

# R 语言基础

经济管理学院, 北京航空航天大学

June 5, 2019

# 目录

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理
- 5 数据读入
- 6 R Function
- 7 R综合应用

# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理
- 5 数据读入
- 6 R Function
- 7 R综合应用

# R


- 是一门用于数据分析和图形展示的语言, 平台和环境;
- 是一种统计分析, 绘图语言, 也是实现该语言的软件;
- 是一种自由软件编程语言与操作环境, 主要用于统计分析、绘图、数据挖掘;
- R语言是在统计和数据科学界广泛应用的编程语言和开发环境, 其[免费](#)、[开源](#)、[灵活](#)的特点


# 为什么学习R?


## 找工作

**Jobs for R-users**  
A job board for people and companies looking to hire R users

Latest

All Jobs Location 

Radius: Auto 

**Featured Jobs** 

<b>Full-Time</b>	<b>Data Analytics Associate</b> <i>Income Discovery - Posted by <a href="#">ANaman</a></i>	Hoboken New Jersey, United States	11 Jul 2016
<b>Full-Time</b>	<b>R Programming Rock Star (10080)</b> <i>Object Systems International - Posted by <a href="#">mideen@objectsystems.com</a></i>	Salt Lake City Utah, United States	7 Jul 2016
<b>Full-Time</b>	<b>Data Scientist / Quantitative Analyst</b> <i>Sporting Data Limited - Posted by <a href="#">sportingdata</a></i>	London England, United Kingdom	27 Jun 2016
<b>Full-Time</b>	<b>Senior Data Scientist</b> <i>Global Strategy Group - Posted by <a href="#">datasaurus</a></i>	New York New York, United States	20 Jun 2016
<b>Part-Time</b>	<b>problem solver</b> <i>IdeaConnection LTD - Posted by <a href="#">NKVanDerwaarden</a></i>	Anywhere	17 Jun 2016



# 为什么学习R?

- 物美价廉:作图颜值高，且完全免费
- 兼收并蓄:算法覆盖广，扩展易
  - 作为统计分析工具，它几乎覆盖整个统计领域的前沿算法。从火到根本停不下来的神经网络（就是那只下围棋的机器狗脑袋里的东西），到经典了一百多年还熠熠生辉的基础回归，数千个R包，上万种算法，你都能找到可直接调用的函数实现。
  - 作为一款软件系统，它有极方便的扩展性。你的数据原来存在Oracle 里？没问题，轻松导入；你的数据在MySQL 里？没问题，照样解决。文本文件、数据库管理系统，统计软件，专门的数据仓库都可兼容。它同样可以将数据输出并写入到这些系统中。同时它还可以轻松于各种语言完成互调，比如你曾经爱的python，还有C，都可无缝对接。
- 集千万力量于一身:强大的社区支持
  - 国外比较活跃的社区有GitHub和Stack Overflow等，通常R包的开发者会先将代码放到GitHub，接受世界各地的使用者提出问题，修改代码等操作，等代码成熟后再放到CRAN上发布
  - 国内最活跃的R社区就属统计之都以及统计之都旗下的COS论坛。统计之都都会经常发布与R相关的优质文章，还会不定期举办线下研讨会，以及规模巨大的R语言会议；而COS论坛则是中文R语言技术问答社区，它们都对R 语言学习者具有很高的参考价值

# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理
- 5 数据读入
- 6 R Function
- 7 R综合应用

# R 软件的下载与安装

<https://cran.r-project.org/>



[Home]

## Download

CRAN

## R Project

About R

Logo

Contributors

What's New?

Mailing Lists

Bug Tracking

Development Site

Conferences

Search

## R Foundation

Foundation

Board

Members

Donors

Donate

## Documentation

Manuals

FAQs

The R Journal

Books

Certification

Other

## The R Project for Statistical Computing

### Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our **answers to frequently asked questions** before you send an email.

### News

- **R version 3.3.0 (Supposedly Educational)** has been released on 2016-05-03.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.



