



数据分析与处理技术——1 基础知识

商学院 徐宁

基础知识

1.绪论

数据科学应用场景

数据分析工具和能力

商业中的数据应用场景



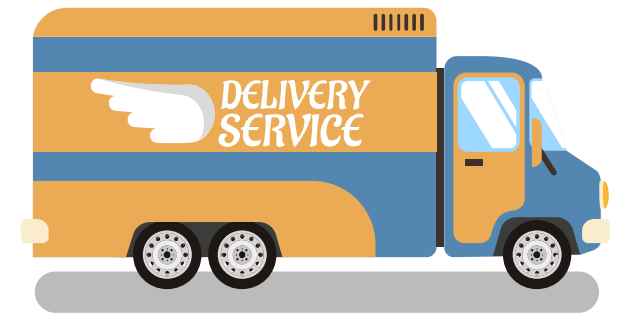
前端

用户画像
商品推荐
广告推荐
风险管理
客户管理



中端

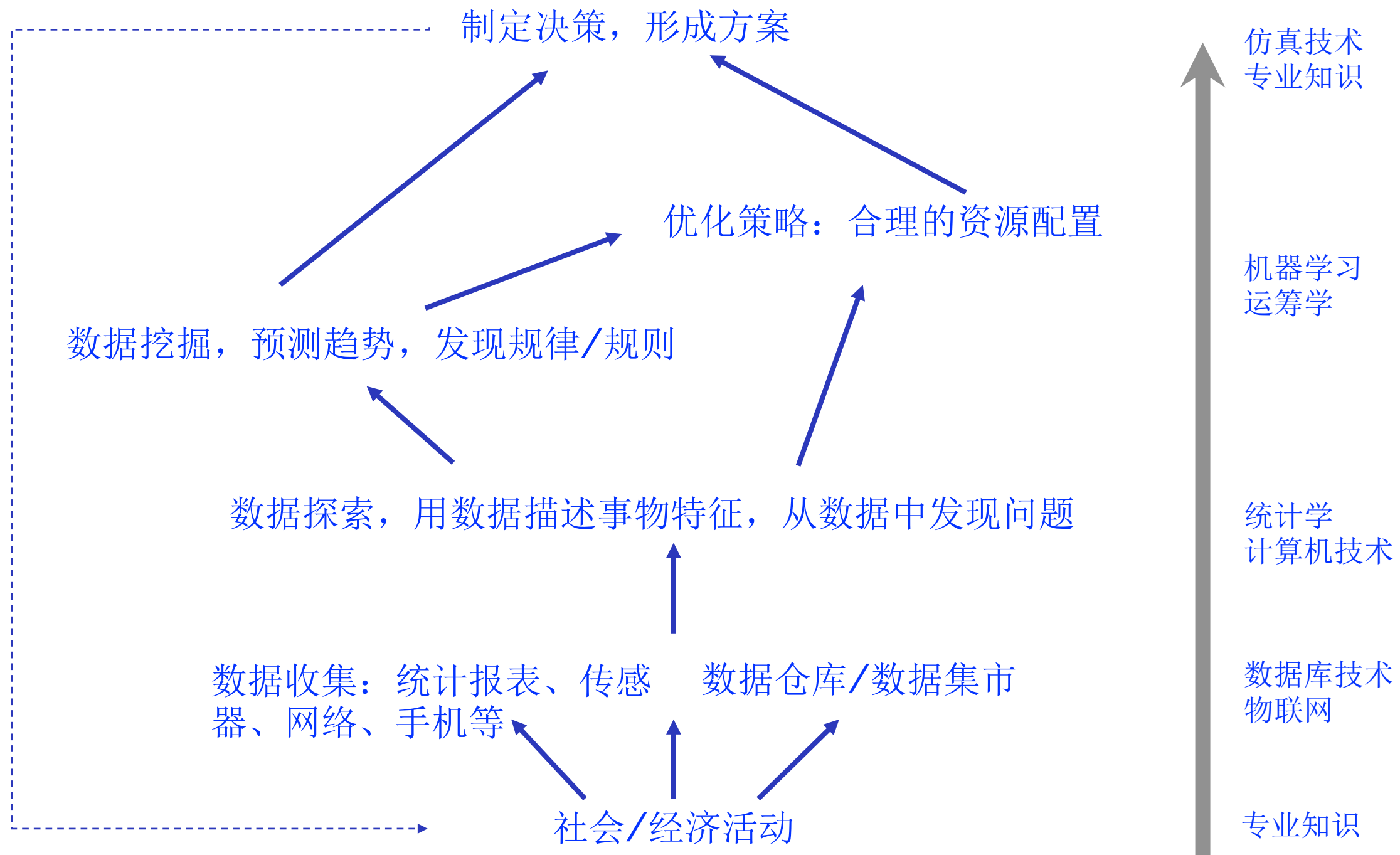
销售预测
库存管理
定价促销
履约优化
