



R语言入门班2020

授课人：丁颖

黄华腾教授团队

2020-09-26



R是什么？

主要用于统计分析、绘图、数据挖掘



有命令行和图形用户界面，RStudio 最为广泛使用的。



可以在多平台运行：Linux、Windows、MacOS 等



免费且开源，丰富的扩展包

为什么要学R?

➤ 面对大量重复的数据分析工作;

```
uce/uce-chr1_1019_s/RAxML_bootstrap.all  
uce/uce-chr1_1020_s/RAxML_bootstrap.all  
uce/uce-chr1_1022_s/RAxML_bootstrap.all  
uce/uce-chr1_10281_s/RAxML_bootstrap.all  
uce/uce-chr1_1036_s/RAxML_bootstrap.all  
uce/uce-chr1_1041_s/RAxML_bootstrap.all  
uce/uce-chr1_1046_s/RAxML_bootstrap.all  
uce/uce-chr1_1050_s/RAxML_bootstrap.all  
uce/uce-chr1_1056_s/RAxML_bootstrap.all  
uce/uce-chr1_1058_s/RAxML_bootstrap.all  
uce/uce-chr1_1060_s/RAxML_bootstrap.all  
uce/uce-chr1_1063_s/RAxML_bootstrap.all  
uce/uce-chr1_10887_s/RAxML_bootstrap.all  
uce/uce-chr1_1089_s/RAxML_bootstrap.all  
uce/uce-chr1_10904_s/RAxML_bootstrap.all
```

```
>1000  
CTGCATGCCAGGATGGCAACTTCAAACCTGGAGAA  
>10000  
-----GAAGATATTAACGCAGAAGAATATCC  
>10001  
AAAACTCCAAAGAGATCATTGAGTAGGAAATCCAG  
>10002  
ATGAGCAGTAATCTAGGAAAAGAAAAGGACTGTAA  
>10003  
ATGTCGGGCTCTTCCAACGTGGCGGCCATGAAGAA
```

批量修改序列名

为什么要学R?

➤用他人方法做分析;

RESOURCE ARTICLE

WILEY

**MOLECULAR ECOLOGY
RESOURCES**

TREESPACE: Statistical exploration of landscapes of phylogenetic trees

We introduce TREESPACE, an R package providing a comprehensive toolkit for the analysis of phylogenetic incongruence. We generalize

BIOINFORMATICS

ORIGINAL PAPER

Vol. 30 no. 16 2014, pages 2280–2287
doi:10.1093/bioinformatics/btu258

Phylogenetics

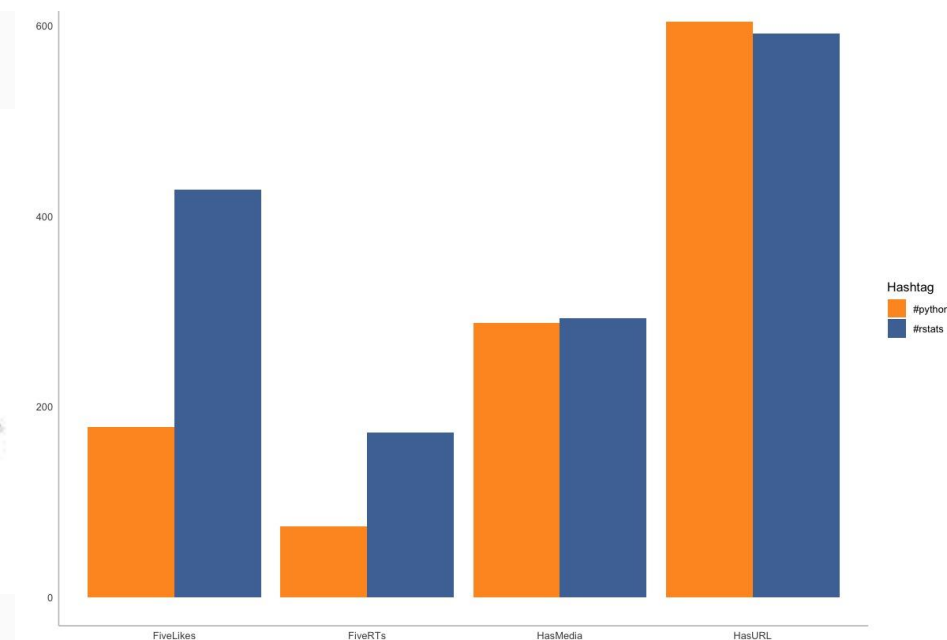
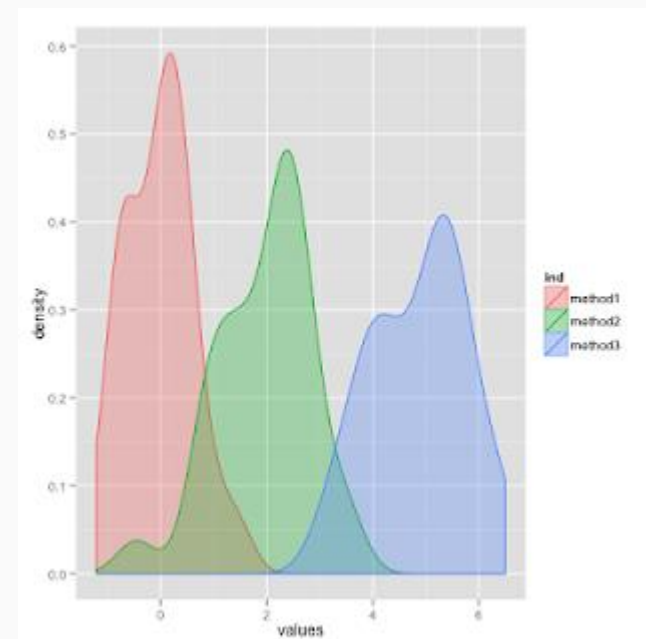
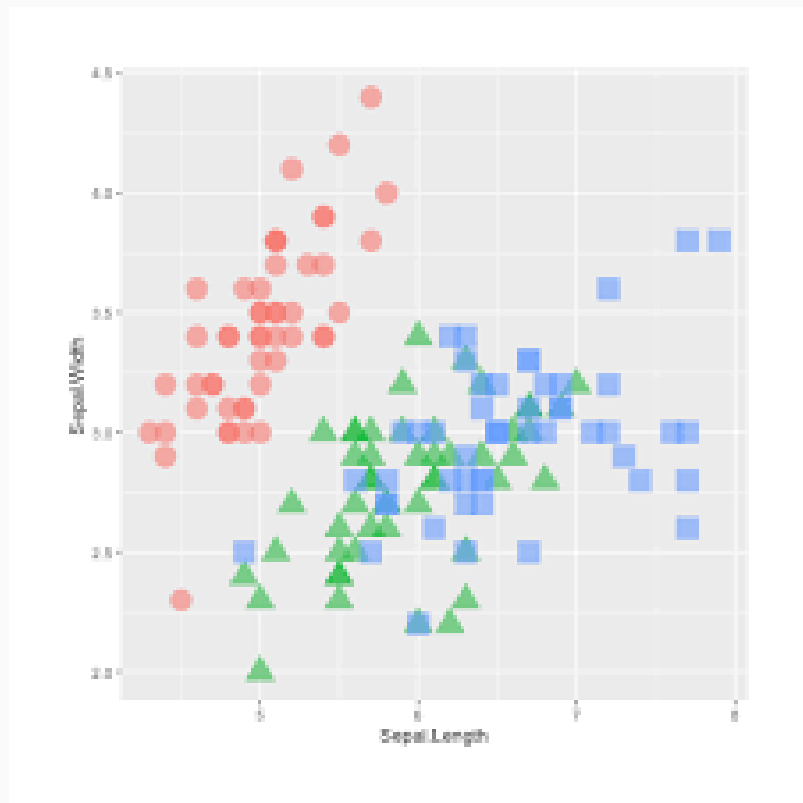
Advance Access publication April 24, 2014

