

第九届中国R语言会议（武汉）

基于R语言的桑基图实现



华南统计科学研究中心 R square

报告人：李媛媛

目录

- ▷ 第一部分 『桑基图概述』
- ▷ 第二部分 『R语言作图』

第一部分

桑基图概述





定义

桑基图是一种特定类型的流程图，图中延伸的分支的宽度对应数据流量的大小

特征

所有主支宽度的总和应与所有分出去的分支宽度的总和相等，保持能量的平衡

优点

桑基图有利于展现分类维度间的相关性，以流的形式呈现共享同一类别的元素数量，特别适合表达集群的发展。

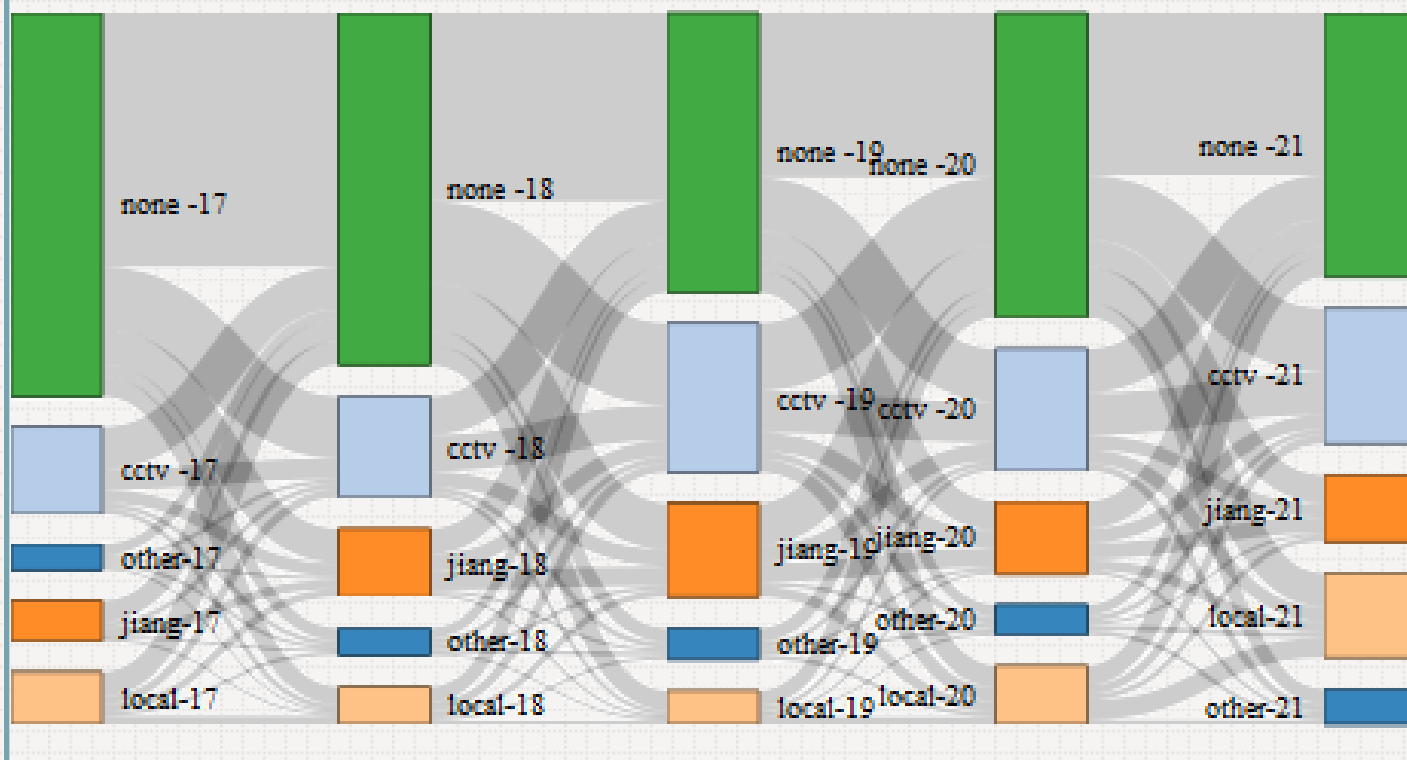
作用

- ▷ 流入流出的可视化分析
- ▷ 分类数据的分布展示

1

桑基图概述

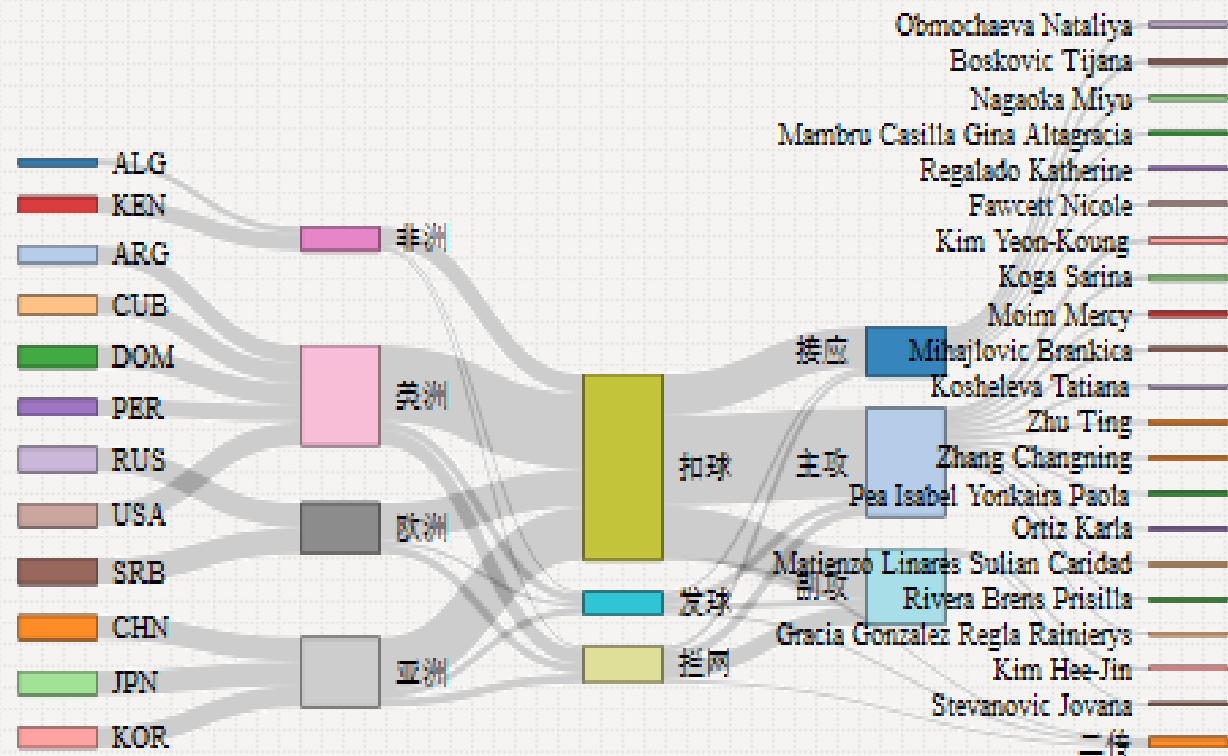
江苏地区某天17:00~21:59频道类型流入流出图
(networkD3)



作用

- ▷ 流入流出的可视化分析
- ▷ 分类数据的分布展示

2015年女排世界杯得分分布图 (networkD3)



作用

▷ 流入流出的可视化分析

▷ 分类数据的分布展示

第二部分

R语言作图





riverplot



networkD3



对比

2 R语言作图

```
riverplot(x, lty = 0, srt = NULL, default_style = NULL, gravity  
= "top", node_margin = 0.1, nodewidth = 1.5, plot_area = 0.5,  
nsteps = 50, add_mid_points = NULL, yscale = "auto")
```

参数	解释	参数	解释
lty	直线类型（0-空，1-实线（默认），2-虚线，3-点线.....）	nodewidth	节点宽度
srt	标签旋转角度	plot_area	画布面积
default.style	默认风格	nsteps	画图的插值步骤数
gravity	节点放置方式（设置y坐标后无效）	add_mid_points	添加节点使图像更为平滑（有y，FALSE）
node_margin	节点间的垂直距离	yscale	按比例决定边缘宽度

目标 x 的三要素

参数	解释
nodestyle	节点类型 (regular、point、invisible)
edgestyle	边缘类型 (sin、straight)
edgecol	边缘颜色 (gradient默认、col同下)
col	节点或边缘的颜色，默认灰色