供应商管理



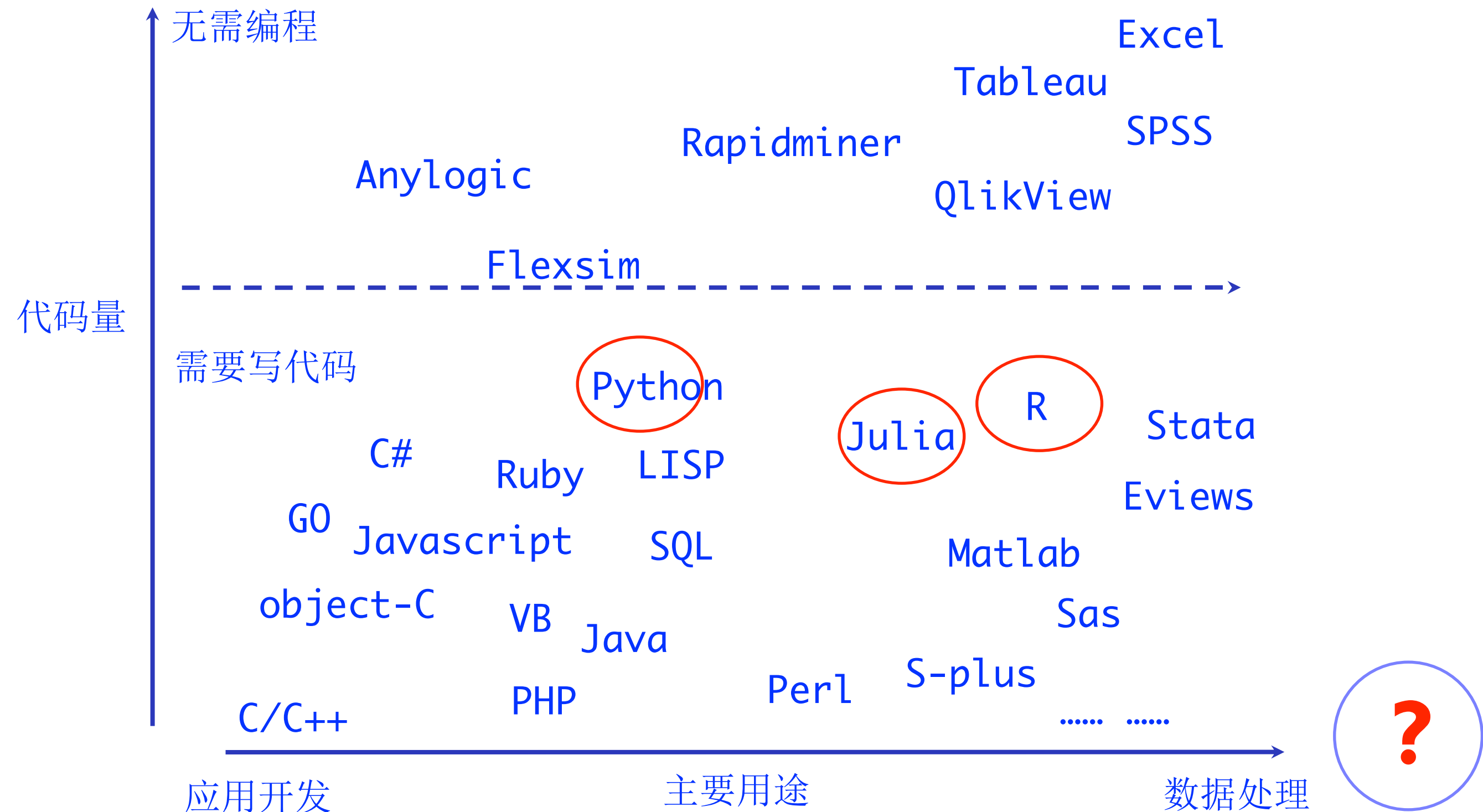
后端

网络优化
末端配送
仓库管理
逆向物流

数据分析全流程的层次结构



数据处理所用工具



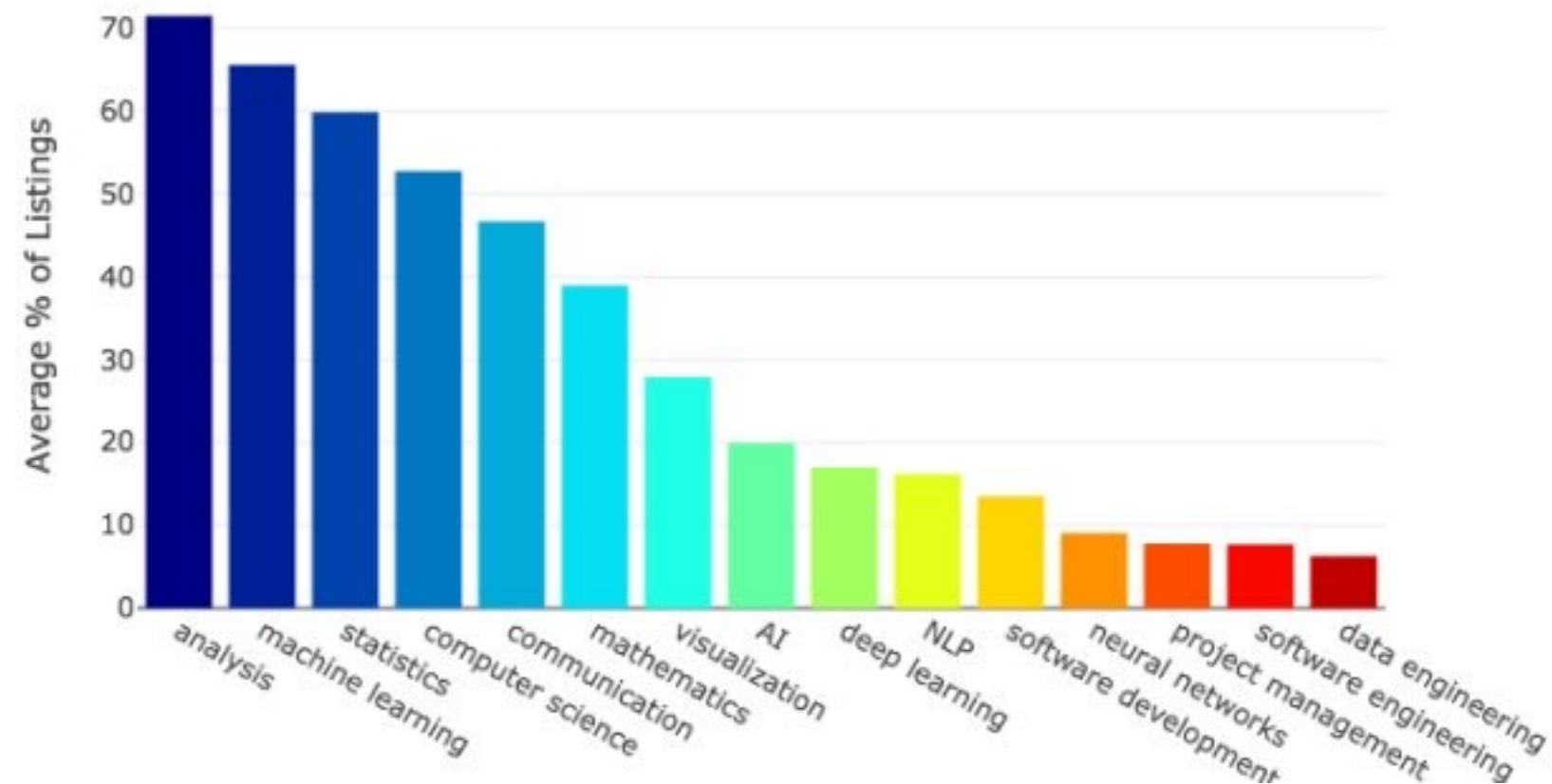
数据科学需要哪些能力

The skills Data Scientists need today

(based on 300 job listings from tech companies in June 2019)



General Skills in Data Scientist Job Listings



<https://www.kdnuggets.com/2019/08/new-poll-data-science-skills.html>

基础知识

2.环境配置

R语言的环境配置

界面操作基础

R的安装

安装R软件:

