

R基础培训

生信帮

云南农业大学

云南省生物大数据重点实验室

2023年8月12日

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图

一、R简介与安装



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

R软件的起源

R是一个开放的统计编程环境，是一门用于统计计算和作图的语言,是S语言的一种实现，诞生于1995年。

- S语言是在1975年至1976年间由AT&T贝尔实验室开发的一种用来进行数据探索、统计分析、作图的解释型语言。
- R是贝尔实验室的Rick Becker, John Chambers和Allan Wilks开发的S语言的一种实现或形式。
- 最先(1995年)是由Auckland大学统计系的Ross Ihaka和Robert Gentleman共同创立，现在由R开发核心小组维护。

一、R简介与安装



雲南農業大學
Yunnan Agricultural University

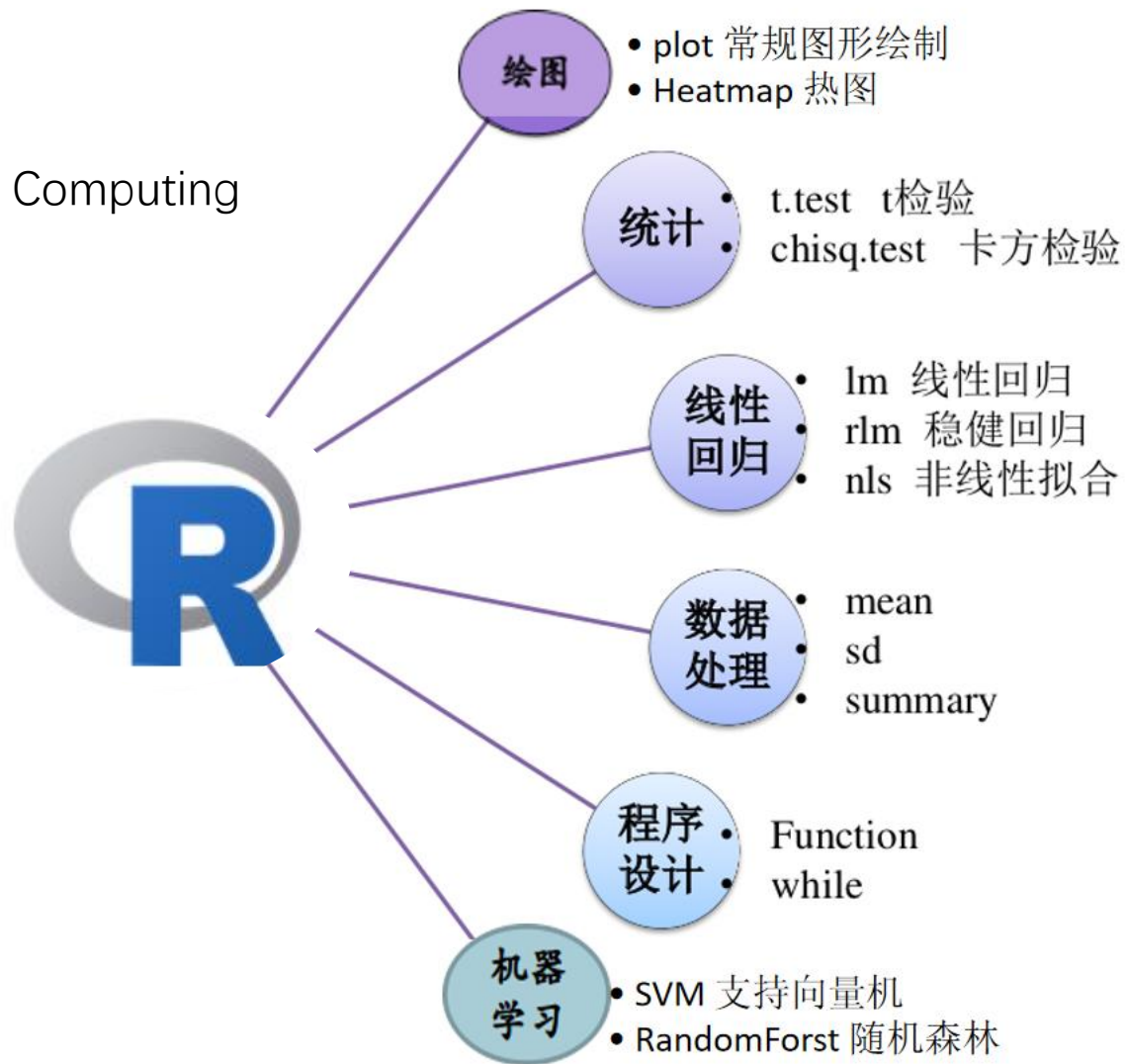
云南省生物大数据重点实验室



生信帮

R语言的功能

The R Project for Statistical Computing



一、R简介与安装



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

R软件下载

下载: <http://www.r-project.org/>

安装 (Windows) : 双击下载的软件即可

更新: 下载新版本软件安装或用以下3个命令

```
install.packages("installr"); library(installr); updateR()
```

R语言的编程环境IDE

R软件Console控制台, 写完一条命令回车即运行

其他文本编辑器

Rstudio, 下载地址:

<https://www.rstudio.com/products/rstudio/download/>

一、R简介与安装

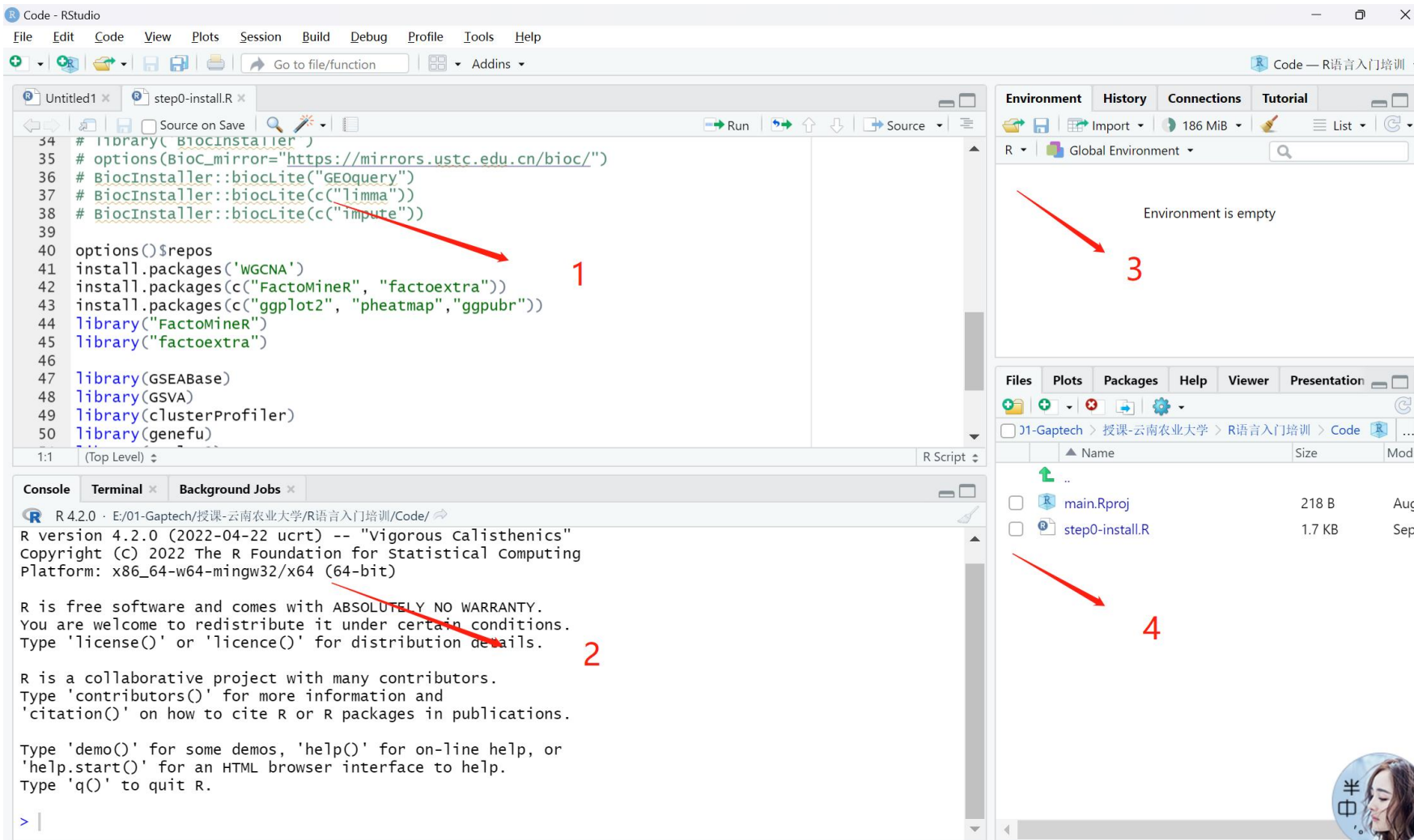


雲南農業大學
Yunnan Agricultural University



生信帮

云南省生物大数据重点实验室



Rstudio界面

- 1: Text Editor 编辑器
- 2: Console 控制台
- 3: Global Environment 显示环境变量的区间
- 4: 展示台, 可以浏览文件目录, 展示绘图结果, 显示已安装和加载的软件包, 显示帮助文档等