ID	x
其他17时	1
央视17时	1
江苏17时	1
地方17时	1
未看17时	1
其他18时	2
央视18时	2
江苏18时	2
地方18时	2
未看18时	2

N2	Value
其他18时	710
央视18时	801
江苏18时	614
地方18时	515
未看18时	1946
其他18时	910
央视18时	3792
江苏18时	2294
地方18时	1020
未看18时	7475

styles

nodes

edges

ID、X坐标、(y坐标)

目标 x 的三要素

ID	x
其他17时	1
央视17时	1
江苏17时	1
地方17时	1
未看17时	1
其他18时	2
央视18时	2
江苏18时	2
地方18时	2
未看18时	2

N1	N2	Value
其他17时	其他18时	710
其他17时	央视18时	801
其他17时	江苏18时	614
其他17时	地方18时	515
其他17时	未看18时	1946
央视17时	其他18时	910
央视17时	央视18时	3792
央视17时	江苏18时	2294
央视17时	地方18时	1020
央视17时	未看18时	7475

参数	解释
点类型 (regular、 point、 invisible)	标签旋转，默认90度
线型 (sin、 straight)	线型
字体颜色 (gradient默 认同下)	字体颜色
点或边缘的 颜色，默认灰	

nodes

edges

styles

N1 (来源)、N2 (目标)、Value

目标 x 的三

N1
其他17时
其他17时
其他17时
其他17时
其他17时
央视17时
央视17时
央视17时
央视17时
央视17时

参数	解释	参数	解释
nodestyle	节点类型 (regular、 point、 invisible)	srt	标签旋转， 默认90度
edgestyle	边缘类型 (sin、 straight)	lty	线型
edgecol	边缘颜色 (gradient默 认、col同下)	textcol	字体颜色
col	节点或边缘的 颜色，默认灰 色		

edges

styles

no

地方18
时
未看18
时

x
1
1
1
1
1
2
2
2
2
2



注意

- ▷ ID、x、y、N1、N2、Value
这些变量名不可更改
- ▷ styles要对每个节点都保存
- ▷ 目标类别要规范为“riverplot”

R code

```
library(riverplot)
library(RColorBrewer)
color = brewer.pal(8, "YlGnBu")[2:7]
styles = lapply (node_wat$x, function(n) {
  list(col = color[n], edgecol="col",
    srt="30") })
names(styles) = node_wat$ID
p <- list(nodes = node_wat, edges =
  edge_wat, styles = styles)
class(p) <- c(class(p), "riverplot" )
riverplot(p, plot_area = 0.75,
  yscale=0.1,nodewidth=3,add_mid_points=
```