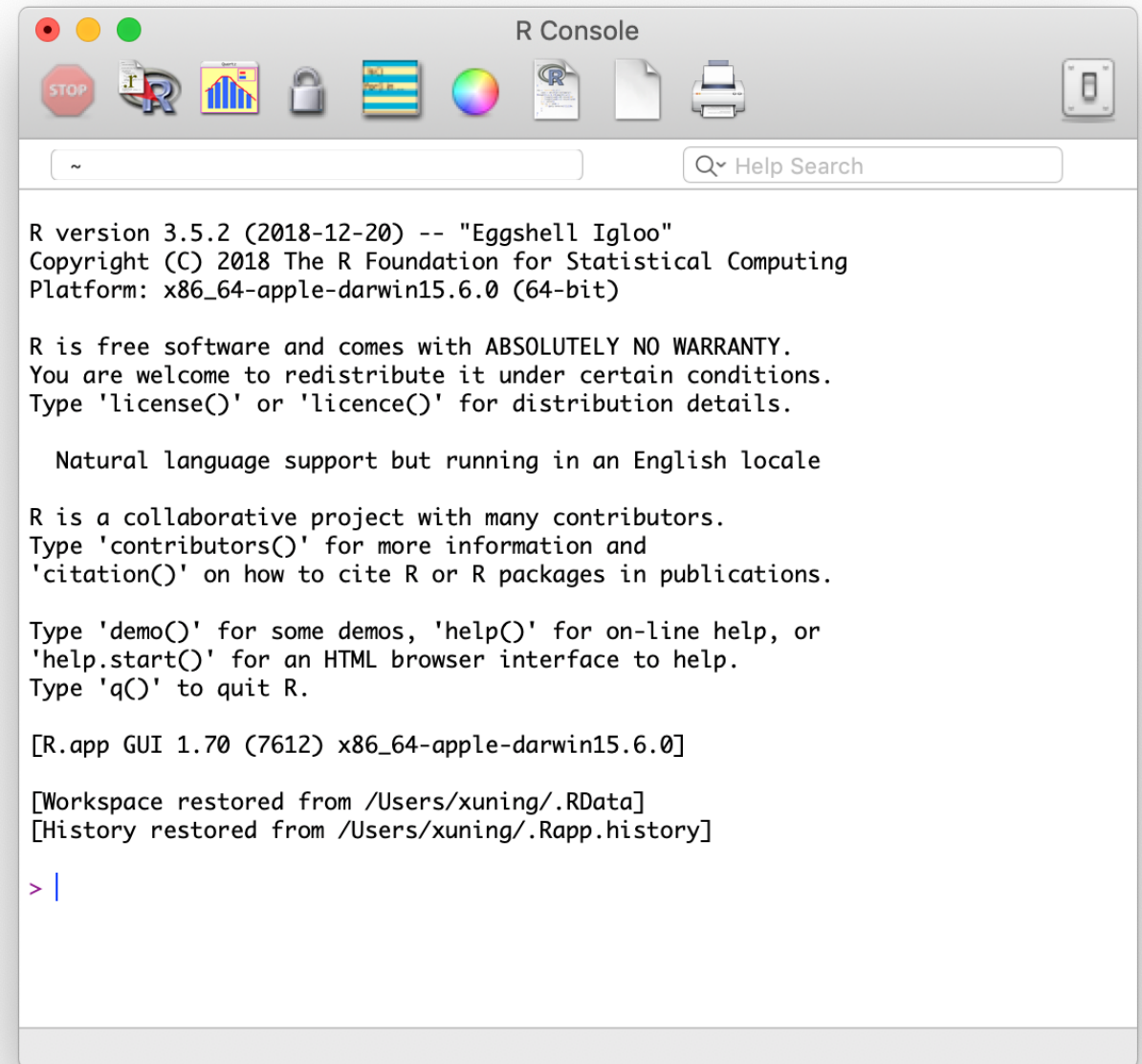
- CRAN版本

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

- Microsoft R

R自带一个简单的代码编辑环境，如右图。