# R 软件的下载与安装

## China

<https://mirrors.tuna.tsinghua.edu.cn/CRAN/>

<http://mirrors.tuna.tsinghua.edu.cn/CRAN/>

<https://mirrors.ustc.edu.cn/CRAN/>

<http://mirrors.ustc.edu.cn/CRAN/>

<https://mirror-hk.koddos.net/CRAN/>

<https://mirrors.eliteu.cn/CRAN/>

<https://mirror.lzu.edu.cn/CRAN/>

<http://mirror.lzu.edu.cn/CRAN/>

<https://mirrors.tongji.edu.cn/CRAN/>

<https://mirrors.shu.edu.cn/CRAN/>

TUNA Team, Tsinghua University

TUNA Team, Tsinghua University

University of Science and Technology of China

University of Science and Technology of China

KoDDoS in Hong Kong

Elite Education

Lanzhou University Open Source Society

Lanzhou University Open Source Society

Tongji University

Shanghai University

# R 软件的下载与安装



CRAN

[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R

[R Homepage](#)  
[The R Journal](#)

Software

[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation

[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2016-05-03, Supposedly Educational) [R-3.3.0.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

### Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

- Windows 下载和安装R-CRAN: Binaries>Windows>base

[Download R 3.2.3 for Windows](#) (62 megabytes, 32/64 bit)  
[Installation and other instructions](#)  
[New features in this version](#)

Figure: R 3.2.3 for Windows 下载页面.

下载完成后, 双击[R-3.2.3-win32.exe](#) 开始安装. 点击下一步, 各选项默认, 语言建议选择中文.

- R 操作界面展示: File, Edit, Packages, Help

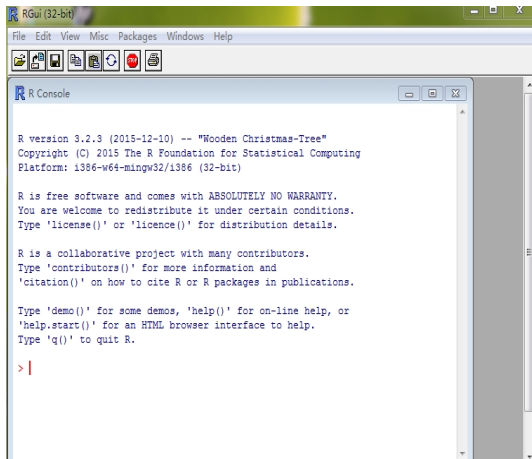
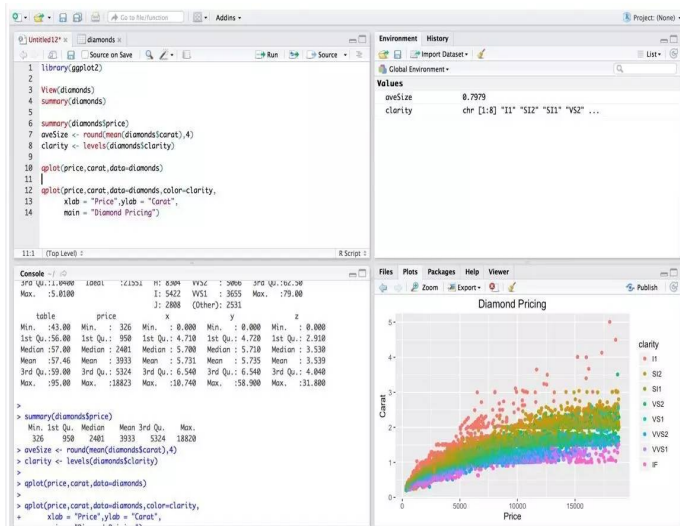


Figure: R 软件界面

- #: 说明语句字符(增加程序的可读性)
- < -: 赋值
- c(): 表示数组
- mean(),sd() 分别表示均值和标准差
- plot(), hist()
- list(), data.frame()
- 帮助文件: pdf帮助文档; help or ? command (?hist, help('hist'))

# R的升级版: RStudio

## R语言的集成开发环境



# RStudio

The screenshot shows the RStudio environment with several handwritten annotations in pink clouds:

- 代码编辑器** (Code Editor): Points to the script editor window.
- 控制台** (Console): Points to the console window.
- 运行** (Run): Points to the 'Run' button in the toolbar.
- 工作空间** (Workspace): Points to the Environment pane.
- 画图与求助台** (Plotting and Help): Points to the Plots pane.