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图

二、包安装大全



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

基础包以外的包都是扩展包

①CRAN上的扩展包的安装（在线安装）

```
install.packages("ggplot2") #每次安装一个
```

```
install.packages(c("ggplot2", "vegan")) #一次安装多个
```

②Bioconductor上的扩展包的安装（在线安装）

```
if (!requireNamespace("BiocManager", quietly = TRUE))
```

```
  install.packages("BiocManager")
```

```
BiocManager::install("limma")
```

③Github上的‘野’包

包+Github网页搜索得到安装方式

④文本型的扩展包（或称为函数）的安装（调用）

```
source("http://bioconductor.org/biocLite.R")
```

```
source("http://zzlab.net/GAPIT/gapit_functions.txt")
```

扩展包是R得以蓬勃发展的主要原因

- R中的所有符号都是英文状态下的；
- 对字母的大小写敏感，所以R中用到的任何名称必须写得完全正确
- 一个包可能在多个地方都有，存在多种安装方式
- 安装软件包的同时一般都会安装其依赖包，如果一次安装不成功，可能是由于其依赖包没有安装成功，重复运行安装命令或者先将其依赖包安装好。其所依赖的包在安装时都会有提示，注意看提示即可

二、包安装大全



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

扩展包的调用

`library(ggplot2)`或者`require(ggplot2)`; 区别:

- `library()`直接加载扩展包, 不成功则返回错误提示, 不返回逻辑值;
- `require()`加载包时可返回逻辑值, 加载成功为TRUE, 否则为FALSE。

一般情况下使用`library()`, 在需要判断加载是否成功是用`require()`。

文本型的扩展包 (或称为函数) 调用:

`source(http://zzlab.net/GAPIT/gapit\_functions.txt)`

关闭已加载的扩展包可以用`detach()`

如`detach(package:ggplot2)`

软件包的更新: 重新运行安装命令即可

二、包安装大全



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

获取帮助文档

对于基础包（如“stats” “graphics” “grDevices” “utils” “datasets” “methods” “base” 安装R软件时就已自动安装，每次启动R都会自动加载这些包）中函数，输入问号加函数名称即可获取该函数的帮助文件：

`?aov()` 或 `help(aov)`

对于扩展包，加载后再用上述命令：

`library(ggplot2); help(geom_abline)`或`?geom_abline`

获取**整个包**的帮助文件

`help(ggplot2)`或`?ggplot2`

模糊搜索：两个问号加搜索主题

`??heatmap; ??glm`

二、包安装大全



帮助文档构成与阅读

构成:

- 描述 (Description)
- **用法 (Usage)**
- **各个参数的具体类型与含义 (Arguments)**
- 某些参数的细节 (Details)
- 返回值 (Value)
- **注意事项 (Note)**
- 作者 (Authors)
- 参考文献 (Reference)
- 其他类似函数 (See also)
- **举例 (Examples)** 等

以上红色加粗部分是应当重点阅读的部分

R本地帮助文档结构说明

plot (graphics) **1** → 函数名 包

Generic X-Y Plotting

Description **2** → 函数功能描述信息

Generic function for plotting of R objects. For more details about the graphical parameter arguments, see [par](#).

For simple scatter plots, [plot.default](#) will be used. However, there are plot methods for many R objects, including [functions](#), [density](#) objects, etc. Use methods (plot) and the documentation for these.

Usage **3** → 函数调用方法

plot(x, y, ...)

Arguments **4** → 参数说明

x the coordinates of points in the plot. Alternatively, a single plotting structure, function or any R object with a plot method

y the y coordinates of points in the plot, optional if x is an appropriate structure.

... Arguments to be passed to methods, such as [graphical parameters](#) (see [par](#)). Many methods will accept the following a

type

Details **5** → 补充说明

The two step types differ in their x-y preference: Going from (x1,y1) to (x2,y2) around.

See Also **6** → 其它类似信息

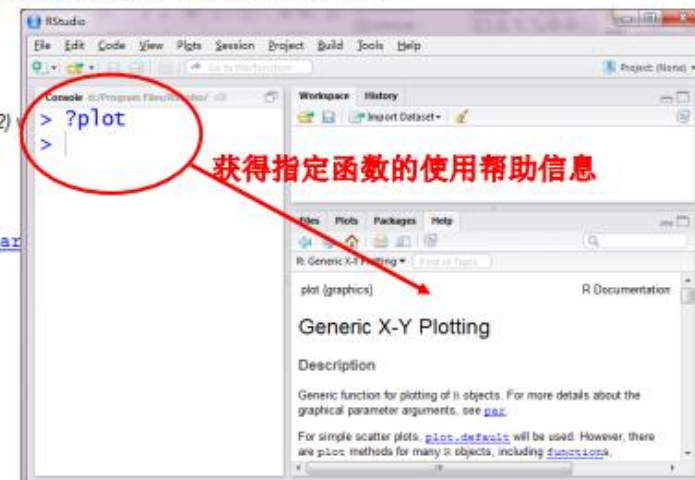
[plot.default](#), [plot.formula](#) and other methods; [points](#), [lines](#), [par](#)

For X-Y-Z plotting see [contour](#), [persp](#) and [image](#).

Examples **7** → 使用例子!

```
require(stats)
plot(cars)
lines(lowess(cars))

plot(sin, -pi, 2*pi) # see ?plot.function
```



二、包安装大全



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

网站推荐：

Quick-R : <http://www.statmethods.net/>

Stackoverflow: <http://stackoverflow.com/>

R-blogger: <http://www.r-bloggers.com/>

使用谷歌、必应搜索、搜狗搜索，放弃百度（你懂的！）

必应搜索: <https://cn.bing.com/>

搜狗搜索: <https://weixin.sogou.com/>

二、包安装大全



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

书籍推荐：

入门书籍：

- 《R语言实战》（R in Action）中英文版
- 《统计建模与R软件》（中国人编的教材）
- 《153分钟学会R》（问题导向型）
- <https://www.runoob.com/r/r-tutorial.html>

高级入门书籍

- 《The R Book》
- 《ggplot2:Elegant Graphics for Data Analysis》（中英文版）
- 《R语言与Bioconductor生物信息学应用》

高级书籍（统计建模的《葵花宝典》级）

- 《An Introduction to Statistical Learningwith Applications in R》
- <http://www-bcf.usc.edu/~gareth/ISL/>
- 《The Elements of Statistical Learning 2nd Edition》
- <http://statweb.stanford.edu/~tibs/ElemStatLearn/>

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

R是一种解释性的语言

- R操作的实体在技术上来说都是**对象(object)**。当R在运行时,所有变量,数据,函数及结果都以对象的形式存在计算机的活动内存中,并有相应的名字对应。
- 对象的属性:
 - 类型, class()
 - 模式, mode()
 - 结构, str()
 - 长度, length()
 - 维度, dim()

```
> x<-matrix(0,4,5)
> x
[1,] [2,] [3,] [4,] [5,]
[1,] 0 0 0 0 0
[2,] 0 0 0 0 0
[3,] 0 0 0 0 0
[4,] 0 0 0 0 0
> class(x)
[1] "matrix"
> mode(x)
[1] "numeric"
> length(x)
[1] 20
>
```


三、数据类型与函数



常用数据结构及其类型

对象	描述	模式	是否允许同一个对象中有多种模式
向量	一系列元素的组合	数值型, 字符型, 复数型, 逻辑型	否
因子	因子是一个分类变量	数值型, 字符型	否
数组	数组是k维的数据表 (k in 1:n, n 为正整数)	数值型, 字符型, 复数型, 逻辑型	否
矩阵	二维的数据表, 是数组的一个特例	数值型, 字符型, 复数型, 逻辑型	否
数据框	是由一个或几个向量和 (或) 因子构成, 它们必须是等长的, 但可以是不同的数据类型。	数值型, 字符型, 复数型, 逻辑型	是
时间序列	由函数ts根据向量 (一元) 或矩阵 (多元) 创建	数值型, 字符型, 复数型, 逻辑型	否
列表	列表可以包含任何类型的对象。	数值型, 字符型, 复数型, 逻辑型, 函数, 表达式, ...	是

