

An Introduction to R

R语言简介

Outline

- ▶ R与Rstudio
- ▶ R Markdown
- ▶ 基础语法与操作
- ▶ 练习

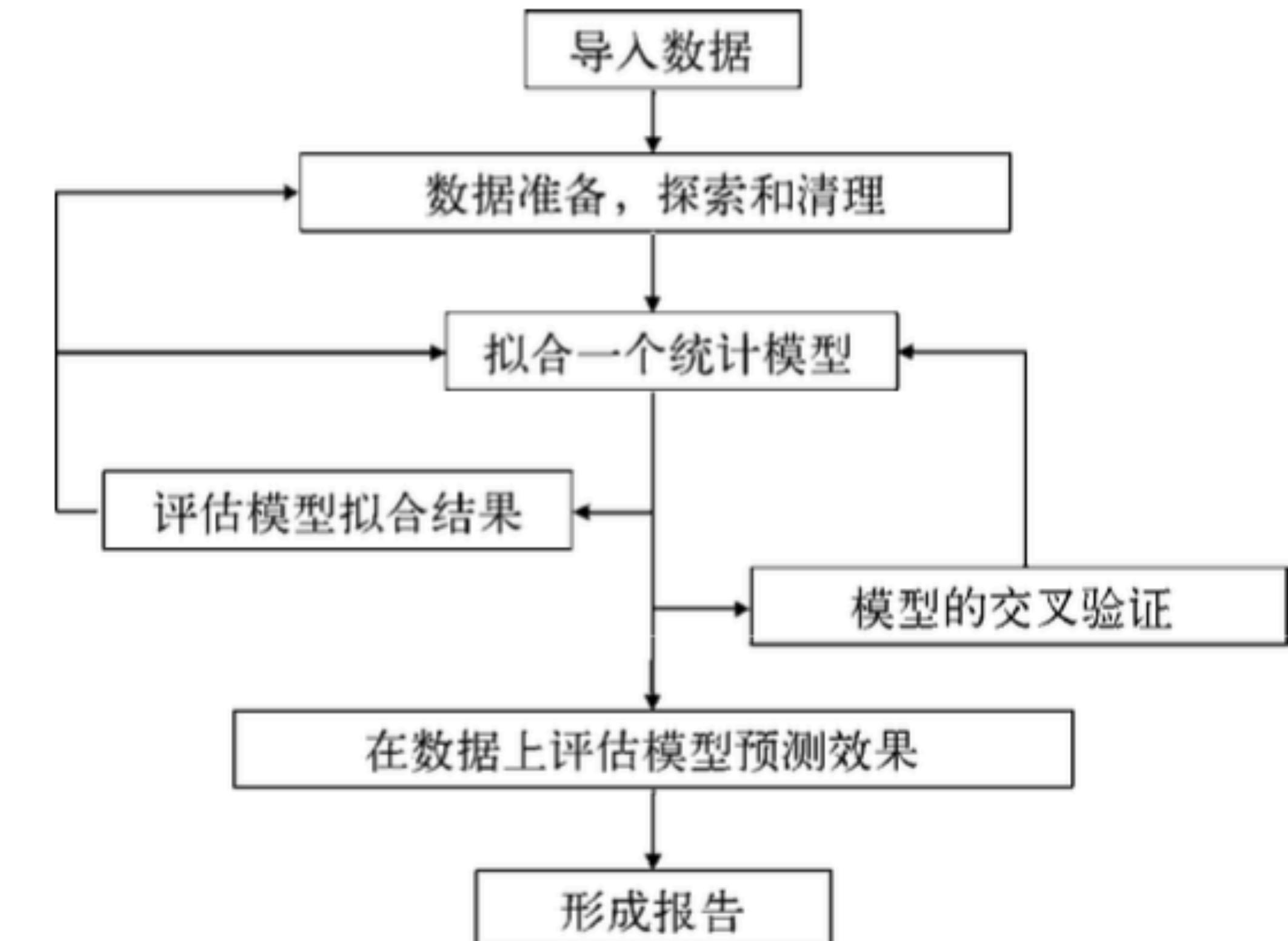
R

- ▶ R is a language and environment for statistical computing and graphics.
- ▶ <https://www.r-project.org/>
- ▶ 免费
- ▶ 开源，上千个包 (package) 可供使用
- ▶ 强大的统计分析能力、优美的图形输出