CRAN中提供的Rtools: 使用R语言开发package时必须的编译工具



环境配置

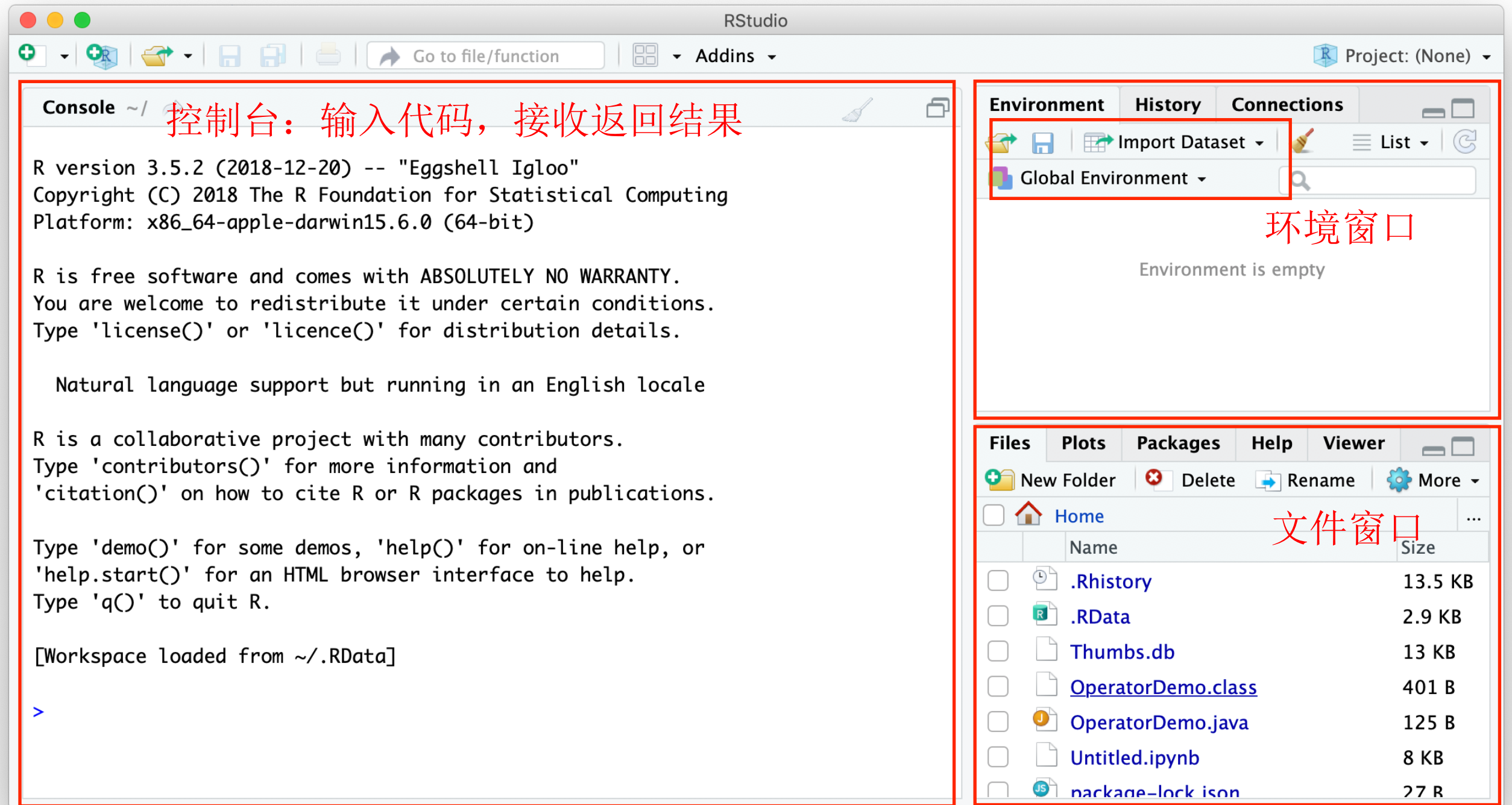
集成开发环境(Integrated Development Environment)