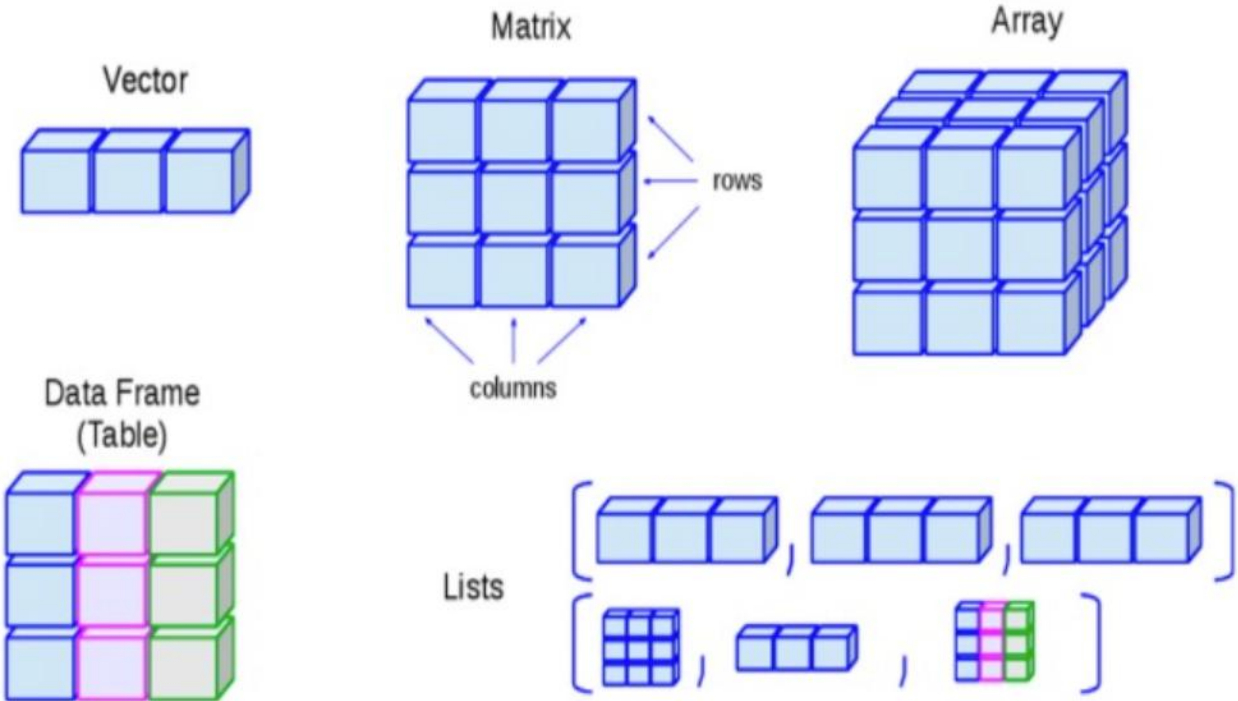
数值型	Numeric	如 100, 0, -4.335	字符型	Character	如 "China"	数据类型
逻辑型	Logical	如TRUE, FALSE	因子型	Factor	表示不同类别	
复数型	Complex	如: 2 + 3i				

三、数据类型与函数



常用数据结构及其类型

对象	描述
向量	一系列元素的组合
因子	因子是一个分类变量
数组	数组是k维的数据表 ($k \in 1:n$,
矩阵	二维的数据表, 是数组的一
数据框	是由一个或几个向量和 (或) 因子
时间序列	是由一个或几个向量和 (或) 因子
列表	列表可以包含任何类型的



数值型	Numeric	如 100, 0, -4.335	字符型	Character	如 "China"
逻辑型	Logical	如TRUE, FALSE	因子型	Factor	表示不同类别
复数型	Complex	如: 2 + 3i			

数据类型

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

向量



向量的生成

- **c()** #若向量没有规律
- **:** #以1为步长的等差序列
- **seq()** #以n为步长的等差序列
- **rep()** #若向量具有较为复杂的规律
- **sequence()** #有规律向量
- **paste()** #转化为字符后链接向量
- **factor()** #因子型向量,也可利用gl()
- **numeric()** #初始化向量为0
- **scan()** #通过键盘逐个输入

向量的下标(index)与向量子集(元素)的提取

- 正的下标提取向量中对应的元素: **x[4:10]**
- 负的下标去掉向量中对应的元素: **x[-1]**
- 逻辑运算提出向量中元素的值满足条件的元素: **x[x>0]**

向量统计

- **sum()** #求和
- **mean()** #求平均值
- **var()** #方差
- **sd()** #标准差
- **min()** #最小值
- **max()** #最大值
- **range()** #取值范围 (二维向量, 最大值和最小值)

注:R中向量的下标从1开始, 不同于某些高级编程语言的向量下标则从0开始!

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

因子(factor)



- **定义：**因子是一个对等长的其它向量元素进行分类(分组)的**向量对象**，这种变量通常被称为**类型变量**。有有序和无序因子之分。
- **作用：**分类,最重要的用途之一是应用于统计建模。
- **属性：**水平，因子可能的取值集合 levels()
- R中的因子存储为整数值的向量，有一组相应的字符值与其相对应，当显示因子时要用到这些字符值。

创建因子：函数factor()

```
factor(x = character(), levels, labels = levels, exclude = NA,  
ordered = is.ordered(x))
```

参数说明：

- x：向量。
- levels：指定各水平值，不指定时由x的不同值来求得。
- labels：水平的标签，不指定时用各水平值的对应字符串。
- exclude：排除的字符。
- ordered：逻辑值，用于指定水平是否有序。
- nmax：水平的上限数量。

注：R中向量的下标从1开始，不同于某些高级编程语言的向量下标则从0开始！

三、数据类型与函数



数组：数组是指维数 ≥ 1 的数据表；而**矩阵**是数组的一个特例(维数=2)

数组的属性: 模式(mode)、长度(length)、维(dim)

数组的建立:

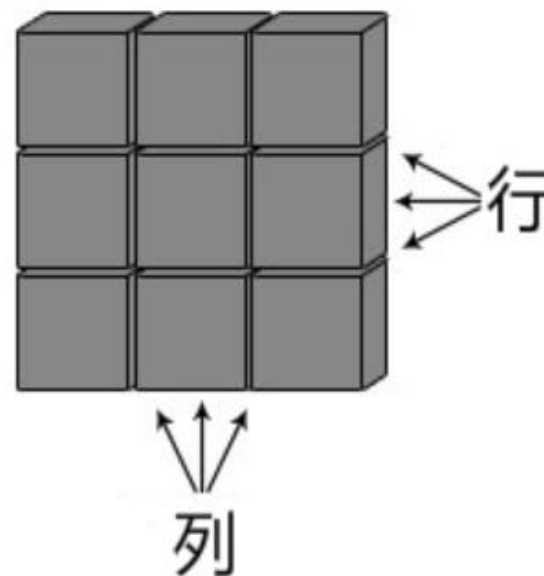
```
array(data = NA, dim = length(data), dimnames = NULL)
```

矩阵的建立:

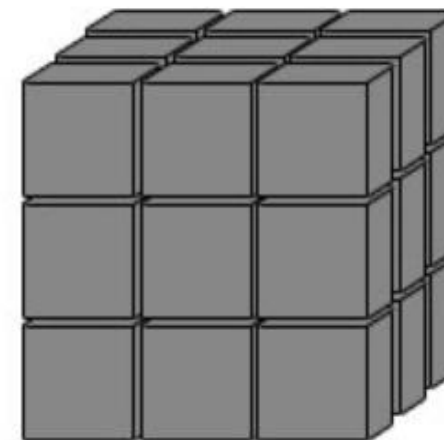
```
matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)
```

约束条件:数组中所有元素必须是同一种模式的。

矩阵



数组



三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

数组：数组是指维数 ≥ 1 的数据表；而**矩阵**是数组的一个特例(维数=2)

数组的属性: 模式(mode)、长度(length)、维(dim)

数组的建立:

```
array(data = NA, dim = length(data), dimnames = NULL)
```

矩阵的建立:

```
matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)
```

约束条件:数组中所有元素必须是同一种模式的。

其他函数:

- `nrow(x)` #行数
- `ncol(x)` #列数
- `rownames(x)` #行名
- `colnames(x)` #列名
- `rowsum(x,group)` #行加和
- `colsum(x,group)` #列加和
- `rowMean()` #行均值
- `colMean()` #列均值
- `rbind()` #行合并
- `cbind()` #列合并

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

数据框(data frame)

数据框 (Data frame) 可以理解成我们常说的"表格"

- 是特殊的二维列表
- 每一列的数据必须是同一模式,且长度要相等
- 数据框架子集的提取:与矩阵基本相同,不同的是对于列还可以使用变量的名称。# df\$colname

data frame

1	"S"	TRUE
7	"A"	FALSE
3	"U"	TRUE
numeric	character	logical

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

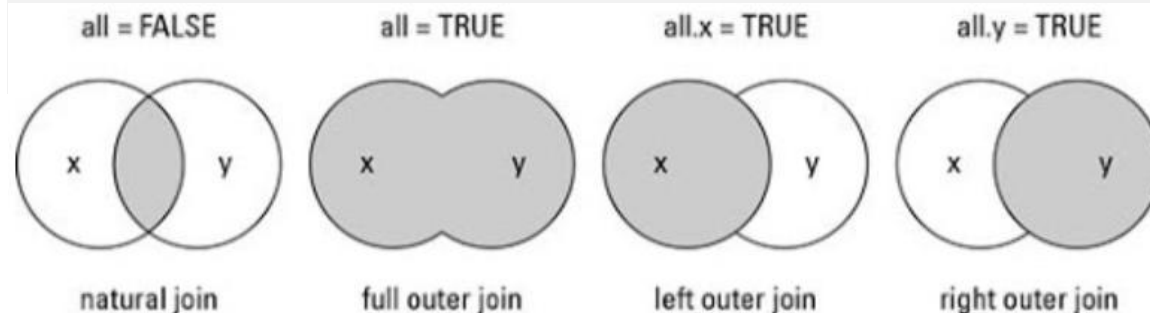
数据框(data frame)

