

R 语言概述

周世祥

2020 年 5 月

初始 R 语言

R 语言是一款在开源世界里集万千宠爱于一身的软件，你能想到的地方都有它的身影！

看看权威的 TIOBE 开发语言排行榜，<http://www.tiobe.com>，可以看到 R 的热度。

** 2020 年 5 月 TIOBE 指数 ** 五月标题：编程语言 C 再次排名第一

Java 和 C 在 4 月已经非常接近，但是本月 C 再次超越 Java。C 上一次成为第一名是在 2015 年。我们只能猜测 C 为什么再次成为第一名。原因之一可能是冠状病毒。这听起来很愚蠢，但是某些编程语言确实可以从这种情况中受益。例子是数据科学领域的 **Python 和 R**，因为每个人都在寻找该病毒的解毒剂。但是嵌入式软件语言（例如 C 和 C++）也越来越流行，因为它们被用于医疗设备的软件中。另一方面，值得一提的是，Rust 实际上已经接近前 20 名（一个月内从第 27 位上升到第 21 位）。

R 目前排第十位。

TIOBE 编程社区索引是编程语言受欢迎程度的指标。索引每月更新一次。评级基于全球熟练的工程师，课程和第三方供应商的数量。诸如 Google, Bing, Yahoo!, Wikipedia, Amazon, YouTube 和 Baidu 等流行的搜索引擎用于计算评分。重要的是要注意，TIOBE 索引与最佳编程语言或大多数代码行所用的语言无关。

R 语言是什么

R 是一个强大的统计分析功能及作图功能的软件系统。S 语言是 AT&T 贝尔实验室开发的一种商业统计软件，1992 年 Auckland 大学为了更有效地开展统计教学，决定开发一套类似的统计软件，就是 R 的前身。

为什么要用 R 语言

物美价廉：作图颜值高且完全免费。ggplot 包，影响了 Python 的可视化包风格。

可以把北京出租车轨迹画出地图。

动态可视化，静态可视化都可以做。

兼收并蓄：算法覆盖广，软件扩展易。从经典的回归分析，机器学习分类聚类算法，到深度学习算法都可直接找到调用的函数。

方便从 Oracle, mysql, 统计表格中导入导出数据。

甚至可以和各种语言互调。

R 语言具有强大的社区支持，R 包的开发者通常先将代码放入 GitHub，接受世界各地使用者提出问题，修改代码，等成熟后再放入 CRAN(Comprehensive R Archive Network 简写) 上发布。Stack Overflow 是一个优质的技术问答网站。国内活跃的 R 社区有统计之都旗下的论坛 COS:<https://d.cosx.org/t/r>；狗熊会：http://www.xiong99.com.cn/article_category.php?id=8；李东风老师 R 语言电子教程：http://www.math.pku.edu.cn/teachers/lidf/docs/Rbook/html/_Rbook/index.html；R 结巴分词：<http://qinwenfeng.com/jiebaR/.tinytex> 文档帮助：<https://yihui.org/tinytex/cn/>；暨南大学王斌会老师的博客<http://rstat.leanote.com>，有很多 R 语言，Python 语言资料；出版书籍参考：<https://github.com/rstudio/bookdown> 动态可视化参考：<https://plotly.com/r/reference/>。

R 语言的安装

<https://www.r-project.org>点击下载 R，选择合适的镜像，China 下离你最近的一个即可。

网站上提供两类 Windows 下的 R 安装文件，base 和 contirb，后者是一个包含所有扩展包的二进制安装文件，前者仅包含基本功能，一般我们安装阶段选 base 即可。

Rstudio 是 R 语言的集成开发环境 IDE，代码编辑，编译，调试等工具集成。

选中某行，点击 run，便可运行，或者 ctrl+enter 键。写代码具有智能提示功能，自动补全，比如 mean。

R 包的安装与获取

R 包就是一个把 R 函数，数据，预编译代码以一种定义完善的格式组织在一起的集合。R 在安装时会自带一系列默认包，如 base，datasets，stats，methods，graphics 等，提供了很多功能丰富的函数和数据，命令 search() 可以告诉你工作空间里有哪些包可以直接使用。

```
search()
```

```
## [1] ".GlobalEnv"      "package:stats"    "package:graphics"
## [4] "package:grDevices" "package:utils"    "package:datasets"
## [7] "package:methods"  "Autoloads"        "package:base"
```

如何安装其他扩展包？

一种方式是通过 CRAN 服务器安装，另一种是从 Github 上安装。

第一种更常见：方法是直接通过命令安装：install.packages(“包名”)；另一种方法是 Rstudio 中通过图形化方式安装（注意从 global 选项中更换镜像）；

Github 上包并不需要从网站下载再安装，R 里有一些方便与 github 交互的包如 devtools，其中的 install_github() 函数就专为从 github 安装包设计。

```
library(devtools)
install_github("gaborcsardi/praise")
```

注意要写清作者名和包名信息。

更新包：update.packages()。

安装完的包就可以直接使用了。

如何用好一个包，搜索，<http://cran.r-project.org/web/views/>，可以寻找跟你研究相近的主题。

在各种可用的包中如何选择？Crantastic 网站：<http://www.Crantastic.org/popcon>中有各种包的热度排行。

可以从 rstudio 右下角界面搜索框中看帮助信息，也可以 help(package=“包名”)，打开帮助指南。

函数的帮助文档很长，从哪里看起，两个方面切入，一是先读 Description 和 Usage，整体了解基本功能及参数设置，二是帮助文档最后附带的 examples。

命令 demo(“函数名”) 查看已经加装到内存的包里有例子的函数列表。

Quick-R 网站，<http://www.statmethods.net/index.html>，内容丰富，极易上手。

出现问题，bug 用搜索引擎，建议英文的用 google。

github 中也可以搜索资源，“R Exploratory Data Analysis”。

参考文献

朱雪宁，《R 语言从数据思维到数据实战》，中国人民大学出版社，2019