- Rstudio, Eclipse+R工具集, VS code, Atom

远程IDE:

- Kaggle的Notebook: <https://www.kaggle.com/notebooks>
- 阿里的天池实验室: <https://tianchi.aliyun.com/notebook-ai/>
- Rstudio的云环境: <https://rstudio.cloud/projects>

Rstudio功能区

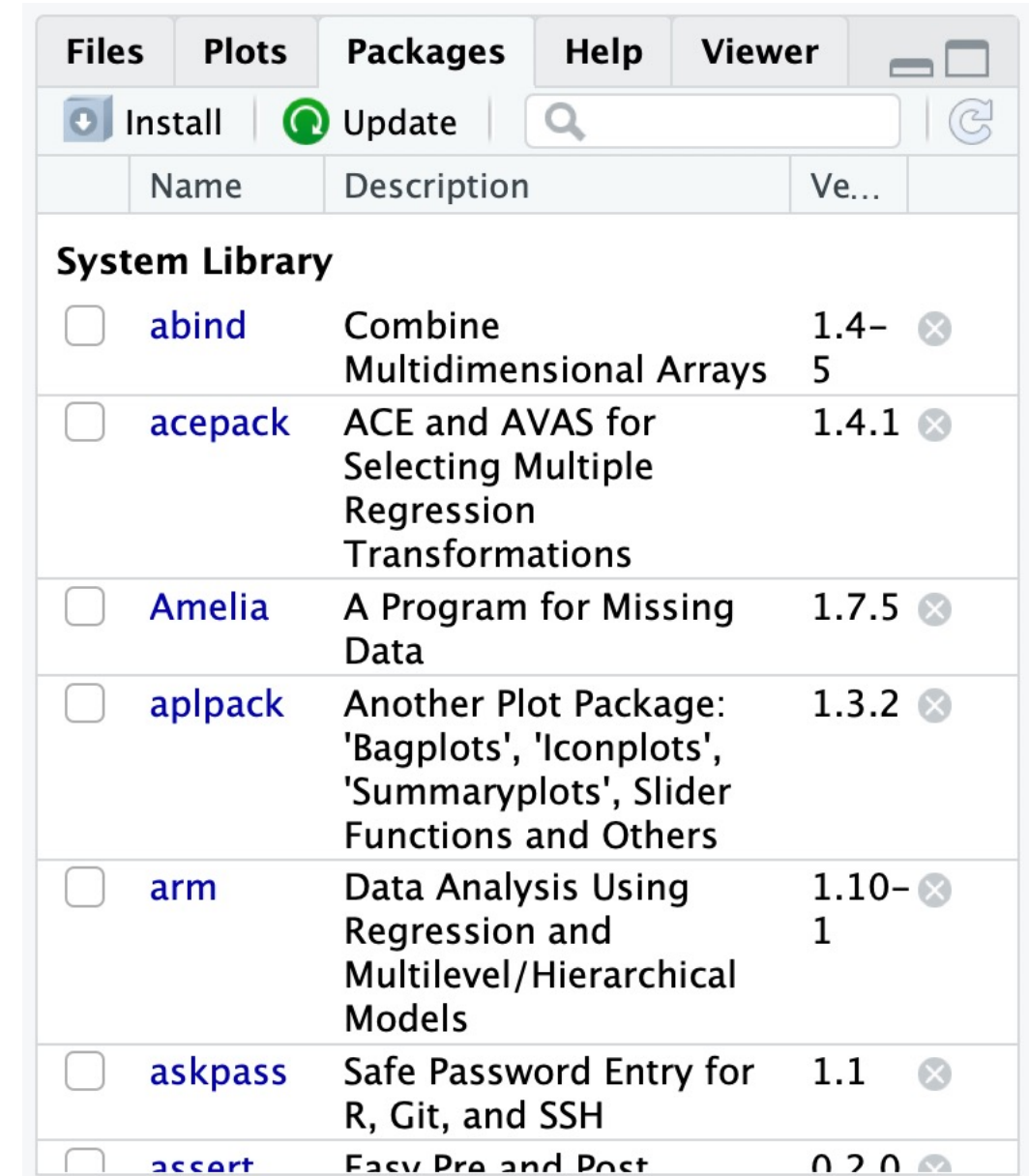
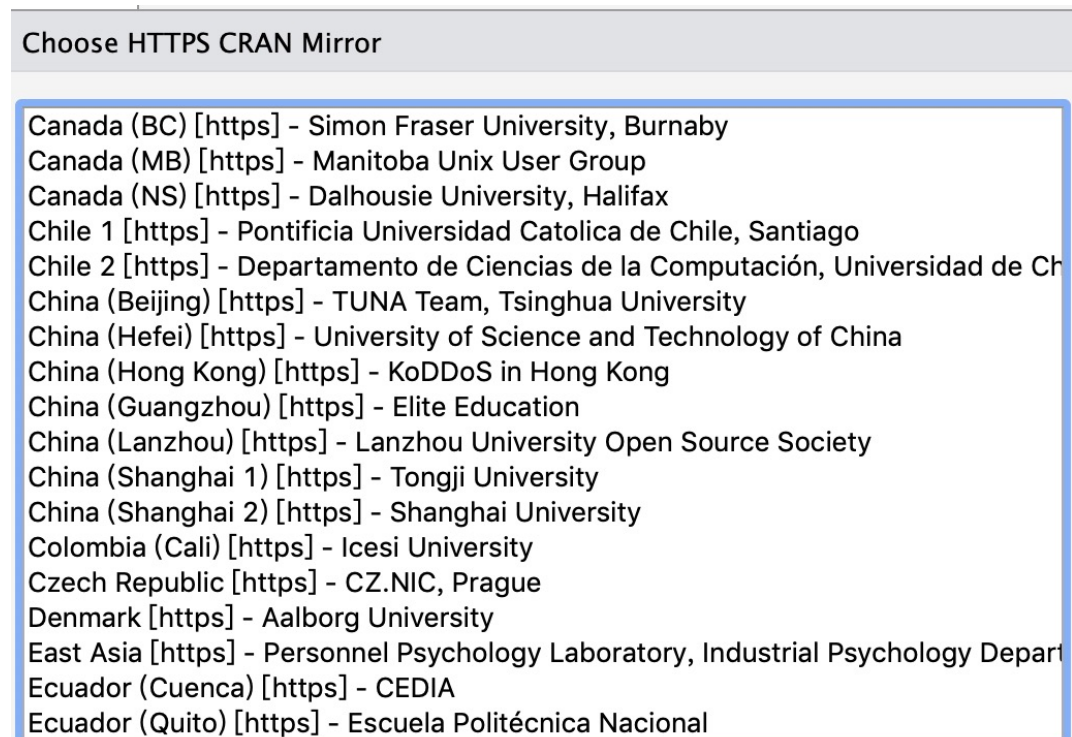