数据框 (Data frame) 可以理解成我们常说的"表格"

- 是特殊的二维列表
- 每一列的数据必须是同一模式,且长度要相等
- 数据框架子集的提取:与矩阵基本相同,不同的是对于列还可以使用变量的名称。# df\$colname

建立数据框架的方法(函数):

- 通过函数从变量, `data.frame()`
- 从文件:
 - `read.table()` 表格式文件默认空格分割
 - `read.csv()` 逗号分隔的文件
 - `read.delim()` tab键分隔的文件
- `data.frame()` #将指定列合并数据框
- `merge()` #根据共有的行或列名合并数据框



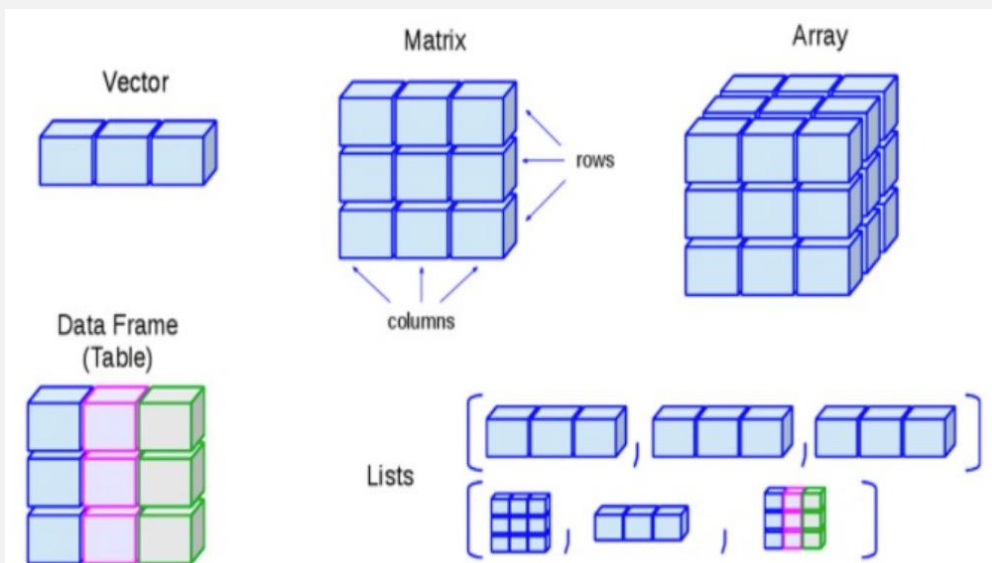
三、数据类型与函数



列表(list)

列表是 R 语言的对象集合，可以用来保存不同类型的数据，可以是数字、字符串、向量、另一个列表、矩阵、数据框等，当然还可以包含矩阵和函数。

列表是一种灵活的数据结构，可以存储和操作多种类型的数据对象



- **创建**: 使用 `list()` 函数
- **提取**: 列表的下标与子集的提取与数据框很相似，提取一个子对象如foo的x,下面三种方式:

```
foo = list(x = 1:6, y = matrix(1:4, nrow = 2))
```

- `foo$x` `#[1] 1 2 3 4 5 6`
- `foo[1]` `#[1] 1 2 3 4 5 6`
- `foo[[1]]` `#[1] 1 2 3 4 5 6`

三、数据类型与函数



变量命名规则

R 语言的有效变量名称由字母，数字以及点号 . 或下划线 _ 组成。变量名称以字母或点开头。

变量名	是否正确	原因
var_name2.	正确	字符开头，并由字母、数字、下划线和点号组成
var_name%	错误	% 是非法字符
2var_name	错误	不能数字开头
.var_name, var.name	正确	可以 . 号开头，但是要注意 . 号开头后面不能跟着数字
.2var_name	错误	. 号开头后面不能跟着数字
_var_name	错误	不能以下划线 _ 开头

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

R语法：条件判断与循环

条件判断

形式1:

```
if (条件) {  
    表达式1  
}
```

形式2:

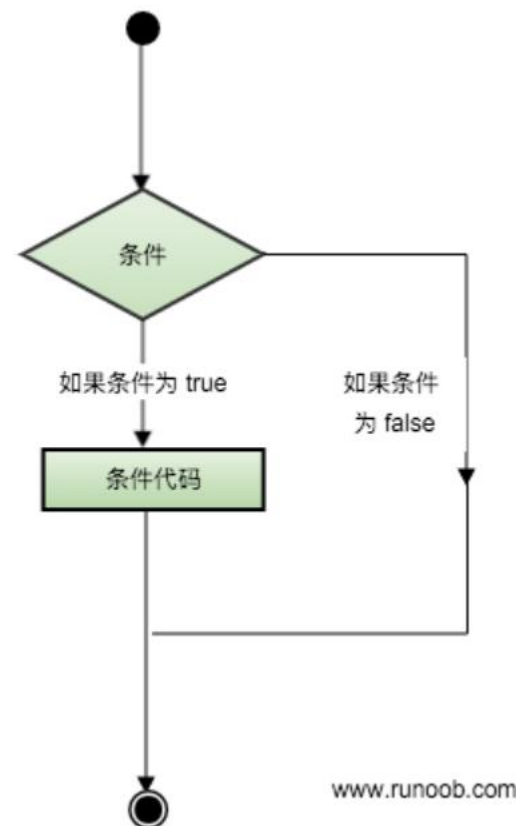
```
if (条件) {  
    表达式1  
}else {  
    表达式2  
}
```

形式3:

```
if(条件1){  
    表达式1  
}else if(条件2){  
    表达式2  
}else{  
    表达式3  
}
```

形式4:

```
ifelse(条件, yes, no)
```



三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

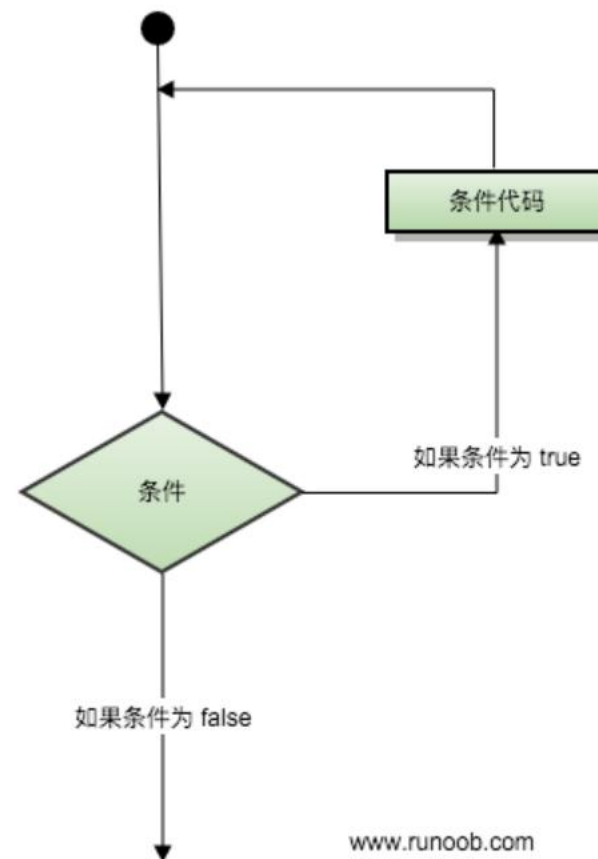


生信帮

R语法：条件判断与循环

循环

- ✓ **for**: 若知道终止条件
for (变量 in 向量) 表达式
- ✓ **while**: 若无法知道运行次数
while(条件) 表达式
- ✓ **repeat**: 若无法知道运行次数
repeat 表达式 #利用break退出循环



www.runoob.com

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

文件操作相关函数

创建和删除

file.create(...)
file.remove(...)
file.append(file1, file2)

文件测试

file.access(names, mode = 0)
file_test(op, x, y)
file.exists(...)

改变和获得路径

setwd()
getwd()

文件信息

file.info(...)