riverplot



networkD3

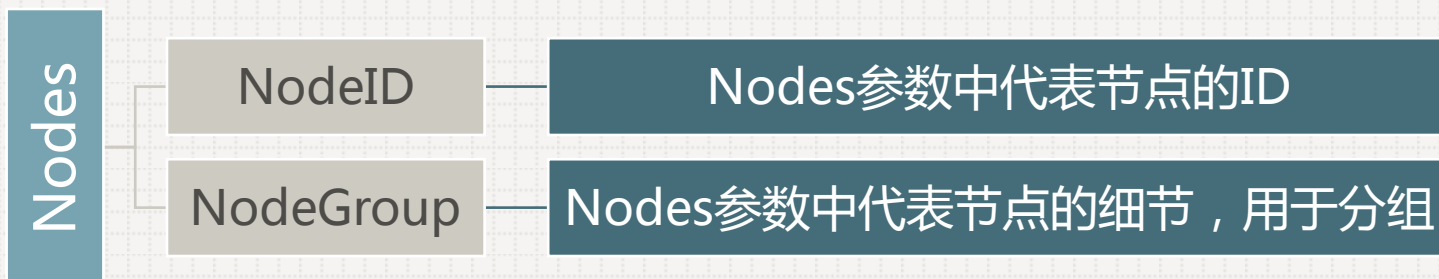


对比

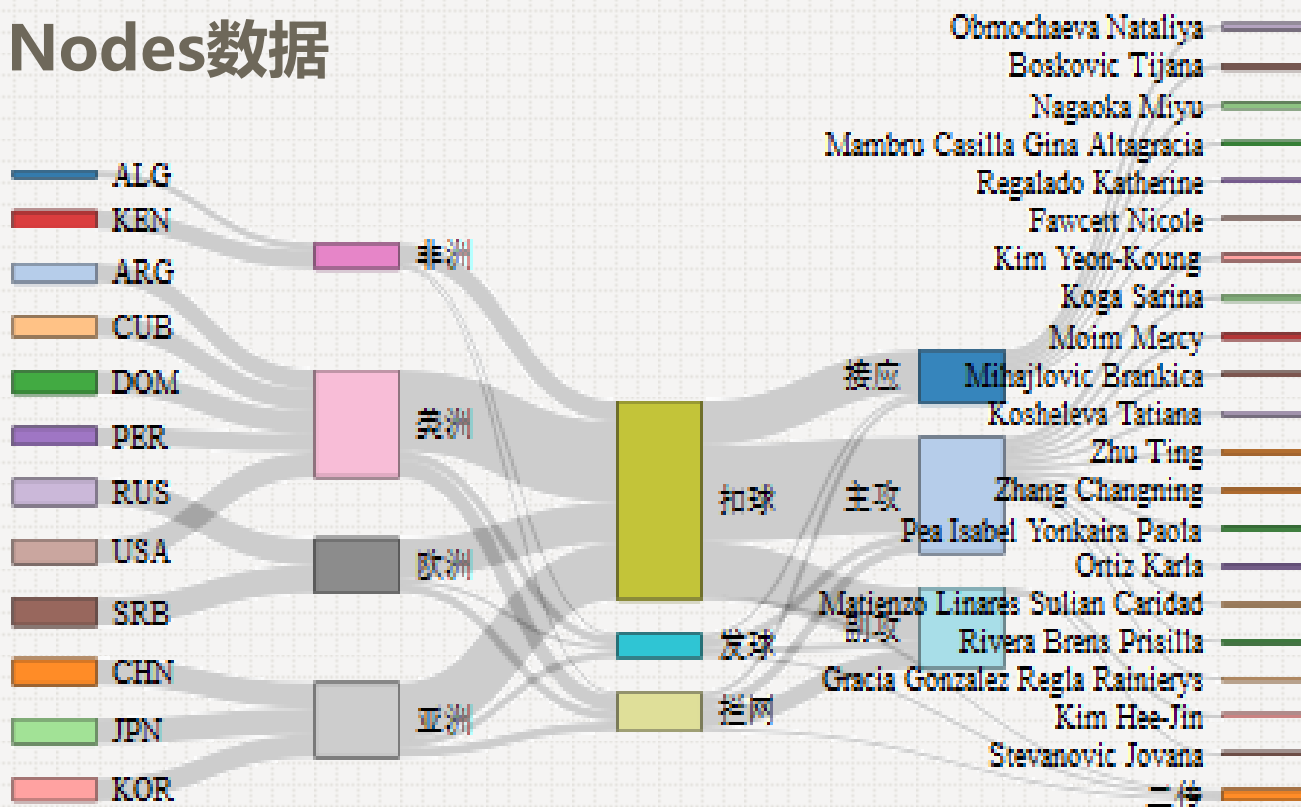


2 R语言作图

```
sankeyNetwork(Links, Nodes, Source, Target, Value, NodeID,  
NodeGroup = NodeID, LinkGroup = NULL, units = "", colourScale =  
JS("d3.scale.category20()"), fontSize = 7, fontFamily = NULL,  
nodeWidth = 15, nodePadding = 10, margin = NULL, height = NULL,  
width = NULL, iterations = 32)
```



Nodes数据



NodeID	num	NodeGroup
ALG	0	1
ARG	1	2
CHN	2	3
CUB	3	4
DOM	4	5
...
Zhu Ting	31	3
Kim Hee-Jin	32	8
Zhang Changning	33	3
Pea Isabel Yonkaira Paola	34	5

* NodeGroup 要用因

```
sankeyNetwork(Links, Nodes, Source, Target, Value, NodeID,  
NodeGroup = NodeID, LinkGroup = NULL, units = "", colourScale =  
JS("d3.scale.category20()"), fontSize = 7, fontFamily = NULL,  
nodeWidth = 15, nodePadding = 10, margin = NULL, height = NULL,  
width = NULL, iterations = 32)
```