KDETREES: non-parametric estimation of phylogenetic tree distributions

The KDETREES package is written in R

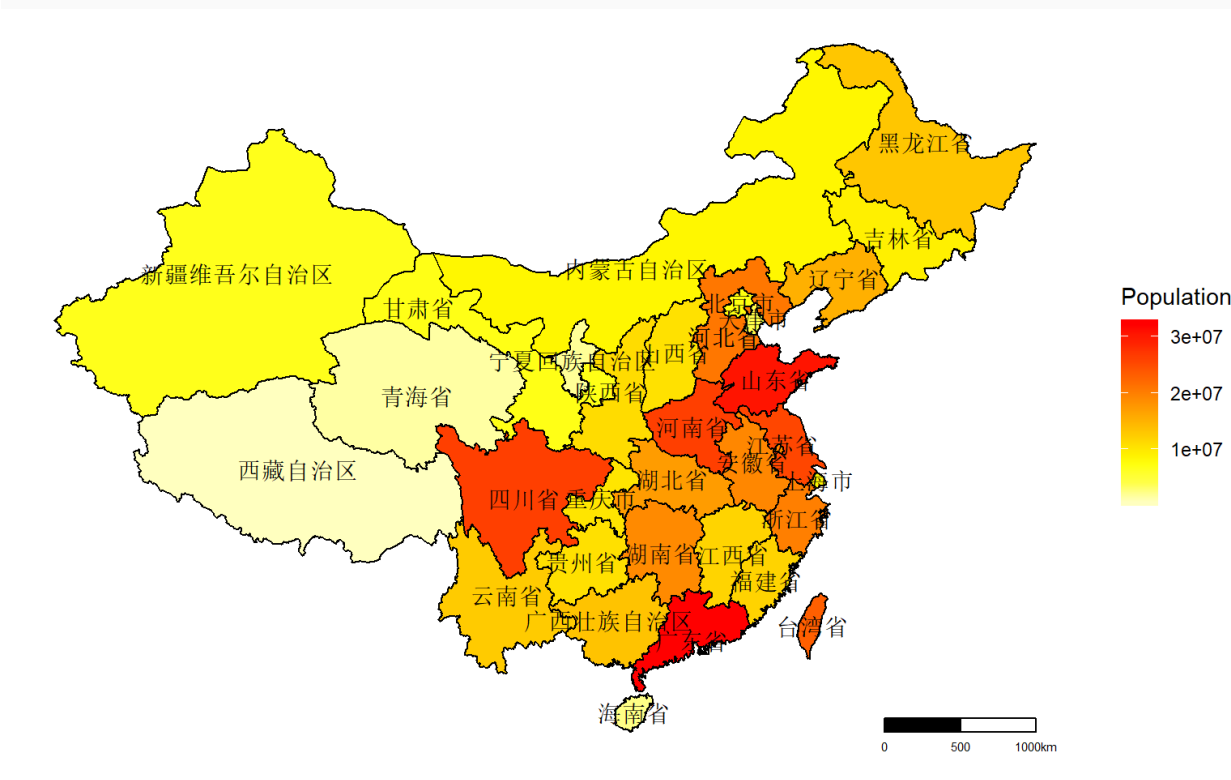