文件显示和编辑

file.show()

file.edit()

创建文件夹

dir.create(path, ...)
list.dirs()
dir()
list.files()

权限更改

Sys.chmod

其它操作

file.rename(from, to)
file.copy(from, to, ...)
file.symlink(from, to)
file.link(from, to)
file.path()
file.choose()

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

程序运行

在R内部: `source("fun.R")`

在命令行: `Rscript fun.R`

三、数据类型与函数



雲南農業大學
Yunnan Agricultural University



生信帮

云南省生物大数据重点实验室

R语法：注释

注释主要用于一段代码的解析，可以让阅读者更易理解，编程语言的注释会被编译器忽略掉，且不会影响代码的执行

阅读者：

- 代码维护更新者
- 你自己
- 未来的你



三、数据类型与函数



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

R语法：注释

一般编程语言的注释分为单行注释与多行注释，但是 R 语言只支持单行注释，注释符号为 **#**，快捷键：**Ctrl+Shift+C**

单行注释

```
# 这是我的第一个编程代码  
myString <- "Hello, World!"  
  
print ( myString )
```

多行注释

```
# R 语言实现两个数相加  
  
# 变量赋值  
a <- 9  
b <- 4  
  
# 输出结果  
print(a + b)
```

多行注释还有一个变通的写法

```
if(FALSE) {  
  "  
  这是一个多行注释的实例  
  注释内容放在单引号或双引号之间  
  "  
}  
  
myString <- "Hello, World!"  
print ( myString)
```

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图

四、文件读写



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

读取文件

读取向量和矩阵格式数据: `scan()`

读取数据框格式数据:

- `read.table()` `#sep=" ",dec="."`
- `read.csv()` `#sep=";",dec="."`
- `read.csv2()` `#sep = ";",dec=","`
- `read.delim()` `#sep = "\t",dec="."`
- `read.delim2()` `#sep = "\t",dec=","`

分别由不同的分割符号: 空格、逗号、分号或制表符分隔的数据

保存文件

□保存工作区数据

`save.image()` #将R工作空间的所有对象保存到指定文件, 默认为 `.Rdata` 中。

`save()` #将指定对象保存到文件中

`>ls()` #查看对象

`>save(x,y,z,file="mydata.rda")`

□保存为ASCII格式文件(文本文件)

`write` #适合于和`scan`同一类型的数据

`write.table()` #适合于通常采用`read.table`读取的数据类型

`write.csv()` #与`read.csv`对应

`write.csv2()` #与`read.csv2`对应

□3 存数为其它软件格式

目录



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

第一章

R简介与安装

第二章

包安装大全

第三章

数据类型与函数

第四章

文件读写

第五章

绘图



ggplot2 简介

ggplot2的基本概念

- 数据(Data)和映射(Mapping)
- 标度(Scale)
- 几何对象(Geometric)
- 统计变换(Statistics)
- 坐标系统(Coordinate)
- 图层(Layer)
- 分面(Facet)

示例数据: mpg

```
> str(mpg)
tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
 $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
 $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
 $ displ      : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
 $ year       : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999 2008 ...
 $ cyl        : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
 $ trans      : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)" "auto(av)" ...
 $ drv        : chr [1:234] "f" "f" "f" "f" ...
 $ cty        : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
 $ hwy        : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
 $ fl         : chr [1:234] "p" "p" "p" "p" ...
 $ class      : chr [1:234] "compact" "compact" "compact" "compact" ...
```

图： 图的本质是数据

五、绘图



雲南農業大學
Yunnan Agricultural University

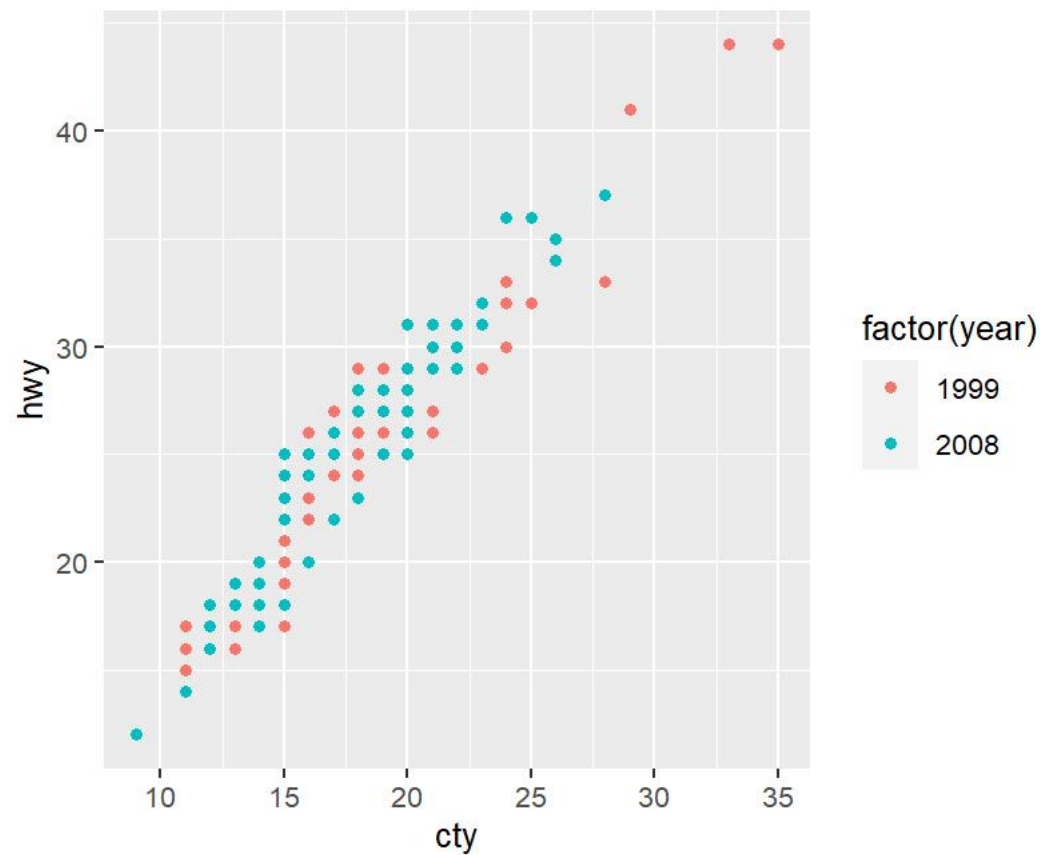
云南省生物大数据重点实验室



生信帮

散点图

```
library(ggplot2)
p <- ggplot(mpg, aes(x=cty, y=hwy, colour=factor(year)))
p <- p + geom_point()
p
```



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

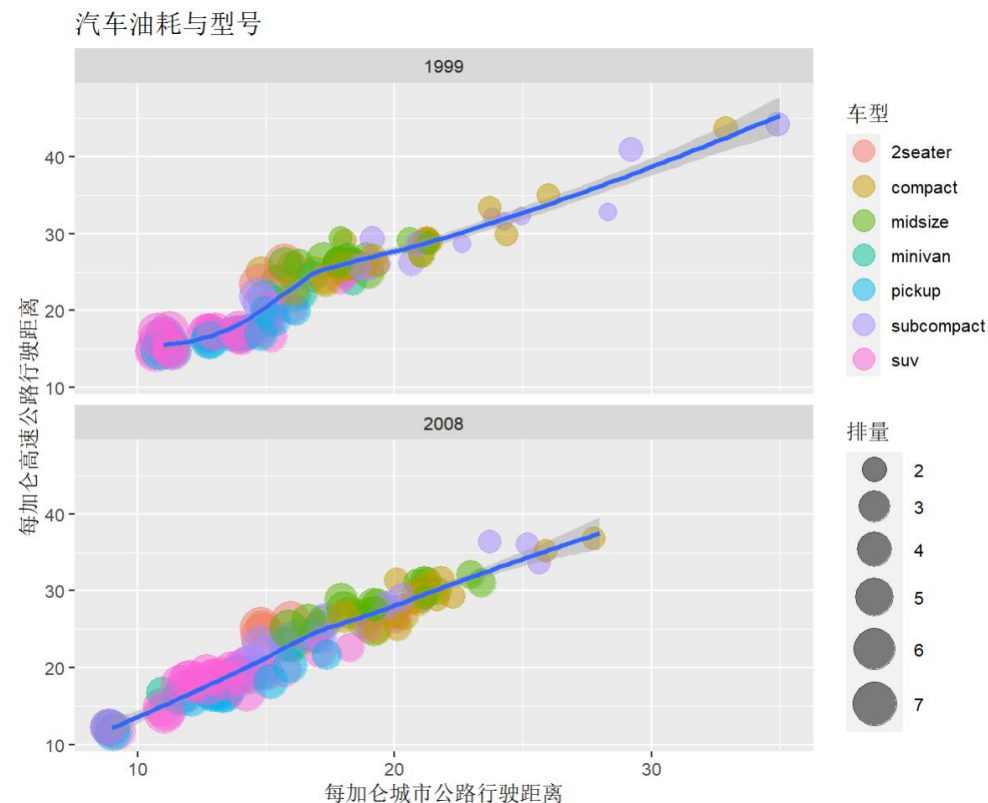


生信帮

散点图

修饰-设置坐标轴

```
p <- ggplot(mpg, aes(x=cty, y=hwy))  
p <- p + geom_point(aes(colour=class,size=displ), alpha=0.5,position = "jitter") +  
  stat_smooth() +  
  scale_size_continuous(range = c(4, 10)) +  
  facet_wrap(~ year,ncol=1) +  
  labs(title='汽车油耗与型号',  
        y='每加仑高速公路行驶距离',  
        x='每加仑城市公路行驶距离') +  
  guides(size=guide_legend(title='排量'),  
         colour = guide_legend(title='车型',override.aes=list(size=5))  
  )
```