工具包配置

通过镜像站点在线加载package

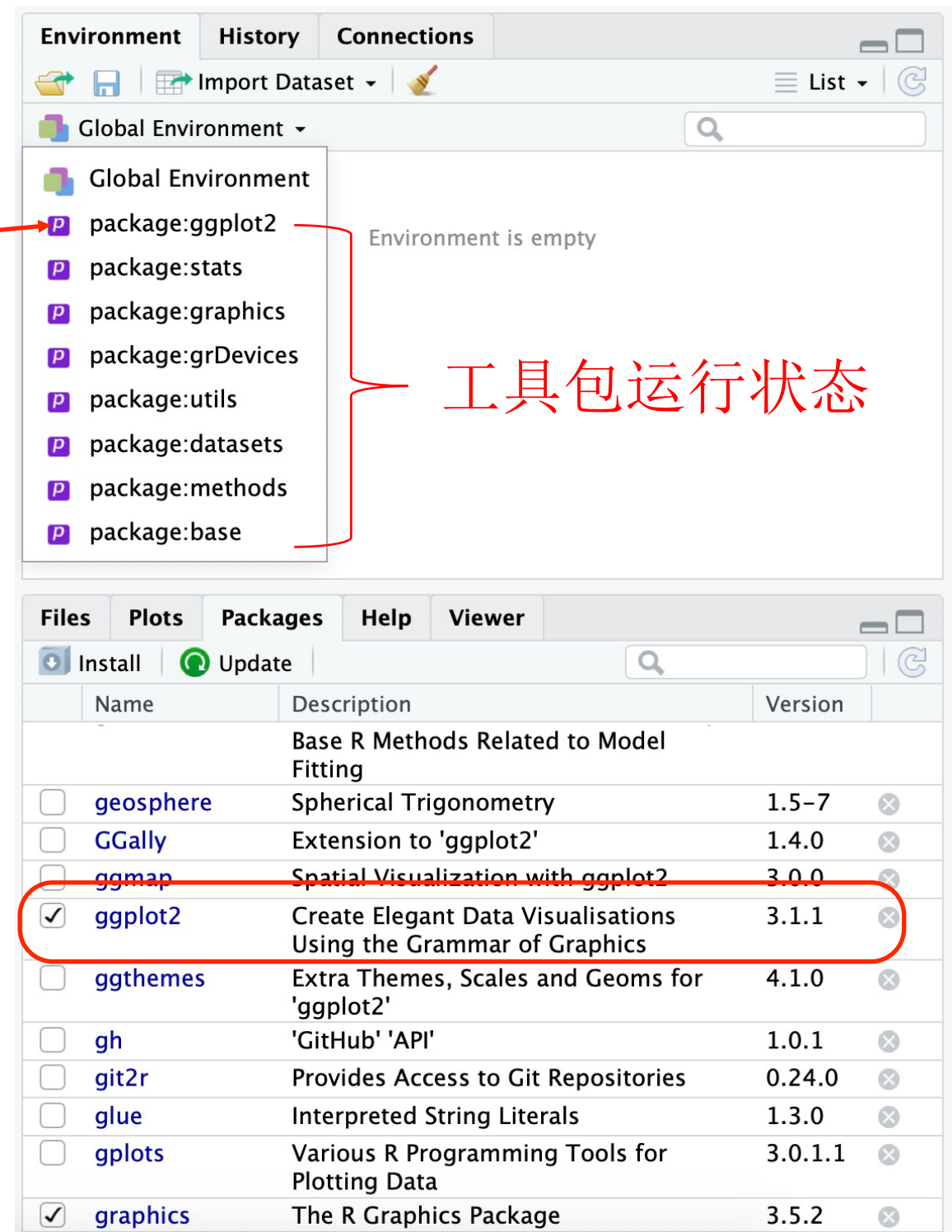
选择CRAN的连接镜像:

Tools->Global Options->Packages->
CRAN mirrors



加载工具包

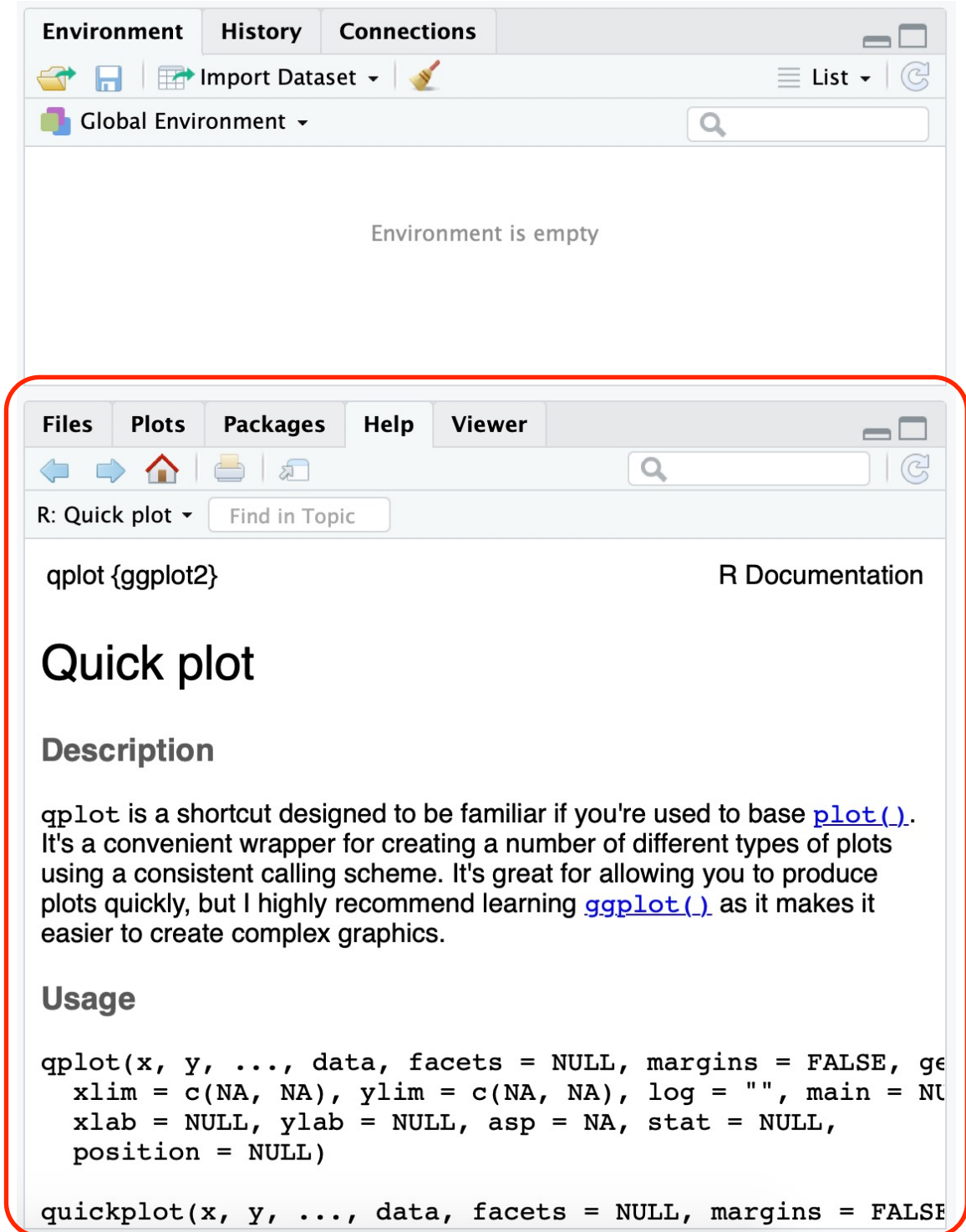
- 加载工具包方法
 - > `install.packages("ggplot2")`
 - > `library(ggplot2)`
- 工具包(package)实质是一系列函数和数据构成的集合