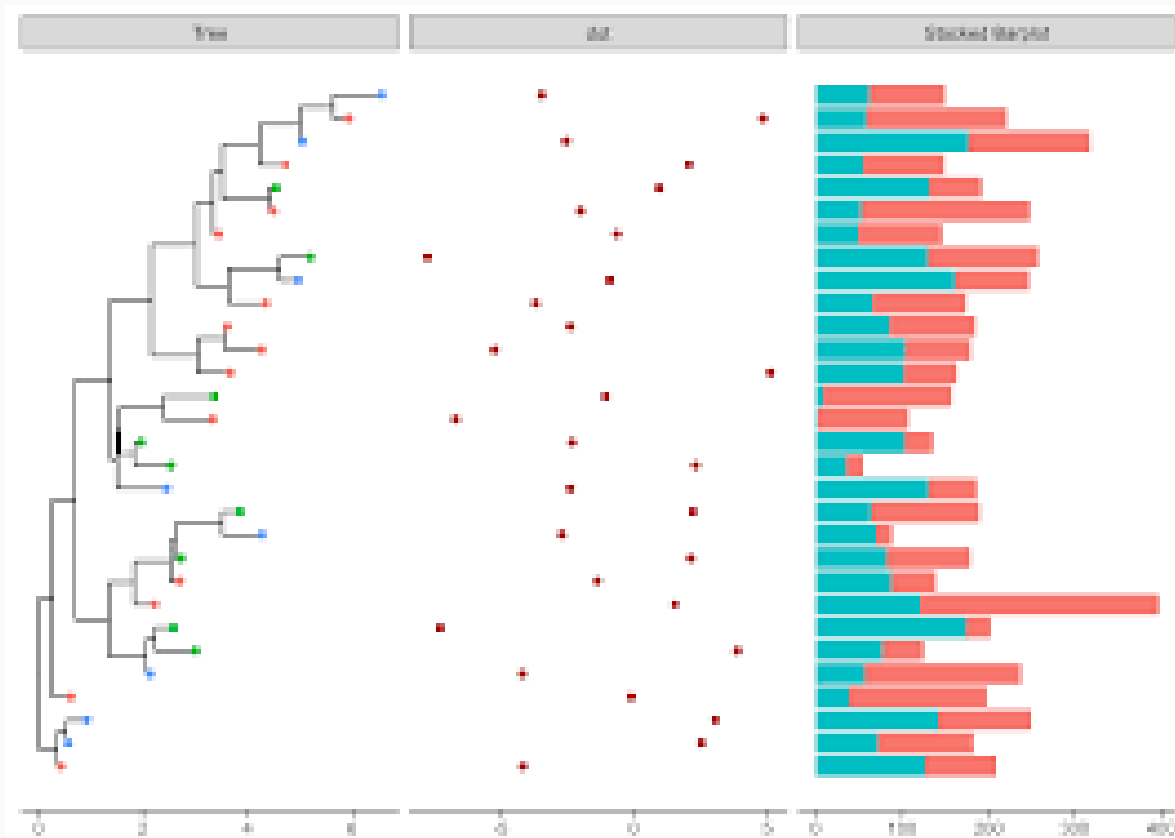
为什么要学R?

➤强大的画图功能;



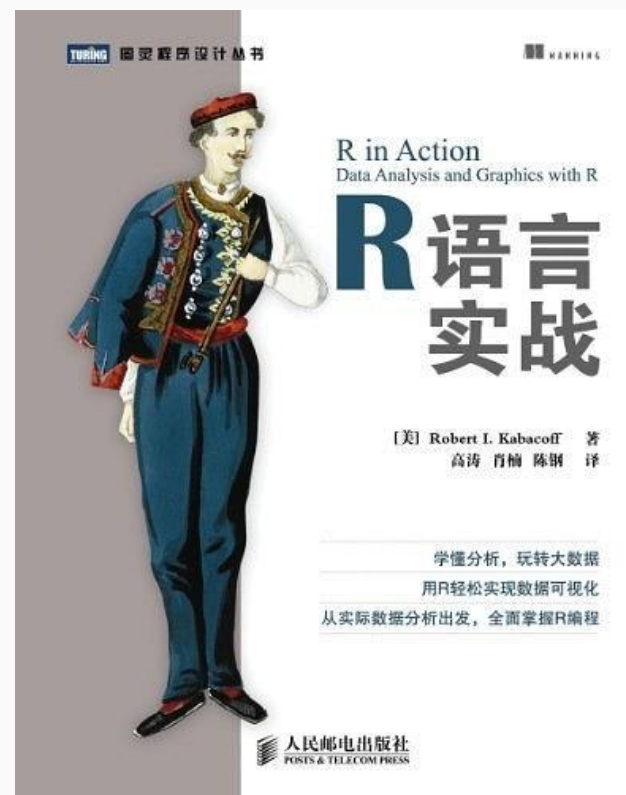
为什么要学R?