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

散点图

实际应用：火山图

```
library(EnhancedVolcano)
```

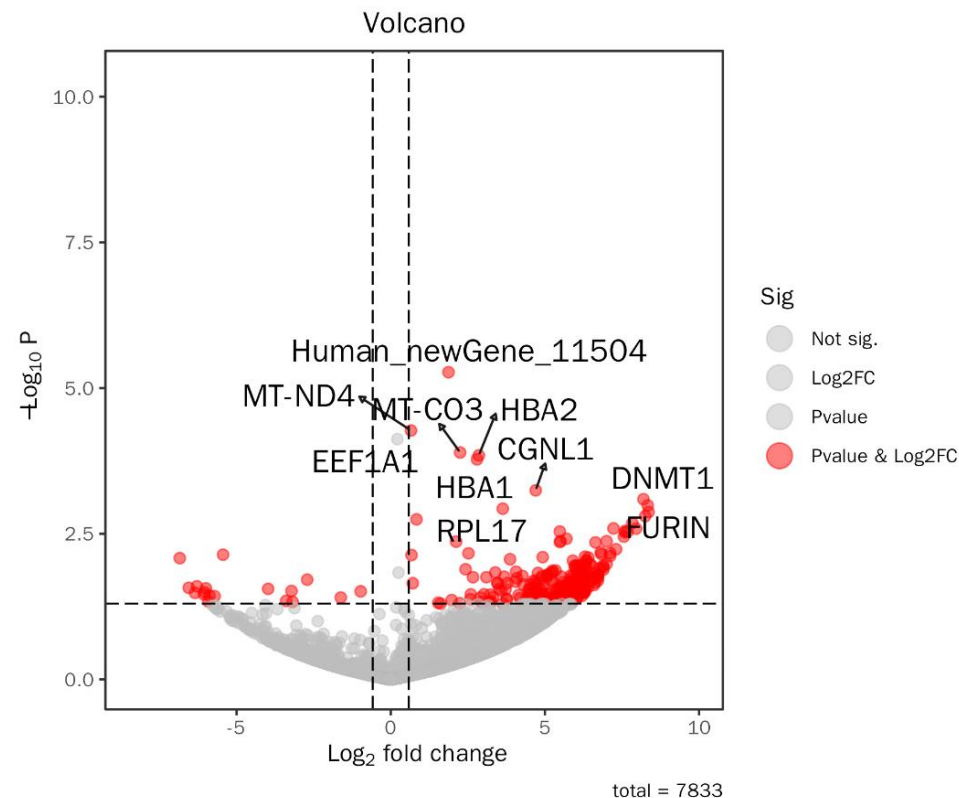
```
mytheme <- theme_bw() +  
  theme(panel.grid=element_blank()) +  
  theme(plot.title = element_text(hjust = 0.5))
```

pvalue

```
draw_data <- read.table("G1_vs_G3.all.xls", sep = "\t", header = T)  
head(draw_data)
```

```
topn <- draw_data %>% top_n(n = 10, wt = (-PValue) )
```

```
p <- EnhancedVolcano(draw_data,  
  lab = draw_data$Symbol,  
  selectLab = topn$Symbol,  
  drawConnectors = TRUE,  
  x = 'log2FC',  
  y = 'PValue',  
  title = "Volcano",  
  subtitle = NULL,  
  caption = paste0("total = ", nrow(draw_data)),  
  pCutoff = 0.05,  
  pCutoffCol = "PValue",  
  FCcutoff = log2(1.5),  
  legendLabels=c('Not sig.', 'Log2FC', 'Pvalue', 'Pvalue & Log2FC'),  
  col = c("grey","grey","grey","red"),  
  legendPosition = 'right',  
  legendLabSize = 10,  
  legendIconSize = 5.0) + mytheme
```



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

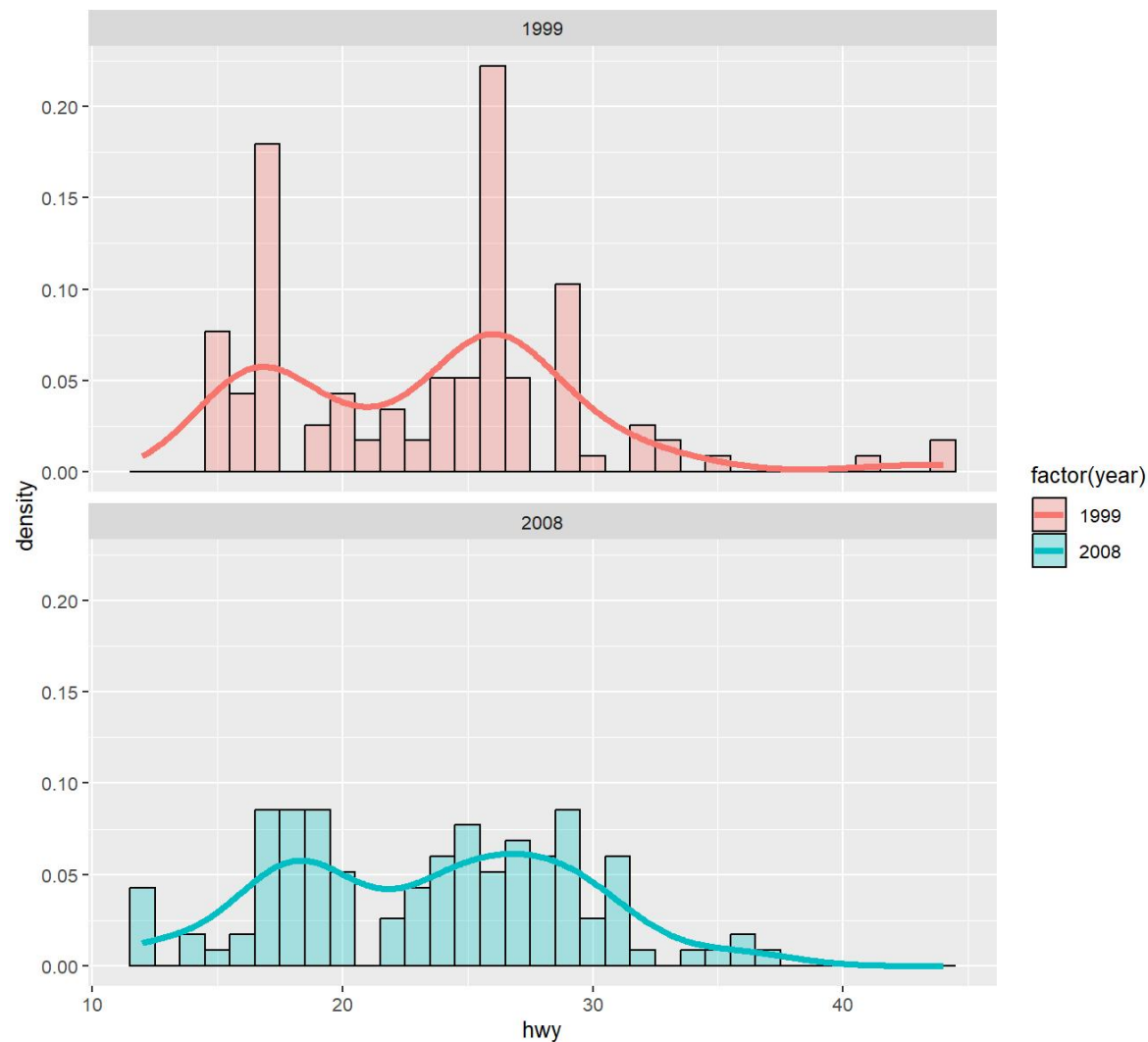


生信帮

直方图

```
p <- ggplot(mpg, aes(x=hwy))  
  
p <- p + geom_histogram(aes(fill=factor(year), y=..density..),  
  alpha=0.3, colour='black', binwidth = 1)+  
  stat_density(geom='line', position='identity', size=1.5,  
  aes(colour=factor(year)))+  
  facet_wrap(~year, ncol=1)
```