Links

Source

Links参数中代表节点数据来源的变量名

Target

Links参数中代表节点数据去处的变量名

value

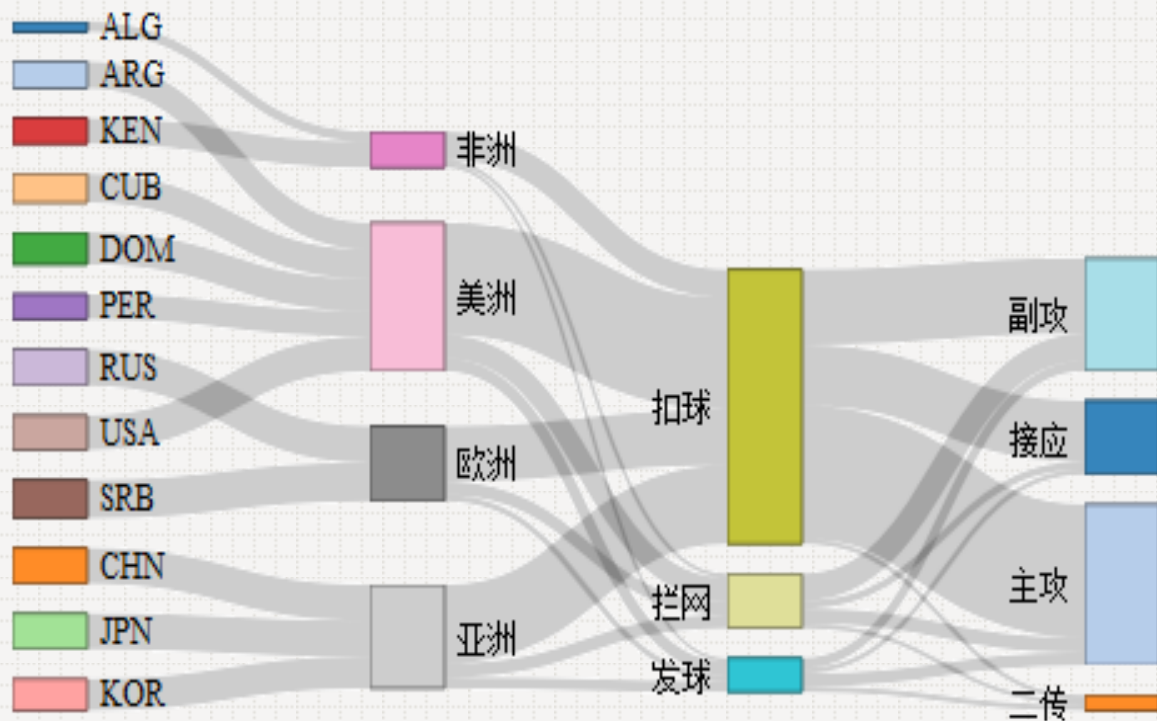
Links参数中节点的数据

LinkGroup

Links参数中代表节点的细节，用于分组

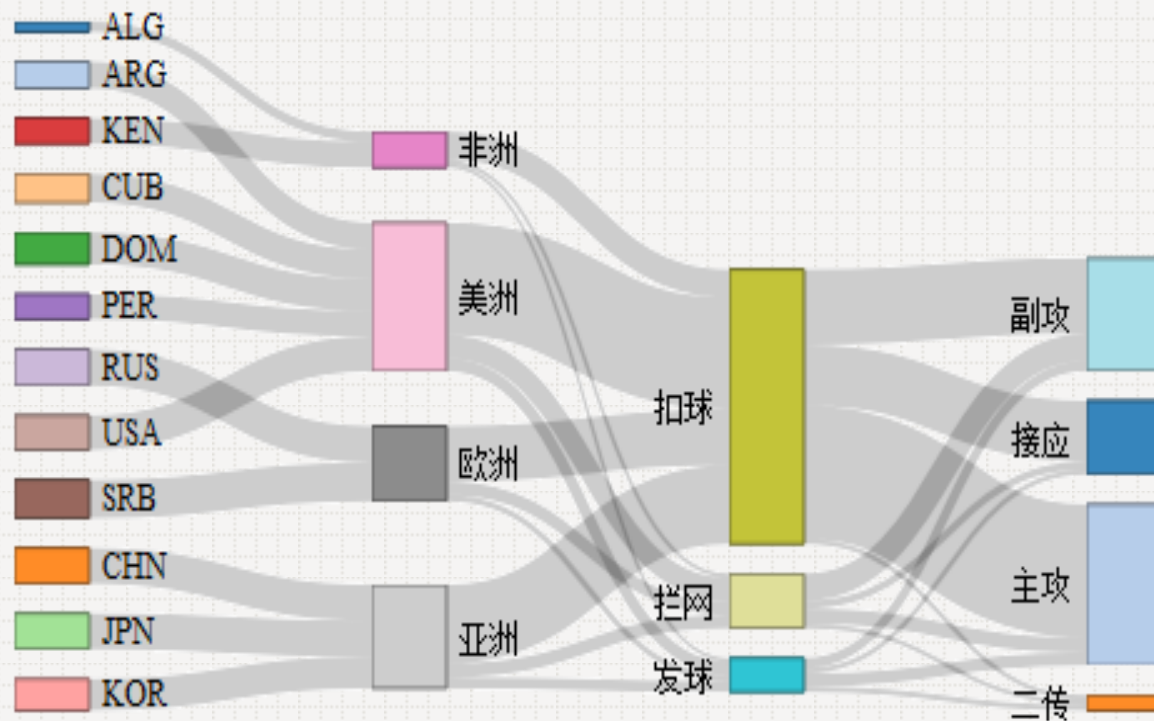
x	y	value
ALG	非洲	200
ARG	美洲	527
CHN	亚洲	702
CUB	美洲	572
DOM	美洲	630
JPN	亚洲	704
KEN	非洲	509
KOR	亚洲	622
PER	美洲	495
RUS	欧洲	719

Links数据



source	target	value
0	12	200
1	13	527
2	15	702
3	13	572
4	13	630
5	15	704
6	12	509
7	15	622
8	13	495
9	14	719

Links数据 (final)



2 R语言作图

▷ 其他参数

参数	解释	参数	解释
units	描述Value的物理单位	nodewidth	节点宽度
colourScale	调色板设置	margin	图像的边缘设置
fontSize	字号设置	height	像素高度
fontFamily	字体设置	width	像素宽度