➤强大的画图功能;



推荐书籍

更适合
入门



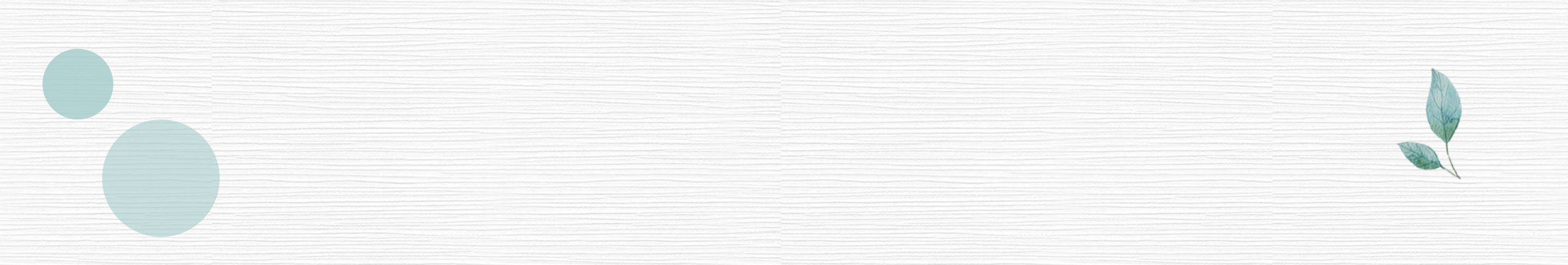
走进R & RStudio

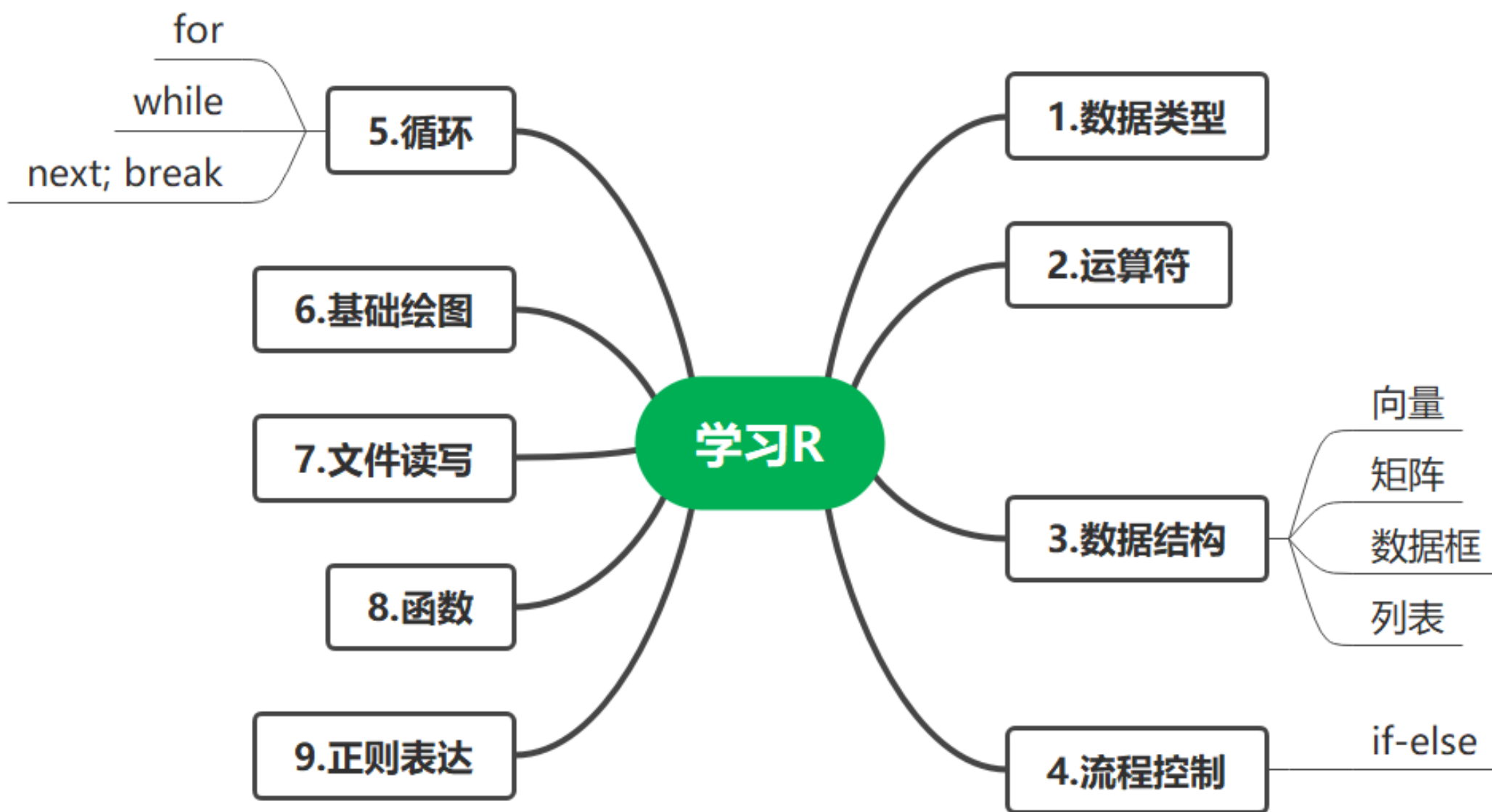
The image shows the RStudio interface with several annotations in Chinese:

