



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS



Data Analytics with R  
语言数据分析



# 数据对象

艾新波 / 2018 • 北京



# 课程体系



## R语言数据分析



### 上部：论道



- 第1章 气象万千、数以等观
- 第2章 所谓学习、归类而已
- 第3章 格言联璧话学习
- 第4章 源于数学、归于工程



### 中部：执具



- 第5章 工欲善其事必先利其器
- 第6章 基础编程
- 第7章 数据对象



- 第8章 人人都爱tidyverse
- 第9章 最美不过数据框



### 下部：博术



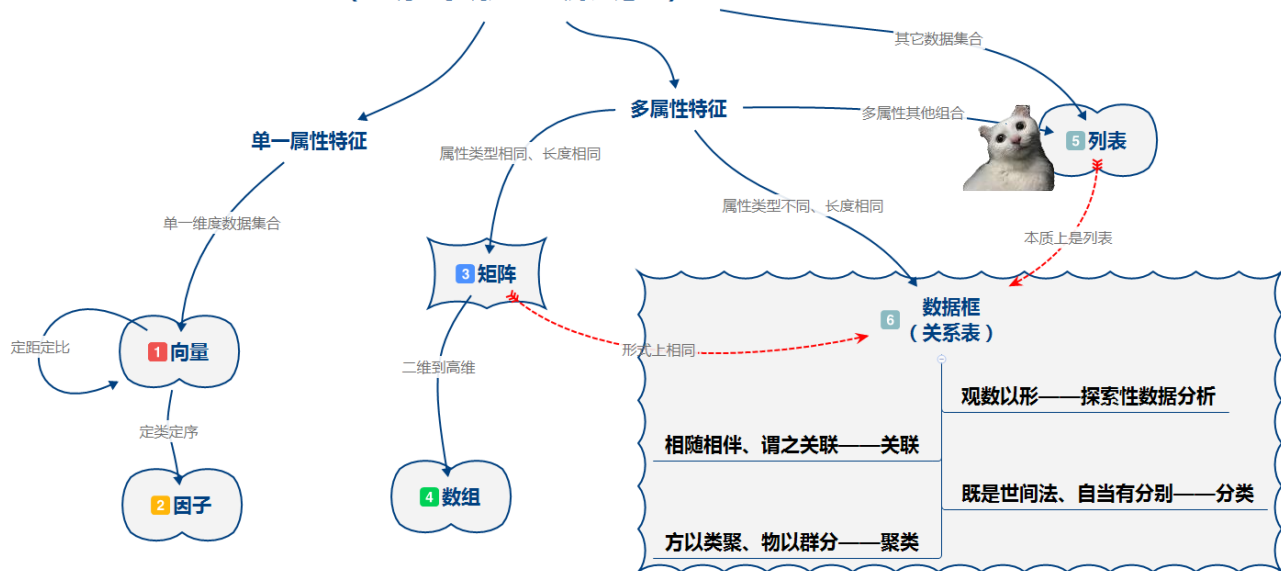
- 第10章 观数以形
- 第11章 相随相伴、谓之关联
- 第12章 既是世间法、自当有分别
- 第13章 方以类聚、物以群分
- 第14章 庐山烟雨浙江潮

# 数据对象



当前位置

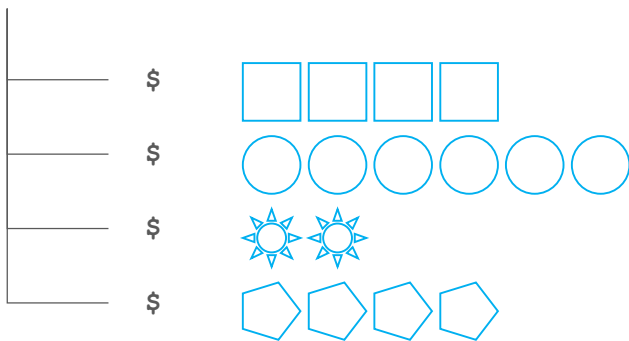
设备或人工采集到的数据  
(对现实系统进行观测、记录)



# 列表

- 列表是对象的**有序集合**，包含的对象又称为它的分量
- 列表最为灵活、最具有包容性
- 对所包含的对象没有限制，可以是不同的类型、不同的长度

list



## 创建列表

#北京邮电大学下设以下学院

```
xue_yuan <- c("信息与通信工程学院", "电子工程学院",  
  "计算机学院", "自动化学院", "软件学院", "数字媒体与设计艺术学院",  
  "现代邮政学院", "网络空间安全学院", "光电信息学院", "理学院",  
  "经济管理学院", "人文学院", "马克思主义学院", "国际学院",  
  "网络教育学院", "继续教育学院", "民族教育学院")
```

