 2013, 33(3): 50-53.
- [23] 纪相禹. 基于 R 语言的差异表达基因检测研究[D]. 长春: 吉林大学, 2011.
- [24] DOBRE G R. R Language: Statistical computing and graphics for modeling hydrologic time series[J]. *Mathematical Modeling in Civil Engineering*, 2014, 10(4): 9-18.
- [25] CRAN. The Comprehensive R Archive Network [DB/OL]. [2017-05-01]. <https://cran.r-project.org/>.
- [26] HORNIK K. The comprehensive R archive network[J]. *Wiley Interdisciplinary Reviews Computational Statistics*, 2012, 4(4): 394-398.
- [27] IHAKA R, GENTLEMAN R. R: A language for data analysis and graphics[J]. *Journal of Computational & Graphical Statistics*, 1996, 5(3): 299-314.
- [28] VANCE A. Data analysts captivated by R's power[N]. *New York Times*, 2009-01-06 (Business Computing).
- [29] ITO K, MURPHY D. Application of ggplot2 to pharmacometric graphics[J]. *Cpt Pharmacometrics & Systems Pharmacology*, 2013, 2(10): 1-16.
- [30] SARKAR, DEEPAYAN. *Lattice: Multivariate Data Visualization with R*[M]. New York: Springer, 2008.
- [31] CHEN H, BOUTROS P C. VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R[J]. *Bmc Bioinformatics*, 2011, 12(1): 35-41.
- [32] FOSTER Z S, SHARPTON T J, GRUNWALD N J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data [J]. *Plos Computational Biology*, 2017, 13(2): e1005404.
- [33] BENNETT D J, SUTTON M D, TURVEY S T. Treeman: An R package for efficient and intuitive manipulation of phylogenetic trees[J]. *Bmc Research Notes*, 2017, 10(1): 30-39.
- [34] MCKENZIE A T, KATSYV I, SONG W M, et al. DGCA: A comprehensive R package for differential gene correlation analysis[J]. *Bmc Systems Biology*, 2016, 10(1): 106-130.
- [35] KNAUS B J, GRUNWALD N J. Vcfr: A package to manipulate and visualize variant call format data in R[J]. *Molecular Ecology Resources*, 2017, 17(1): 44-53.
- [36] TITECA K, MEYSMAN P, LAUKENS K, et al. Sfinx: An R package for the elimination of false positives from affinity purification - mass spectrometry datasets[J]. *Bioinformatics*, 2017, 33(12): 1902-1904.
- [37] MEI H, LI L, JIANG F, et al. snpGeneSets: An R package for genome-wide study annotation[J]. *G3 (Bethesda, Md.)*, 2016, 6(12): 4087-4095.
- [38] NOTA B. Gogadget: An R package for interpretation and visualization of GO enrichment results[J]. *Molecular Informatics*, 2017, 36: 5-6.
- [39] 中国测绘网. 地图制图概念(词汇解释) [EB/OL]. (2013-08-28) [2017-05-15]. <http://www.cehui8.com/zhuantu/map/20130828/1191.html>.
- [40] 房世波, 齐月, 韩国军, 等. 1961-2010 年中国主要麦区冬春气象干旱趋势及其可能影响[J]. *中国农业科学*, 2014, 47(9): 1754-1763.
- [41] 国家统计局. 国家数据[EB/OL]. [2017-05-15]. <http://data.stats.gov.cn/easyquery.htm?cn=E0103>.
- [42] 丁香园. 热图的解读与应用[EB/OL]. (2016-02-01) [2017-05-15]. <http://www.dxy.cn/bbs/topic/32836042?source=rss>.
- [43] CARUSO S, MIRAGLIA R, MARUZZELLI L, et al. Chaperone activity of tobacco HSP18, a small heat-shock protein, is inhibited by ATP[J]. *Plant Journal*, 2000, 23(6): 703-713.
- [44] 张厚品. 韦恩图的起源[J]. *数学教学*, 2005(7): 48-49.
- [45] MICALLEF L, RODGERS P. Euler APE: Drawing area-proportional 3-venn diagrams using ellipses[J]. *Plos One*, 2014, 9(7): e101717.