- Top Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Source Editor (Top Left):** Contains a script file named 'Untitled1'. Annotations include:
 - 保存 (Save) - pointing to the Save icon.
 - 查找和替换 (Find and Replace) - pointing to the Find icon.
 - 检查脚本 (Check Script) - pointing to the Check icon.
 - 运行 (选中的部分) 快捷键ctrl+enter (Run (selected part) shortcut ctrl+enter) - pointing to the Run icon.
 - 再次运行 (Run Again) - pointing to the Run Again icon.
 - 运行全部 (Run All) - pointing to the Run All icon.
- Environment/History/Connections Panel (Top Right):** Annotations include:
 - 环境 (对象及数值) (Environment (objects and values)) - pointing to the Environment tab.
 - 在控制台敲过的历史代码 (History of code entered in the console) - pointing to the History tab.
- Console/Terminal (Bottom Left):** Contains the R startup message. Annotations include:
 - 看到大于号才可以写命令 (See the greater-than sign before writing commands) - pointing to the prompt character '>'.
 - 2.控制台, 写命令的地方 (2. Console, the place to write commands)
 - 相当于linux的命令行 (Equivalent to the Linux command line)
- Plots/Viewer Panel (Bottom Right):** Annotations include:
 - 显示作的图 (Show the plot) - pointing to the Plots tab.
 - 帮助信息 (Help information) - pointing to the Help tab.
 - 包 (扩展包) (Package (extension package)) - pointing to the Packages tab.
 - 工作目录里的文件 (Files in the working directory) - pointing to the Files tab.
 - 3
 - 4



打开Rstudio, 让我们开始学习R...





学习R

• R常用技巧

- 井号表示注释信息，不会执行的语句。加注释/取消注释快捷键：shift+ctrl+c
- 运行一行或一段Run or ctrl+enter，对于初学者建议一行一行运行
- 清空控制台ctrl+l
- `rm(list=ls())` # 清空环境
- `? + 函数名` --- 查看help文件

- R常用数据类型

数据	类型
1, 2, 3, ...	“numeric”
1L, 2L, 3L, ...	“integer”
“A”, “a”, “Good”, ...	“character”
TRUE, FALSE	“logical”

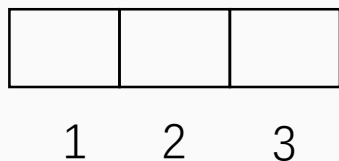
- R常用运算符

算术运算符	$+$, $-$, $*$, $/$, $\%\%$, $\%/ \%$, $\wedge$
关系运算符	$<$, $>$, $==$, $<=$, $>=$, $!=$
逻辑运算符	$\&$, $ $, $!$
赋值运算符	$<-$, $=$
其他	$\%in\%$

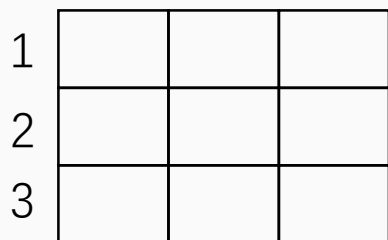
优先级：算术 > 关系 > 逻辑 > 赋值

• R常用数据结构

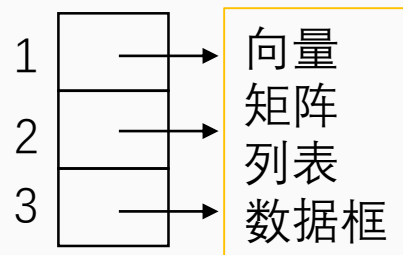
(a) 向量



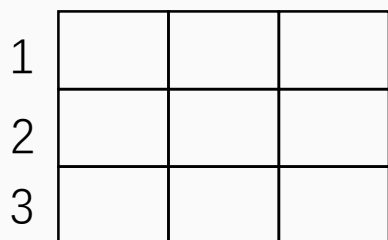
(b) 矩阵



(c) 列表



(d) 数据框

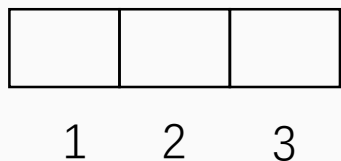


操作:

- 创建
- 引用
- 增删
- 修改元素
- 结构转换

• R常用数据结构

(a) 向量



Practice

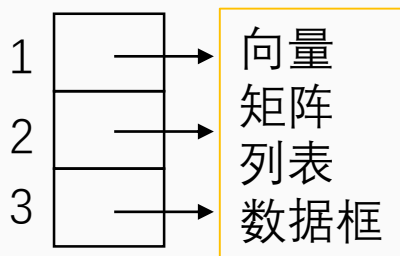
1: 删除data向量中的"Y","u","l","am".两种方法1.找到下标;2.which 与 %in% 结合