#拥有以下基地

```
ji_di <- c(国家重点实验室 = 2, 国家工程实验室 = 5, 部级重点实验室 = 9)  
xiao_qu <- c("西土城路校区", "沙河校区", "宏福校区") #校区分布  
xue_sheng <- c(全日制 = 30000, 非全日制 = 45000) #学生数量
```

## 创建列表

#集合在一起，变成一个整体

```
bupt <- list(xue_yuan = xue_yuan,  
            xiao_qu = xiao_qu,  
            ji_di = ji_di,  
            xue_sheng = xue_sheng)
```

```
length(bupt)
```

```
#> [1] 4
```

```
names(bupt)
```

```
#> [1] "xue_yuan" "xiao_qu"  "ji_di"    "xue_sheng"
```

```
typeof(bupt)
```

```
#> [1] "list"
```

## 列表的子集

#通过美元\$符号访问

```
bupt$xue_sheng
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

```
bupt$xue_sheng["全日制"]
```

```
#> 全日制
```

```
#> 30000
```

```
sum(bupt$xue_sheng)
```

```
#> [1] 75000
```

## 列表的子集

#通过[]来访问

```
bupt[4]
```

```
#> $`xue_sheng`
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

```
typeof(bupt[4]) #单个的[], 看到的依然是包装箱
```

```
#> [1] "list"
```

```
bupt[[4]]
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

```
typeof(bupt[[4]])
```

```
#> [1] "double"
```

一般而言  
X[]与X类型相同



## 列表的子集

```
sum(bupt[4])
```

```
#> Error in sum(bupt[4]) : invalid 'type' (list) of argument
```

```
sum(bupt[[4]]) #正确的打开方式
```

```
#> [1] 75000
```

```
bupt["xue_sheng"]
```

```
#> $`xue_sheng`
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

```
bupt[["xue_sheng"]]
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

## 列表的子集

#以下三种方式效果相同

```
bupt$xue_sheng
```

```
bupt[[4]]
```

```
bupt[["xue_sheng"]]
```

```
#> 全日制 非全日制
```

```
#> 30000      45000
```

## 更改列表元素

```
bupt$A_xueke <- c("信息与通信工程", "计算机科学与技术", "电子科学与技术")
```

```
length(bupt)
```

```
#> [1] 5
```

```
names(bupt)
```

```
#> [1] "xue_yuan" "xiao_qu" "ji_di" "xue_sheng"
```

```
#> [5] "A_xueke"
```

```
bupt$A_xueke <- NULL #注意NULL/NA/""之间的区别
```

```
names(bupt)
```

```
#>[1] "xue_yuan" "xiao_qu" "ji_di" "xue_sheng"
```

## 列表的操作

#对列表的每一个组成部分，执行某种操作

```
(component_length <- lapply(bupt, length))
```

```
#> $`xue_yuan`
```

```
#> [1] 17
```

```
#> $xiao_qu
```

```
#> [1] 3
```

```
#> $ji_di
```

```
#> [1] 3
```

```
#> $xue_sheng
```

```
#> [1] 2
```

## 列表的操作

```
unlist(component_length)
```

```
#> xue_yuan      xiao_qu      ji_di xue_sheng
```

```
#> 17           3           3           2
```

```
sapply(bupt, length)
```

```
#> xue_yuan      xiao_qu      ji_di xue_sheng
```

```
#> 17           3           3           2
```

```
sapply(bupt, typeof)
```

```
#> xue_yuan      xiao_qu      ji_di      xue_sheng
```

```
#> "character" "character" "double"    "double"
```

A decorative blue border with rounded corners frames the entire slide. Two thin blue crosshair symbols are positioned on the right and left sides of the text.

**谢谢聆听**  
**Thank you**

# 教师个人联系方式

艾新波

手机: 13641159546

QQ: 23127789

微信: 13641159546

E-mail: [13641159546@126.com](mailto:13641159546@126.com)

[axb@bupt.edu.cn](mailto:axb@bupt.edu.cn)

地址: 北京邮电大学科研楼917室

课程网址: <https://github.com/byaxb/RDataAnalytics>