R可以做什么？

- ▶ 从数据科学家的角度，处理、分析和理解数据，包括：
 - ▶ 获取数据(从各种数据源将数据导入程序)
 - ▶ 整理数据(编码缺失值、修复或删除错误数据、将变量转换成更方便的格式)
 - ▶ 注释数据(以记住每段数据的含义)
 - ▶ 总结数据(通过描述性统计量了解数据的概况)
 - ▶ 数据可视化
 - ▶ 数据建模(解释数据间的关系，检验假设)
 - ▶ 整理结果(创建具有出版水平的表格和图形)



RStudio

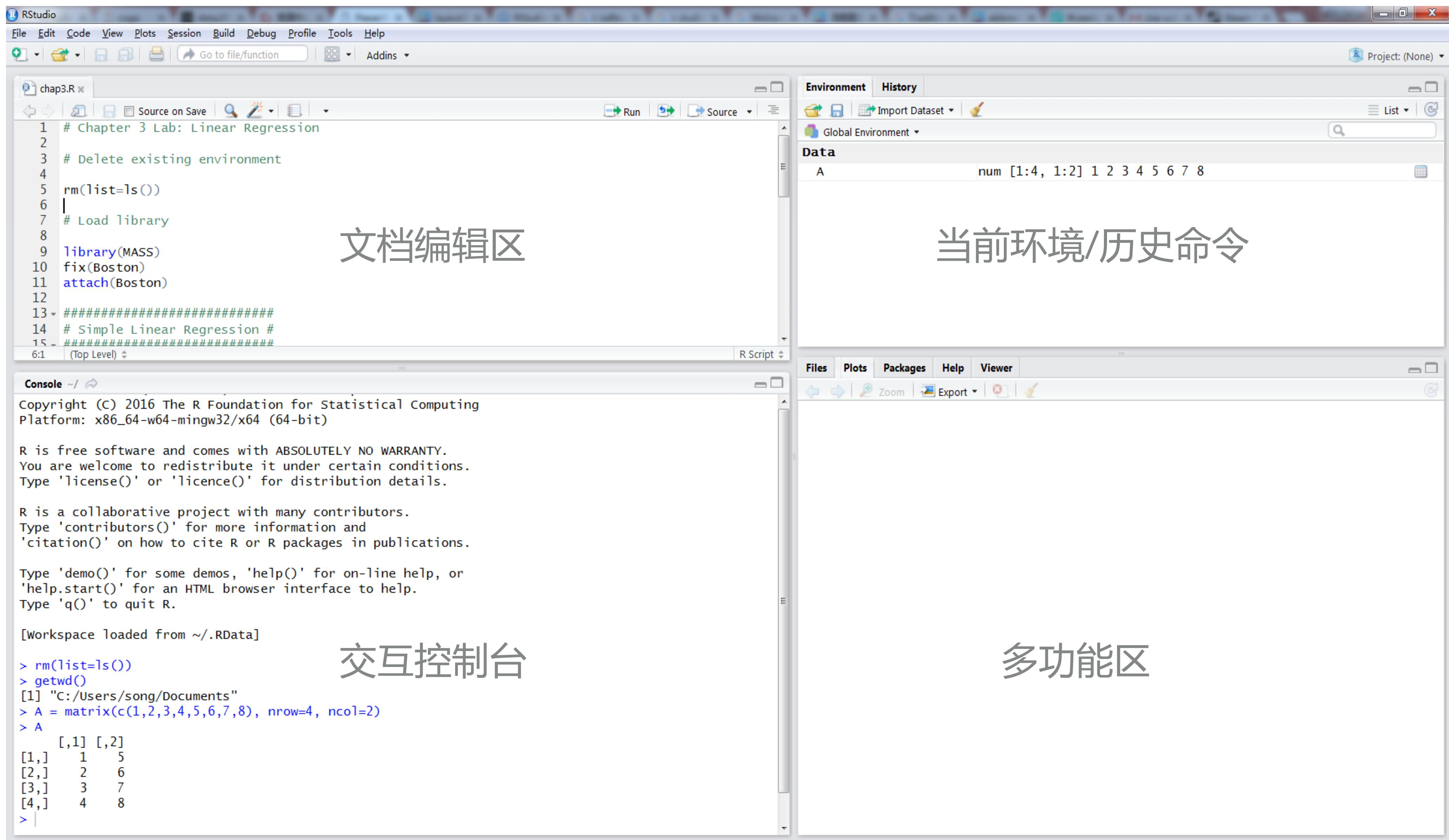
- ▶ RStudio provides open source and enterprise-ready professional software for data science.
- ▶ <https://www.rstudio.com/>
- ▶ IDE (integrated development environment) , 2011年发布



