

《数据分析与处理技术》附录2
南京审计大学2015级物流管理专业

R语言部分程序包总结

版本：2017.9.11

自动升级包-installr

- `require("stringr")`
- `library("stringr")`
- `library("installr")`
- `updateR()`

通过github获取package

- devtools 从非官方地址加载R包，由于github是目前最流行的R包获取源，github被默认为优先源

```
install.packages("devtools")
```

```
library("devtools")
```

```
install_github("sinhrks/ggfortify") #从github的sinhrks仓库中加载包  
ggfortify
```

- gi

- thubinstall从github上直接搜索可用的包, 并提供备选仓库源

```
install.packages("githubinstall")
```

```
library("githubinstall")
```

```
githubinstall("ggfortify") #给出推荐的github上的ggfortify下载源
```

```
gh_suggest_username("hadly") # 模糊搜索与hadly相似的R包
```

```
gh_search_packages("lasso") #按关键词lasso搜索github上的相关R包
```

```
gh_update_package_list() #升级来自github上的R包
```

数据清洗-dplyr

- 筛选

`filter()`

`filter(flights_df, month==1, day==1)`

- 选出n行数据

`slice(iris, 1:10)`

- 排列

`arrange(iris, Sepal.Length)`

`arrange(iris, desc(Sepal.Length))`

- 选择列

`select(women, height)` 选择数据集内某列

`select(iris, Sepal.Length:Petal.Length)` 选择连续的几个列

`select(iris, -(Sepal.Length:Petal.Length))` 删除这几列

- 生成新列

`mutate(iris, gain=Sepal.Length-Petal.Length)`

- 分组

```
group_by(iris,Species)
```

线性规划包-Rglpk

- `require("slam")`
- `library("Rglpk")`
- `obj<-c(2,4,3)`
- `mat<- matrix(c(3,2,1,4,1,3,2,2,2),nrow=3)`
- `dire<-c("<=", "<=", "<=")`
- `rhs<-c(60,40,80)`
- `Rglpk_solve_LP(obj,mat,dire,rhs,max=TRUE)`
- 其中参数含义请对照右侧线性规划公式，`max=TURE`时求最大化目标，`max=False`则为最小化目标函数

$$\begin{aligned} \max : z &= 2x_1 + 4x_2 + 3x_3 \\ \begin{cases} 3x_1 + 4x_2 + 2x_3 \leq 60 \\ 2x_2 + x_2 + 2x_3 \leq 40 \\ x_1 + 3x_2 + 2x_3 \leq 80 \\ x_1, x_2, x_3 \geq 0 \end{cases} \end{aligned}$$

图像数据包-jpeg

- `install.packages("jpeg")`
- `library("jpeg")`
- `img <- readJPEG("F:\\NauCloud\\1.jpg,`

云型文字图

- 云词的包有wordcloud 和wordcloud2
- 以wordcloud2为例

```
install.packages("wordcloud2")
```

```
wordcloud2(demoFreq,size=1,shape='pentagon')
```

数据demoFreq包含三个列，即默认的row.names, V2:词 V1：词频

```
letterCloud(demoFreq,"R")
```

生成字母形状的云图