nodePadding	影响宽高的数字	iterations	迭代次数（不宜过高）



▷ 保存为html格式

```
library(magrittr)
S1=SankeyNetwork(...)
s1 %>%
saveNetwork(file =
'Net1.html')
```

▷ 基于shiny的交互式

```
library(shiny)
output$sankey <- renderSankeyNetwork(
SankeyNetwork(...)) #sever.R文件

sankeyNetworkOutput( "sankey" ) #u
文件
```




riverplot



networkD3



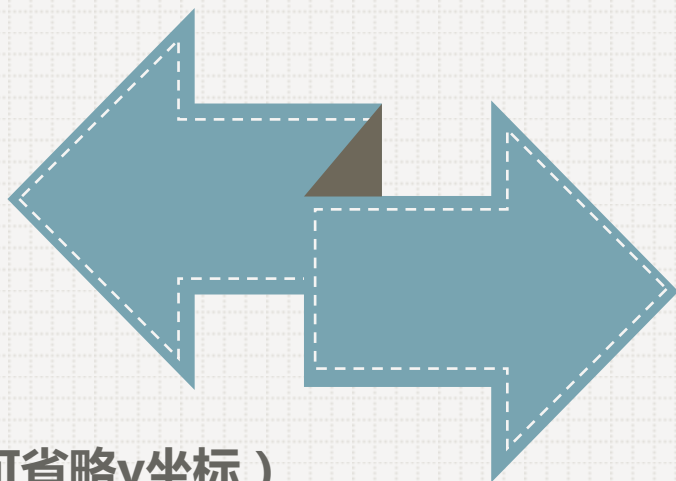
对比





riverplot

- ▷ 位置标号从1开始
- ▷ 节点位置以坐标描述（可省略y坐标）
- ▷ 变量名不可随意更改（N1、N2、Value、ID、x）
- ▷ 画图目标为列表型
- ▷ 仅有图片格式



networkD3

- ▷ 位置标号必须从0开始
- ▷ 所有节点位置为同一序列
- ▷ 自定义变量名
- ▷ 作图函数作用于数据框
- ▷ 可存为html格式



第九届中国R语言会议（武汉）

Thank you for watching !



华南统计科学研究中心 R square



报告人：李媛媛

liyy_circle@163.com