p



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

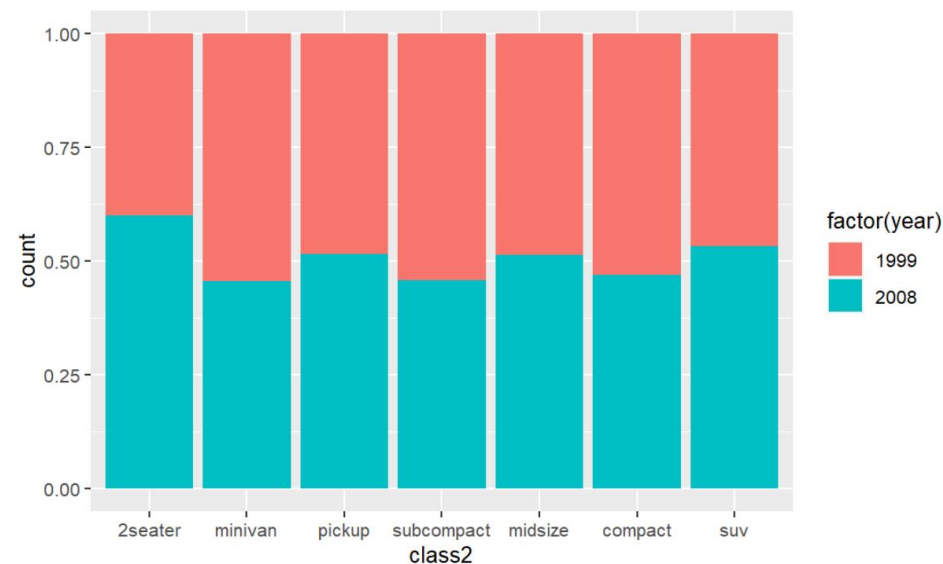
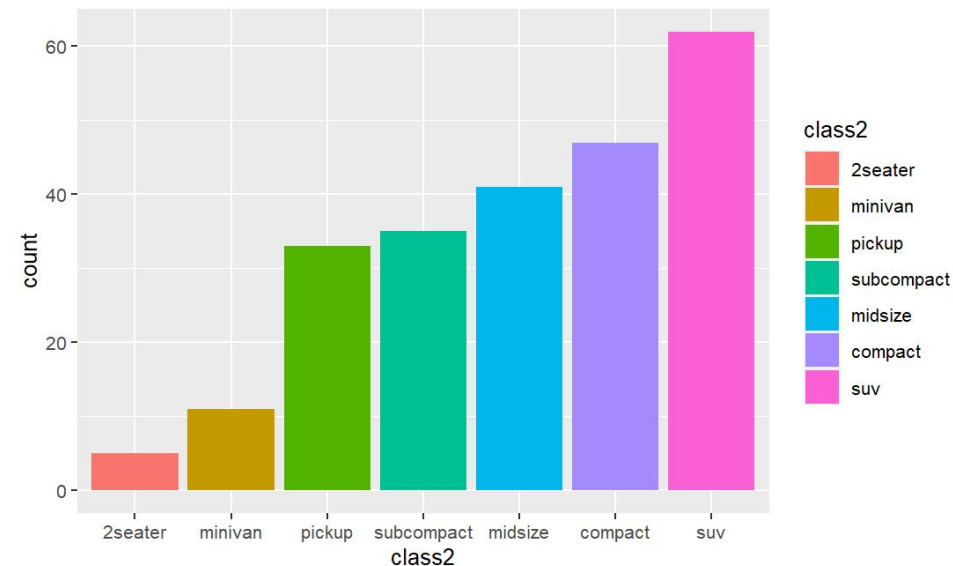
条形图

#条形图

```
class2 <- mpg$class; class2 <-reorder(class2,class2,length)
mpg$class2 <- class2
p <- ggplot(mpg, aes(x=class2))
p + geom_bar(aes(fill=class2))
```

堆积条形图

```
p <- ggplot(mpg, aes(class2,fill=factor(year)))
p+geom_bar(position='fill')
```



五、绘图



条形图

根据年份分别绘制条形图,position控制位置调整方式

```
p <- ggplot(mpg, aes(class2,fill=factor(year)))
```

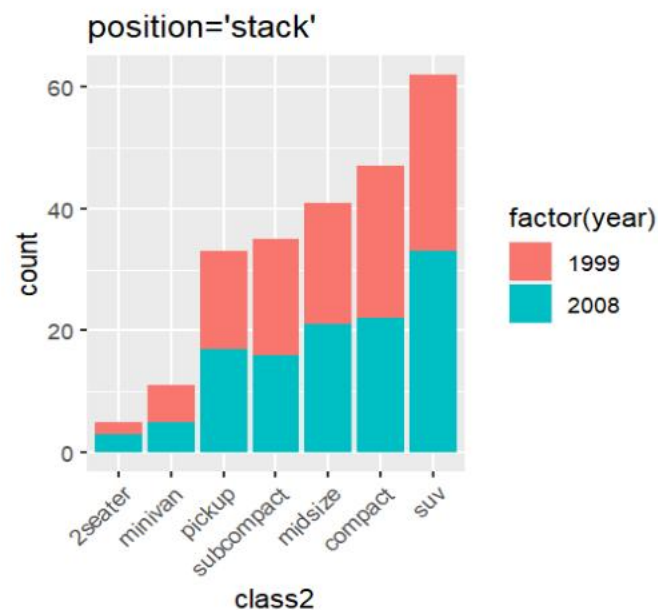
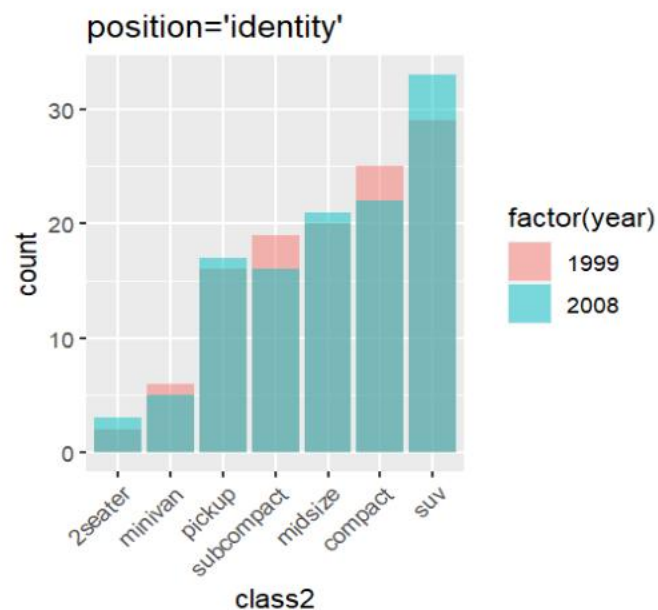
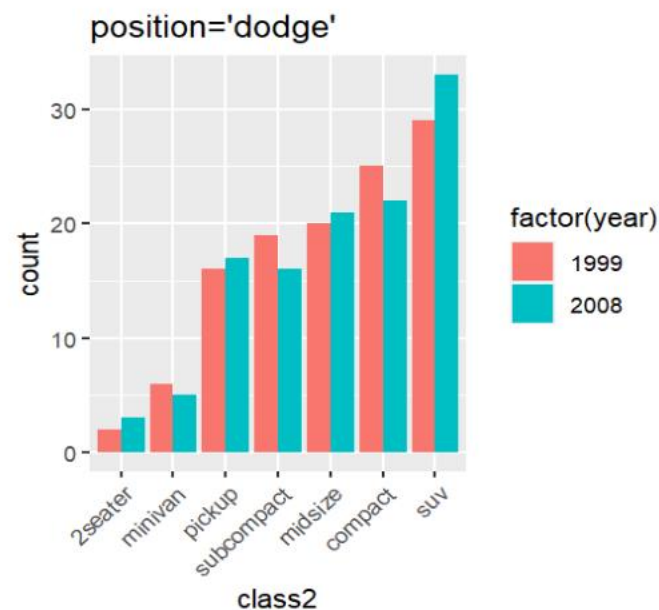
#position: 设置bar的排放位置

```
p1 <- p + geom_bar(position='dodge')
```

```
p2 <- p + geom_bar(position='identity', alpha=0.5)
```

```
p3 <- p+geom_bar(position='stack')
```