制定用水效率目标时, 除用水效率现状外, 还需综合考虑当地水资源禀赋条件、水环境质量、生态用水需求、用水技术水平等其他条件, 目标约束强度分析的客观可比性仍存在不足. ③分区讨论仅以工、农业用水效率进行目标约束强度为基础, 指标选取不够全面, 且以全国整体约束水平作为标准无法反映区域差异. 各地用水现状和约束情况属复合型, 因此除工、农业用水效率外, 分区还应综合考虑当地水资源禀赋、用水总量、水环境质量等多方面的评价指标, 全面分析当地用水现状、目标及目标约束情况, 以完善分区结果.

[参 考 文 献]

- [1] 操信春. 中国粮食生产用水效率及其时空差异研究 [D]. 陕西 杨凌: 西北农林科技大学, 2015.
- [2] 李静, 孙有珍. 资源与环境双重约束下的粮食生产用水效率研究 [J]. 水资源保护, 2015, 31(6): 67-75.
- [3] 刘涛. 我国农业用水效率的时空差异 [J]. 节水灌溉, 2016(3): 75-79.
- [4] 潘经韬. 中国农业用水效率区域差异及影响因素研究 [J]. 湖北农业科学, 2016(11): 2943-2947.
- [5] 姜蓓蕾, 耿雷华, 卞锦宇, 等. 中国工业用水效率水平驱动因素分析及区划研究 [J]. 资源科学, 2014, 36(11): 2231-2239.
- [6] 李静, 马潇琛. 资源与环境双重约束下的工业用水效率——基于 SBM-Undesirable 和 Meta-frontier 模型的实证研究 [J]. 自然资源学报, 2014, 29(6): 920-933.
- [7] 雷玉桃, 黄丽萍, 张恒. 中国工业用水效率的动态演进及驱动因素研究 [J]. 长江流域资源与环境, 2017, 26(2): 159-170.
- [8] 魏文龙, 曾思育, 杜鹏飞, 等. 一种兼顾目标总量和容量总量的水污染物排放限值确定方法 [J]. 中国环境科学, 2014, 34(1): 136-142.
- [9] 刘年磊, 蒋洪强, 卢亚灵, 等. 水污染物总量控制目标分配研究——考虑主体功能区环境约束 [J]. 中国人口·资源与环境, 2014, 24(5): 80-87.
- [10] 夏军, 王渺林, 王中根, 等. 针对水功能区划水质目标的可用水资源量联合评估方法 [J]. 自然资源学报, 2005, 20(5): 752-760.
- [11] 刘巧玲, 王奇. 基于区域差异的污染物削减总量分配研究——以 COD 削减总量的省际分配为例 [J]. 长江流域资源与环境, 2012, 21(4): 512-517.
- [12] 秦迪岚, 韦安磊, 卢少勇, 等. 基于环境基尼系数的洞庭湖区水污染总量分配 [J]. 环境科学研究, 2013, 26(1): 8-15.
- [13] 王壬, 陈莹, 陈兴伟. 区域水资源可持续利用评价指标体系构建 [J]. 自然资源学报, 2014, 29(8): 1441-1452.

(责任编辑: 张 晶)

(上接第 135 页)

- [46] KOSTAL V, KORBELOVA J, STETINA T, et al. Physiological basis for low-temperature survival and storage of quiescent larvae of the fruit fly *Drosophila melanogaster*[J]. Scientific Reports, 2016(6): 32346-32356.
- [47] 李红, 张元湖. 应用 DPPH? 法测定苹果提取物的抗氧化能力[J]. 山东农业大学学报(自然科学版), 2005, 36(1): 35-38.
- [48] 孙存善. 自由基生物学导论[M]. 合肥: 中国科学技术大学出版社, 1999.
- [49] JIN S K, HA S R, CHOI J S. Effect of *Caesalpinia sappan* L. extract on physico-chemical properties of emulsion-type pork sausage during cold storage[J]. Meat Science, 2015, 110: 245-252.
- [50] MENTE S, KUHN M. The use of the R language for medicinal chemistry applications[J]. Current Topics in Medicinal Chemistry, 2012, 12(18): 1957-1964.

(责任编辑: 张 晶)