安装

- ▶ 从官网下载R
- ▶ 运行安装文件进行安装
- ▶ 从官网下载RStudio
- ▶ 运行安装文件进行安装
- ▶ 运行RStudio

RStudio界面

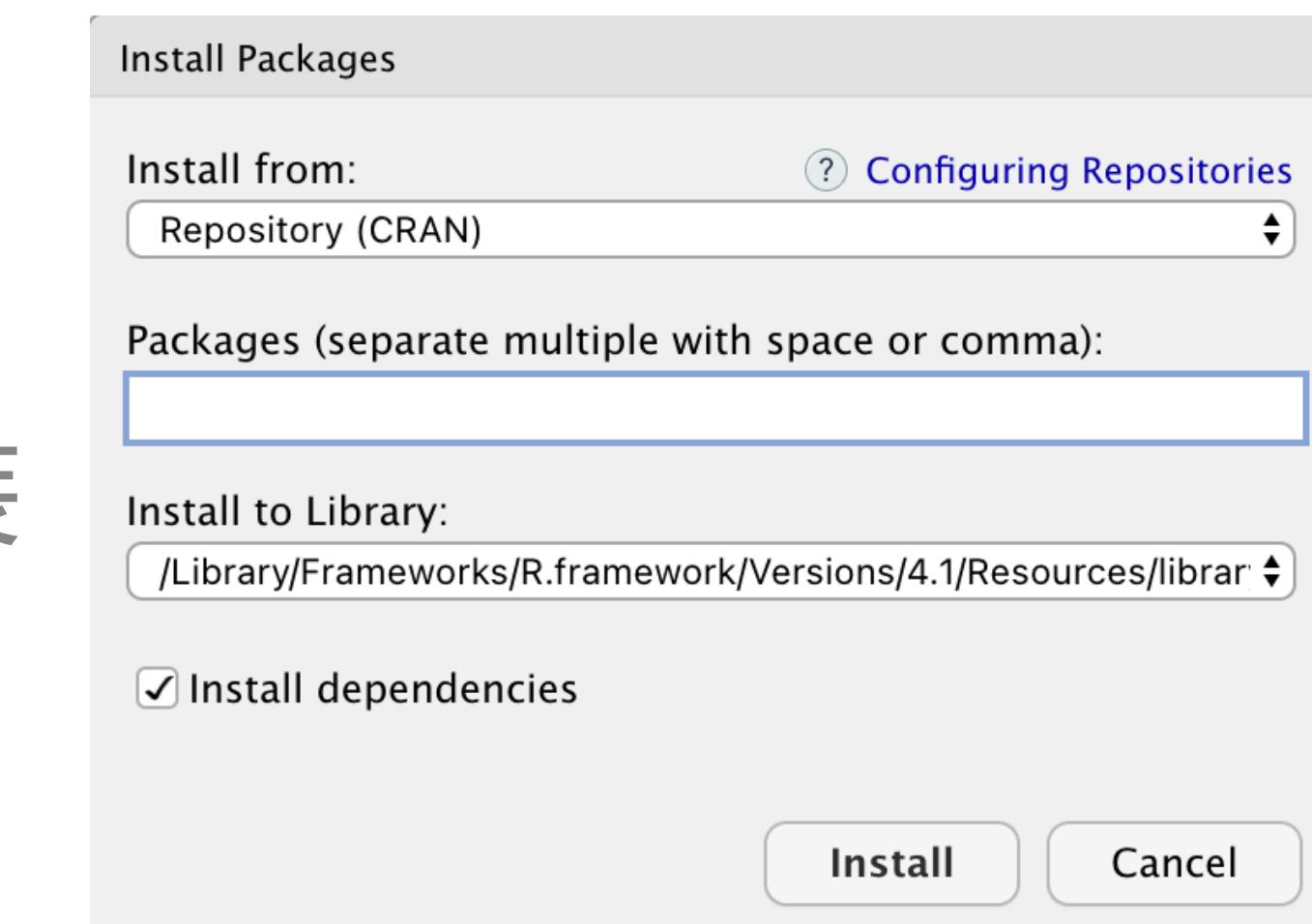


函数 (Function) 与对象

- R 中的多数功能是由程序内置函数、用户自编函数和对对象的创建和操作实现的。
- 对象可以是任何东西：数据、函数、图形、分析结果等。
- 一次交互式会话期间的所有数据对象都被保存在内存中。一些基本函数默认直接可用，更多高级函数则包含于需要预先加载的程序包中。
- 要运行一个函数名为 `func` 的函数，我们需要输入 `func(input1, input2, ..., ...)`，其中 `input1` 和 `input2` 是函数的输入参数。
- 例1：`x = c(1,2,3,4)`，聚合1,2,3,4成一向量，并保存到对象x。
- 例2：`x <- rnorm(5)`，产生5个标准正态分布随机数，并将这一向量保存到对象x。
- =与<-都可使用，<-是标准用法。

包 (Package)

- ▶ R 包是由社区开发的函数和数据集的集合。它们通过改进现有的基本 R 功能或添加新的复杂功能来提升 R 的效率。
- ▶ 包的安装 : `install.packages("package")`
- ▶ 在安装一个包的时候，这个包的附属包也会自动安装
- ▶ 包的加载 : `library(package)`
- ▶ 未加载状态下调用包中函数 : `package::function()`



System Library				
Name	Description	Version		
<input type="checkbox"/> backports	Reimplementations of Functions Introduced Since R-3.0.0	1.3.0		
<input checked="" type="checkbox"/> base	The R Base Package	4.1.2		
<input type="checkbox"/> base64enc	Tools for base64 encoding	0.1-3		
<input type="checkbox"/> boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-28		
<input type="checkbox"/> bridgedist	An Implementation of the Bridge Distribution with Logit-Link as in Wang and Louis (2003)	0.1.0		
<input type="checkbox"/> bslib	Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'	0.3.1		
<input type="checkbox"/> caret	Classification and Regression Training	6.0-90		
<input type="checkbox"/> checkmate	Fast and Versatile Argument Checks	2.0.0		