2: 查看seq函数用法, seq1变量保存100到1的所有奇数, seq2变量保存100到1的所有偶数

3: 查看sort()与order()函数用法, 对sample(1:50,10,replace = TRUE)生成的序列排序, 比较两个函数的区别

• R常用数据结构

(c) 列表



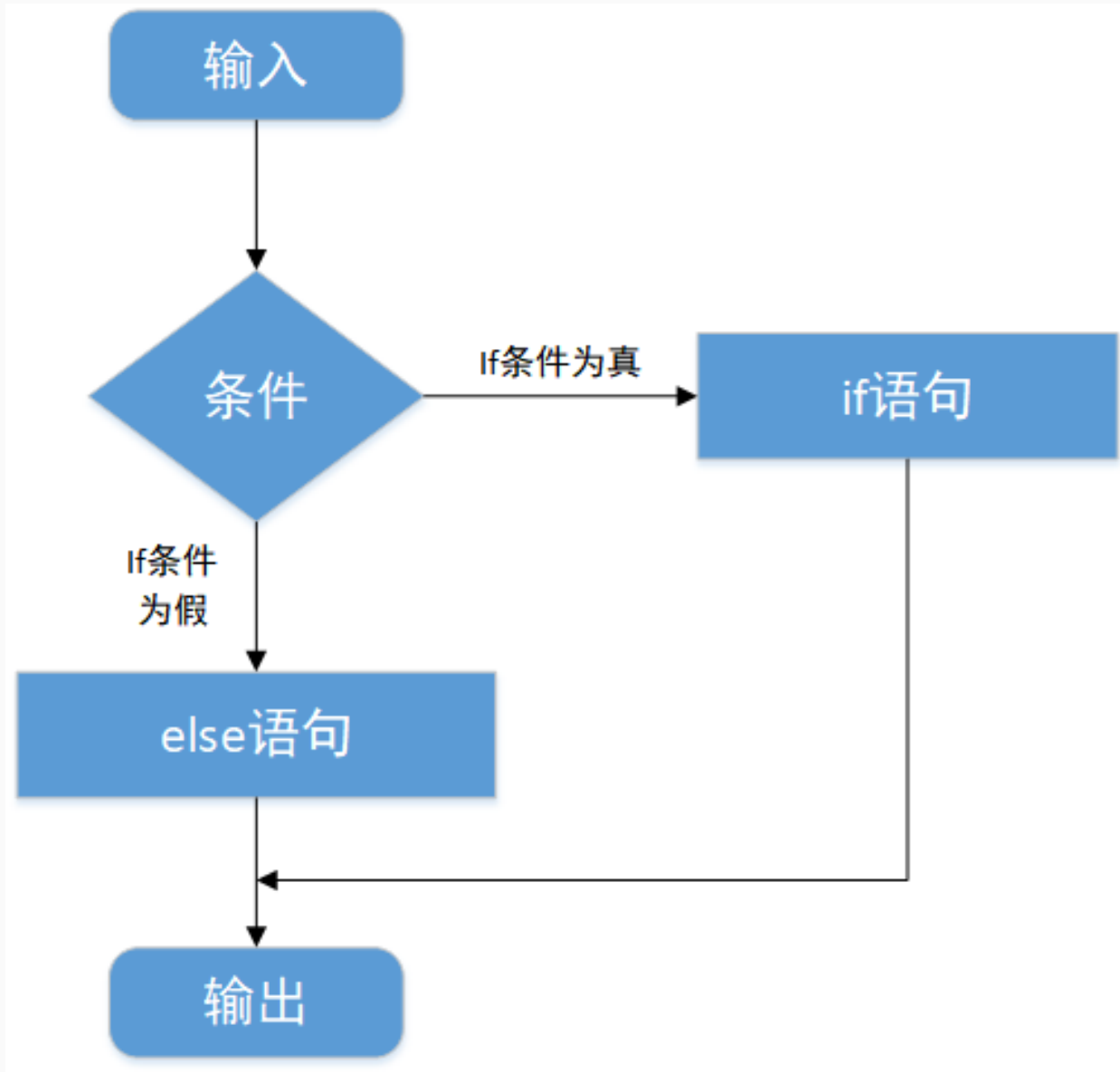
Practice

1. 得到l4列表中的字母“c”和“d”

2. 得到l4列表中的数字3

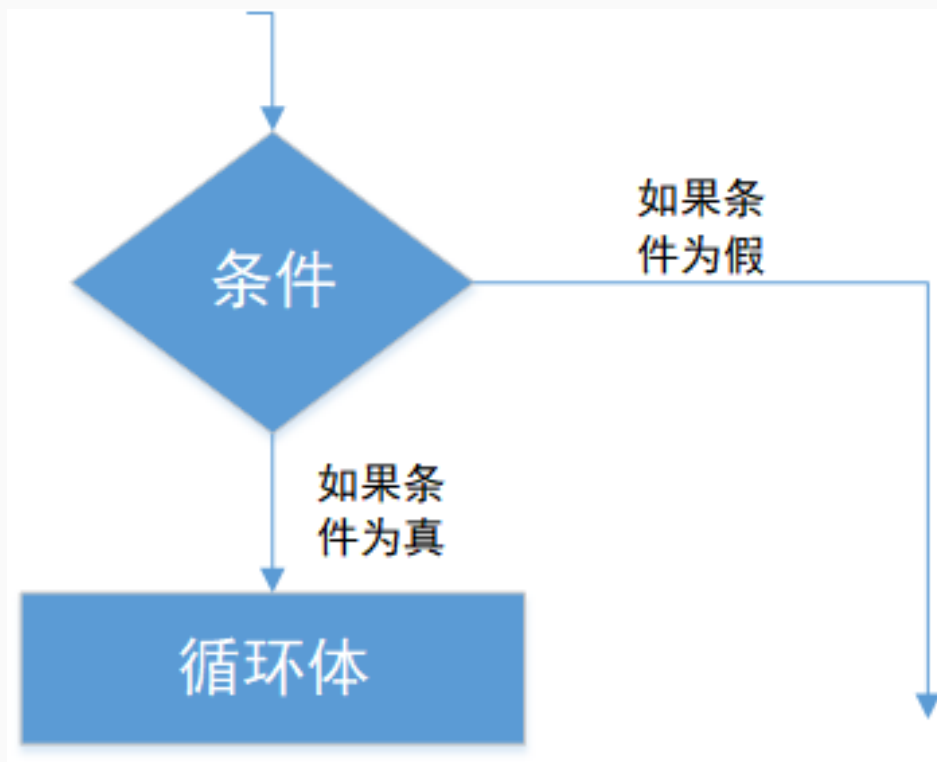
3. 创建一个列表，第一个元素包含1到5，第二个元素包含1到5的平方，第三个元素包含1到5的幂,第4个元素包含1到5的开方