The script editor contains the following R code:

```
1 #读入数据
2 data(cars)
3 #查看前六行数据
4 head(cars)
5 #了解数据的情况
6 summary(cars)
7 #看一下两变量的关系
8 plot(cars,main="速度与刹车距离的关系",xlab="速度",ylab="刹车距离",col=6)
9
```

The console shows the output of the commands:

```
> head(cars)
  speed dist
1    4.0   2
2    4.0  10
3    7.0   4
4    7.0  22
5    8.0  16
6    9.0  10

> #了解数据的情况
> summary(cars)
      speed      dist
Min.   : 4.0   Min.   : 2.00
1st Qu.:12.0   1st Qu.: 26.00
Median :15.0   Median : 36.00
Mean   :15.4   Mean   : 42.98
3rd Qu.:19.0   3rd Qu.: 56.00
Max.   :25.0   Max.   :120.00

> #看一下两变量的关系
> plot(cars,main="速度与刹车距离的关系",xlab="速度",ylab="刹车距离",col=6)
```

The Plots pane displays a scatter plot titled "速度与刹车距离的关系" (Relationship between Speed and Stopping Distance). The x-axis is labeled "速度" (Speed) and ranges from 0 to 25. The y-axis is labeled "刹车距离" (Stopping Distance) and ranges from 0 to 100. The plot shows a positive correlation between speed and stopping distance, with data points colored in light blue.

# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装**
- 4 R的数据组织和整理
- 5 数据读入
- 6 R Function
- 7 R综合应用



# R package

- R 包: R函数,数据,预编译代码以一种定义晚上的格式组织在一起的集合
- 通过CRAN服务器安装: `install.packages('package_name');`  
`library('package_name')`
- 通过Github下载安装(补充安装方式): `library('devtools');`  
`install_github('author/package_name')`
- 检查更新: `update.packages()`
- `help()`

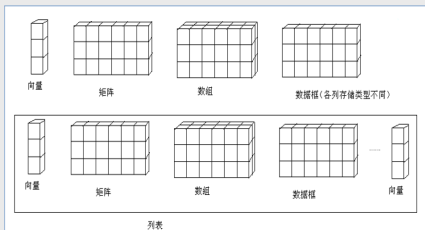
# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理**
- 5 数据读入
- 6 R Function
- 7 R综合应用

# R的数据对象

R的数据对象有哪些类型

- 从存储角度划分R对象：数值型、字符型、逻辑型
- 从结构角度划分R对象：向量、矩阵、数组、数据框、列表



# 创建和访问R的数据对象

- 创建R的对象
  - 对象名  $\leftarrow$  R常量或R函数
- 访问R的数据对象
  - 对象名, 或, `print(对象名)`
- 查看R数据对象的结构
  - `str(对象名)`
- 管理R的数据对象
  - `ls()`、`rm(对象名或对象名列表)`, 或, `remove(对象名)`

# 向量的创建和访问

- 向量可以只包含一个元素，存储一个数据，也可以包含若干个元素，存储多个数据。向量包含的元素可以是数值型、字符串型或逻辑型，对应的向量依次称为数值型向量、字符串型向量或逻辑型向量
- 创建只包含一个元素的向量：对象名 <- R常量
- 创建包含多个元素的向量
  - c函数、rep重复函数、seq序列函数
  - scan键盘数据读入函数
  - vector创建向量函数

# 向量的创建和访问

## ■ 访问向量中的元素

### ■ 访问指定位置上的元素

- 向量名[位置常量]、向量名[位置常量1:位置常量2]、向量名[c(位置常量列表)]

### ■ 利用位置向量访问指定位置上的元素

- 向量名[位置向量名]

### ■ 访问指定位置之外的元素

- 向量名[-位置常量]、向量名[-(位置常量1:位置常量2)]、向量名[-c(位置常量列表)]、向量名[-位置向量名]

# 矩阵的创建和访问

- 矩阵用来组织具有相同存储类型的一组变量。矩阵元素可以是数值型、字符串型或逻辑型，对应的矩阵依次称为数值型矩阵、字符串型矩阵或逻辑型矩阵
- 创建矩阵
  - 向量合并形成矩阵
    - `cbind`(向量名列表)、`dim`(矩阵名)、`colnames`(矩阵名)、`rownames`(矩阵名)
  - 由单个向量派生矩阵
    - `matrix`(向量名, `nrow`=行数, `ncol`=列数, `byrow`=TRUE/FALSE, `dimnames`=list(行名称向量, 列名称向量))