<input type="checkbox"/> class	Functions for Classification	7.3-19		
<input type="checkbox"/> cli	Helpers for Developing Command Line Interfaces	3.1.0		
<input type="checkbox"/> cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.2		
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-18		
<input type="checkbox"/> colorspace	A Toolbox for Manipulating and	2.0-2		

R Markdown

- ▶ R Markdown 是 R 语言环境中提供的 markdown 编辑工具
- ▶ markdown是一种轻量级标记语言，它允许使用者可以通过简单的文本格式编写文档，并且转换成html文档
- ▶ 对于专注于用 R 语言写报告的数据分析师来说，R Markdown既提高了数据分析工作的便捷性，也提高了数据分析报告的复用性
- ▶ `install.packages("rmarkdown")`

R Markdown三大要素

- ▶ Metadata (YAML) : R Markdown的头部文件，定义了R Markdown的性质，比如title、author、date、指定output文件类型等。
- ▶ Text (Markdown)
- ▶ Code
 - ▶ 代码块 (code chunk):
 - ▶ ``{r}
 - ▶ yourcode
 - ▶ ...
 - ▶ 文字中代码: abcdefg `r yourcode`
- ▶ 开始实战！

The screenshot shows the RStudio Source Editor interface with an R Markdown file named "test.Rmd". The file contains the following content:

```
---  
title: "Hello R Markdown"  
author: "Awesome Me"  
date: "2018-02-14"  
output: html_document  
---  
This is a paragraph in an R Markdown document.  
Below is a code chunk:  
```{r}  
fit = lm(dist ~ speed, data = cars)
b = coef(fit)
plot(cars)
abline(fit)
```  
The slope of the regression is `r b[1]`.
```

The code blocks are highlighted with blue boxes. The rendered output "The slope of the regression is `r b[1]`." is also highlighted with a blue box at the bottom right.