- 流程控制语句

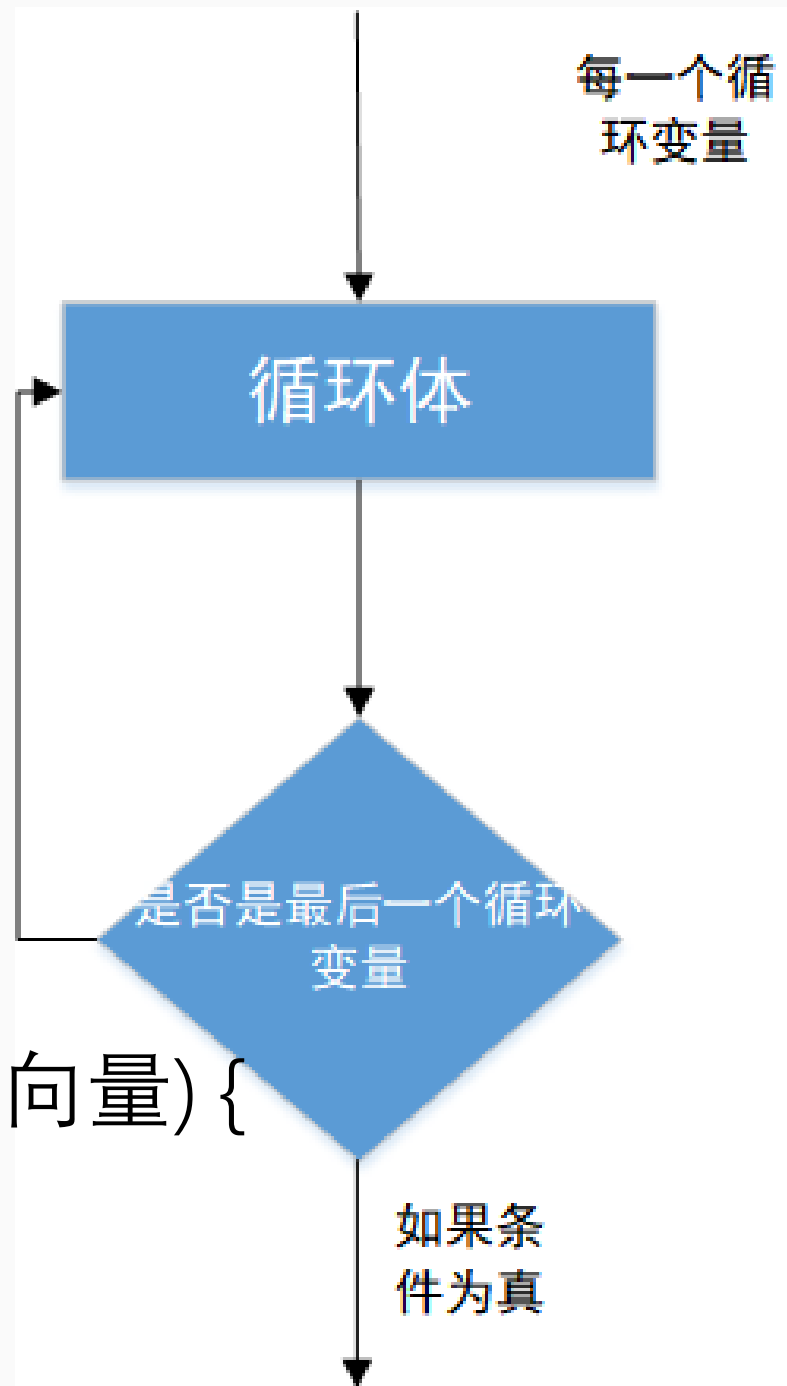


```
If(条件){  
    if语句  
}else{  
    else语句  
}
```

- 循环语句



```
While(条件){  
    循环体  
}
```



```
for(循环变量 in 向量) {  
    循环体  
}
```


- **next & break**

循环语句

{

...

next;

...

}

...

只结束本次循环



循环语句

{

...

break;

...

}

...