# 矩阵的创建和访问

## ■ 访问矩阵

### ■ 访问指定位置上的元素

- 矩阵名[行位置常量, 列位置常量]、矩阵名[行位置常量1:行位置常量2, 列位置常量1:列位置常量2]、矩阵名[c(行位置常量列表), c(列位置常量列表)]

### ■ 访问指定行上的所有元素

- head(矩阵名, n)、tail(矩阵名, n)
- 矩阵名[行位置常量, ]、矩阵名[行位置常量1:行位置常量2, ]、矩阵名[c(行位置常量列表), ]、矩阵名[行位置向量名, ]



# 矩阵的创建和访问

## ■ 访问矩阵

### ■ 访问指定列上的所有元素

■ 矩阵名[, 列位置常量]、矩阵名[, 列位置常量1:  
列位置常量2]、矩阵名[, c(列位置常量列表)]、  
矩阵名[, 列位置向量名]

### ■ 利用编辑窗口访问矩阵元素

■ fix(矩阵名)

# 数据框的创建和访问

- 数据框也是一张二维表格，行和列在统计上分别称为观测和变量，计算机上分别称为记录和域。变量名的对应称谓是域名，变量值对应域值
- 创建数据框
  - `data.frame`(域名1=向量名1, 域名2=向量名2, ...)
  - `names`(数据框名)
- 访问数据框
  - 数据框名\$域名、数据框名[[“域名”]]、数据框名[[域编号]]
  - `attach`函数和`detach`函数、`with`函数

# 列表的创建和访问

- 列表是对象的集合，可包含向量、矩阵、数组、数据框甚至列表等等。其中的每个对象称为列表的一个成分，且均有一个成分名
- 创建列表
  - `list(成分名1=对象名1, 成分名2=对象名2, ……)`
- 访问列表
  - 列表的访问方式与数据框完全相同

# 数据对象的相互转换

- 不同存储类型之间的转换
  - 判断数据对象的存储类型
    - is. 存储类型名(数据对象名)、typeof(数据对象名)
  - 数据对象存储类型的转换
    - as. 存储类型名(数据对象名)
- 不同结构类型之间的转换
  - 向量和矩阵之间的互转
    - as.matrix(向量名)、as.vector(矩阵名)

# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理
- 5 数据读入**
- 6 R Function
- 7 R综合应用

# 读、写数据文件

- 纯文本文档 `read.table()`
- SPSS, SAS, S-PLUS, Stata 数据文件: `library('foreign')`
  - SPSS: `read.spss()`
  - SAS: `read.xport()`
  - S-PLUS: `read.S()`
  - Stata: `read.dta()`
- Excel 数据文件
  - Excel 表转化为“文本文件(制表符分隔)”: `read.delim()`
  - Excel 表转化为“CSV(逗号分隔)”: `read.csv()`
  - 直接读取: `library('RODBC'); odbcConnectExcel(); sqlFetch()`
- 数据库: `data()`; e.g., `data('CO2')`
- `write.table()`; `write.csv()`

# Contents

- 1 什么是R?
- 2 初识R软件
- 3 R包的获取与安装
- 4 R的数据组织和整理
- 5 数据读入
- 6 R Function**
- 7 R综合应用

- 如何提高R程序处理的能力：流程控制
- 分支结构的流程控制：分支结构的流程控制是指R程序在某处的执行取决于某个条件。当条件满足时执行一段程序，当条件不满足时执行另外一段程序。因程序的执行在该点出现了“分支”，因而得名分支结构的流程控制
  - if结构
  - if-else结构
  - switch结构



## ■ 循环结构的流程控制

- 循环结构的流程控制是指R程序在某处开始，根据条件判断结果决定是否反复执行某个程序段

- for结构

- while结构

- repeat结构

- 函数的基本形式: 变量名 = function(变量列表) 函数体
- 控制结构
  - 条件语句
    - if (条件) 表达式1 else 表达式2
    - ifelse(条件, yes, no)
    - if(x>=0) sqrt(x) else NA  
ifelse(x>=0,sqrt(x),NA)
  - 循环语句
    - for(变量in 向量)表达式
    - while(条件)表达式
    - repeat 表达式(break跳出循环)

See R code in RFun.R.

Much thanks !