p1|p2|p3



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



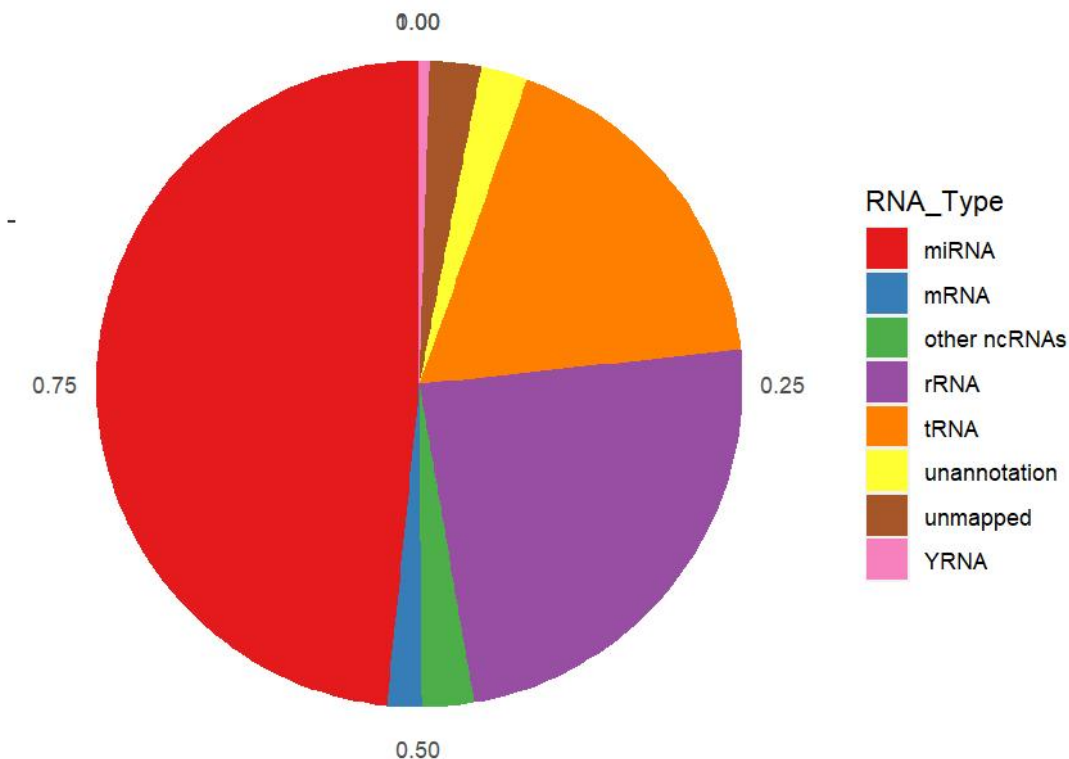
生信帮

饼图

```
library(ggplot2)
library(RColorBrewer)
library(grid)
p <- ggplot(df, aes(x="", y=V3, fill=V1))
p <- p + geom_bar(width = 1, stat="identity")
p <- p + labs(title="Pie_plot",x="",y="")+xlab("")
p <- p + scale_fill_manual(values = col)
p <- p + coord_polar(theta = "y")
p <- p + theme(axis.text = element_blank(),
               axis.ticks = element_blank(),
               panel.grid = element_blank())
p <- p + guides(fill = guide_legend(ncol = 1,
title="RNA_Type"))
p <- p + theme( panel.background =
element_rect(colour="white", size=1, fill="white"),
               panel.grid = element_blank())
```

p

Pie_plot



五、绘图



雲南農業大學
Yunnan Agricultural University

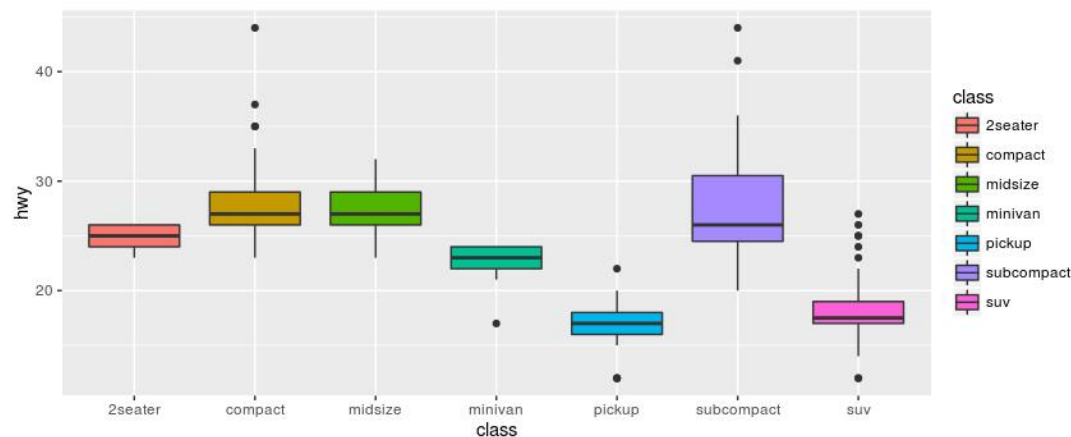
云南省生物大数据重点实验室



生信帮

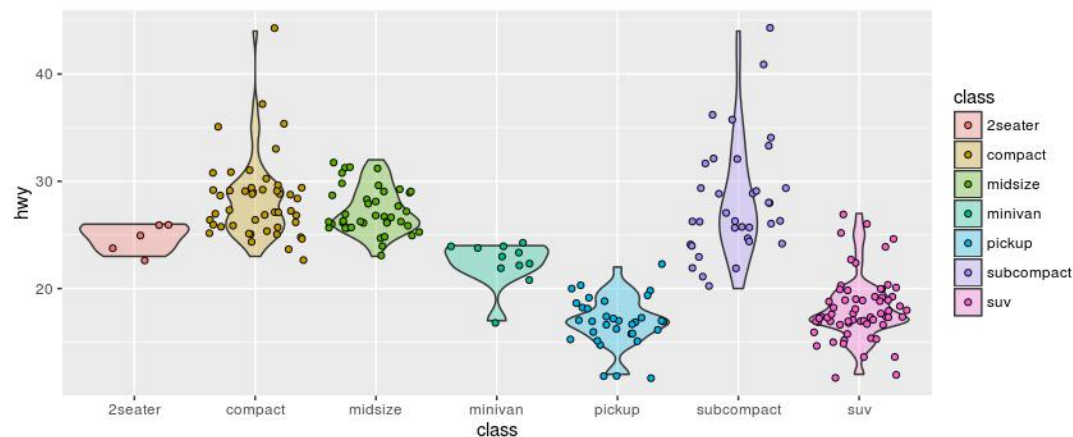
箱线图

```
p <- ggplot(mpg, aes(class,hwy,fill=class))  
p+geom_boxplot()
```



小提琴图

```
p+ geom_violin(alpha=0.3,width=0.9) +  
geom_jitter(shape=21)
```



五、绘图



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



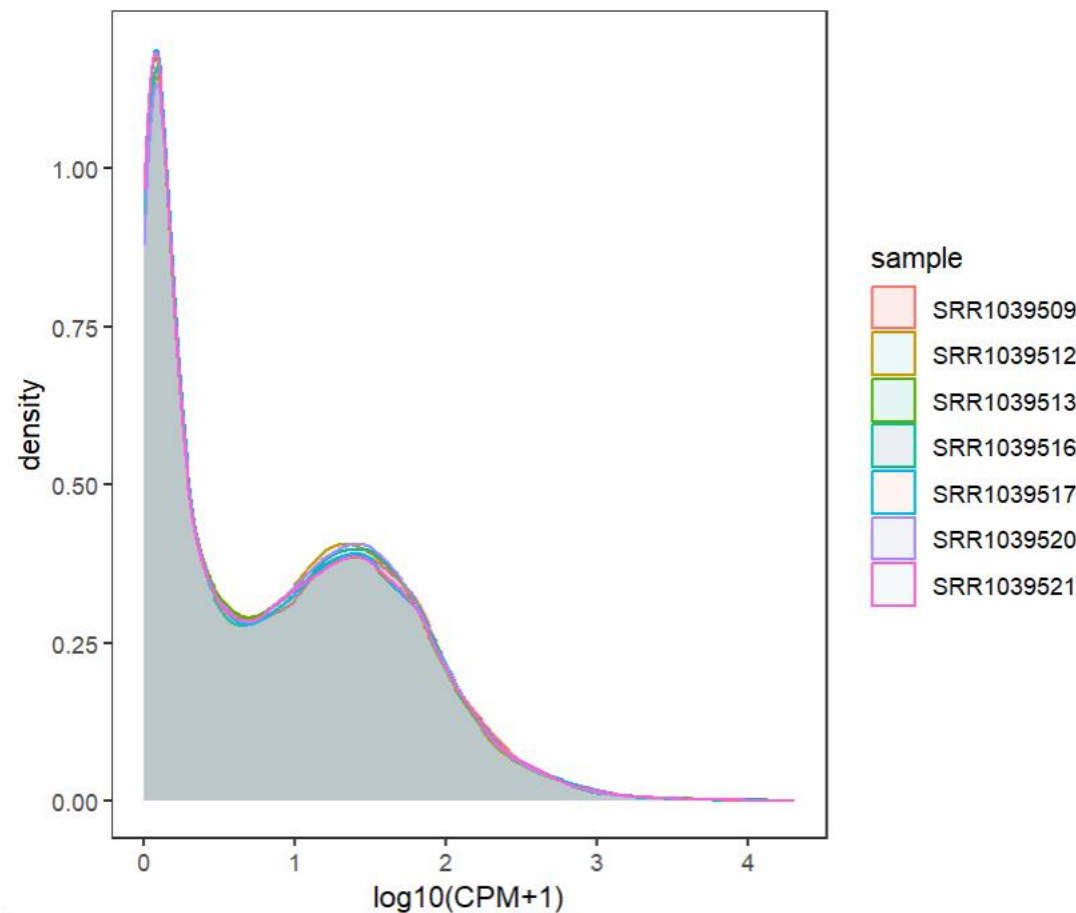
生信帮

密度曲线图

```
m <- ggplot(data=data1, aes(x=expression))
```

```
p3 <- m +  
  geom_density(aes(fill=sample,  
    colour=sample),alpha = 0.1) +  
  xlab("log10(CPM+1)") +  
  theme_bw() +  
  theme(panel.grid.major=element_blank(),  
    panel.grid.minor=element_blank()) +  
  scale_fill_npg()
```

p3



总结



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

- 掌握R包的安装
- 掌握R基本数据类型，了解如何写函数
- 掌握数据的读取和存储，并对数据进行简单的处理
- 使用ggplot2进行基础绘图：散点图、条形图、直方图等
- 课后习题：多加以练习



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室

生信帮，答你所问