结束整个循环



Practice

1. 结合条件语句和循环，输出1到100的所有奇数
2. 将上步生成的所有奇数存储为向量,所有偶数存储为列表
3. 用sample可放回的从1到20生成30个随机数，输出向量中第一个为5的倍数及其前面所有的数，例随机数向量(4,6,12,5,7,10,20),则结果为(4,6,12,5)

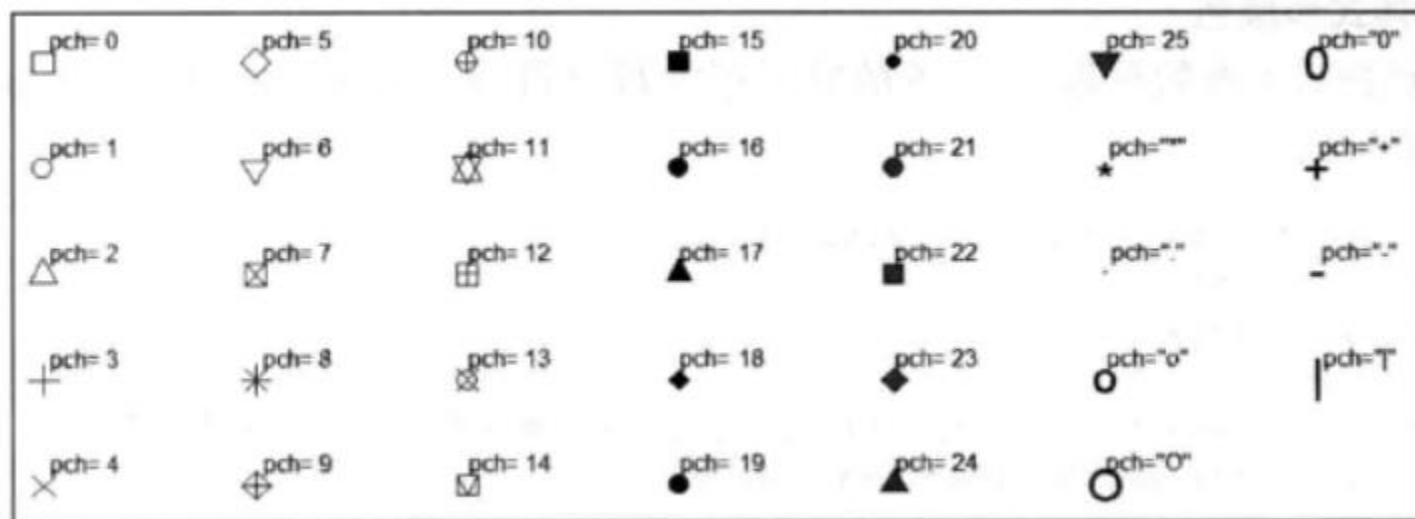
- 文件读写

- `read.csv(file,sep="")`
- `read.table(file,sep=",")`
- `scan(file)`
- `readLines(file)`

- `write.csv(data,file)`
- `write.table(data,file)`
- `write.table(data,file,sep=",")`
- `sink()+cat()`

• 基础绘图

- plot()
- points()
- lines()
- text()
- pdf()
- dev.off()



• 基础绘图

Practice

1. 读入Rworkshop文件夹中“read.data.csv”文件，以svl为横坐标，tail为纵坐标，蓝色的折线图(type="l"), 改变坐标范围，使图缩小（提示xlim, ylim），即横纵坐标范围扩大。
2. 随机生成100个服从均值为10，标准差为5的正态分布的点，画出来，颜色为“lightblue”,大小为5。
3. 在此基础上增加10个介于5-10之间的均分布的点，颜色为“purple”,形状为三角形，提示用points()函数，类似于上面提到的lines函数。
4. 保存为test.pdf
5. 读入“lizardcommLG.xls”，按date列先后顺序排列，并写到"lizardcommLG-sort.xls"文件中
6. 获取date列年份，统计每个年份的记录，画条形图
麻烦点的方法：unique函数获取非重复的年份，%in%--统计非重复的年份在所有年份中的数量
快捷版：用table函数

- # 函数

```
function_name <- function (arg_1,arg_2,……) {  
    函数体  
}
```

难点重点:

- 区分实参、形参
- 调用函数

Practice

1. paste练习，输出今天的日期，格式为年-月-日 注意区分paste函数的参数sep与collapse
2. 自定义函数，实现x的阶乘，假定x=10，提示：用到循环

• 正则表达式

- grep()
- gsub()

表达式: <https://tool.oschina.net/uploads/apidocs/jquery/regexp.html>

Practice

1. 读入文件中的“genefile.fasta”文件; 将所有序列号提取出来写入“seqname.txt”文件;

这道题有点难



• 案例课堂练习

1. 已从NCBI下载Orthoptera CO1 591条序列，存入sequence.fasta。（练习读写）
2. 假设序列已经alignment,接下来对下载的序列要建树，那我们希望每条序列名只保留物种名，每条序列写成一行（练习正则表达式，循环语句，数据结构操作）

还可以在学R?

慕课网-程序员的梦工厂



下载APP 慕课网APP - 随时随地学编程App Store下载 Android下载
扫描下载慕课网APP购物车 登录 / 注册 Java入门 前端入门 ...
<https://www.imooc.com/> - 百度快照 - 187条评价

生信技能树



哔哩哔哩(゜-゜)つロ 干杯~-bilibili 官网



bilibili是国内知名的视频弹幕网站,这里有最及时的动漫新番,最棒的ACG氛围,最有创意的Up主。大家可以在这里找到许多欢乐。
哔哩哔哩 - V1 - 百度快照 - 2705条评价