查询帮助

- 查询帮助
- ```
> ?qplot
```

发布在网上的工具包需要经过学术审核，这保证了任何一个CRAN发布的包都有完备的帮助文档。



基础知识

---

## 3.数值运算

数值运算符

数值函数

参考书目

# 数值运算符

- 常用数学运算符:

`+` `-` `*` `/` `^` `%%` `%/%`

- 常用数学函数运算:

`sin()` `cos()` `exp()`  
`log()` `log10()`

- 比较和判断也是运算

```
> 10%%3
[1] 1
> exp(10)
[1] 22026.47
> log(exp(1))
[1] 1
> log10(10)
[1] 1

> 1>=2
[1] FALSE
```

# 赋值运算

- 赋值运算符 = 或 <- ->
- 命名规则
  - 中间无空格
  - 首字符非数字
- 命名规范
  - 变量用名词，函数用动词
  - 下划线取代空格，避免驼峰式命名
  - 用 <- 而非 =

```
> a <- 1+2
```

```
> a
[1] 3
```

```
> a = 1+2
```

```
> a
[1] 3
```

```
> 1+2 -> a
```

```
> a
[1] 3
```



# 常用运算符

| 算术符号 |       | 关系符号 |      | 逻辑符号 |     |
|------|-------|------|------|------|-----|
| 符号   | 含义    | 符号   | 含义   | 符号   | 含义  |
| +    | 加     | >    | 大于   | &    | 与   |
| -    | 减     | <    | 小于   |      | 或   |
| *    | 乘     | >=   | 大于等于 | !    | 非   |
| /    | 除     | <=   | 小于等于 | &&   | 与运算 |
| %%   | 求余数   | ==   | 等于   |      | 或运算 |
| %/%  | 求模或整除 | !=   | 不等于  |      |     |
| ^    | 指数幂   |      |      |      |     |
|      |       |      |      |      |     |
|      |       |      |      |      |     |

# 常见数值函数

| 函数                 | 功能描述                      |
|--------------------|---------------------------|
| abs(x)             | 绝对值                       |
| sqrt(x)            | 开方，等效于 $x^{0.5}$          |
| ceiling(x)         | 不小于x的最小整数，ceiling(3.4)得到4 |
| floor(x)           | 不大于x的最大整数                 |
| trunc(x)           | 截取x的整数部分                  |
| round(x,digits=n)  | 保留n位小数的四舍五入               |
| signif(x,digits=n) | 保留n位有效数字的四舍五入             |
| sin(x) tan(x)      | 正弦 正切                     |
| asin(x) atan(x)    | 反正弦 反正切                   |
| log(x,base=n)      | 以n为底取对数                   |
| log(x) log10(x)    | 取自然对数，取以10为底的对数           |
| exp(x)             | 指数函数                      |

# 参考书目

《R数据科学》Hadley Wichham, 人民邮电出版社

《R语言入门经典》安迪.尼克拉斯, 人民邮电出版社

《R语言 实用数据分析和可视化技术》Jared P. Lander, 曾益强译, 机械工业出版社

《R语言编程艺术》Norman Matloff, 机械工业出版社

《R语言编程指南》任坤, 人民邮电出版社

《R语言实战》Robert I kabacoff, 机械工业出版社

《高级R语言编程指南》Hadley Wichham, 机械工业出版社

此次以外, 涉及各专题的优秀教材将在随后章节中作为参考书推荐

## 相关电子资料

---

- 工具包搜索: <https://www.rdocumentation.org/>
- 数据科学资讯: <https://www.kdnuggets.com/>
- R的电子教程平台: <https://bookdown.org/>

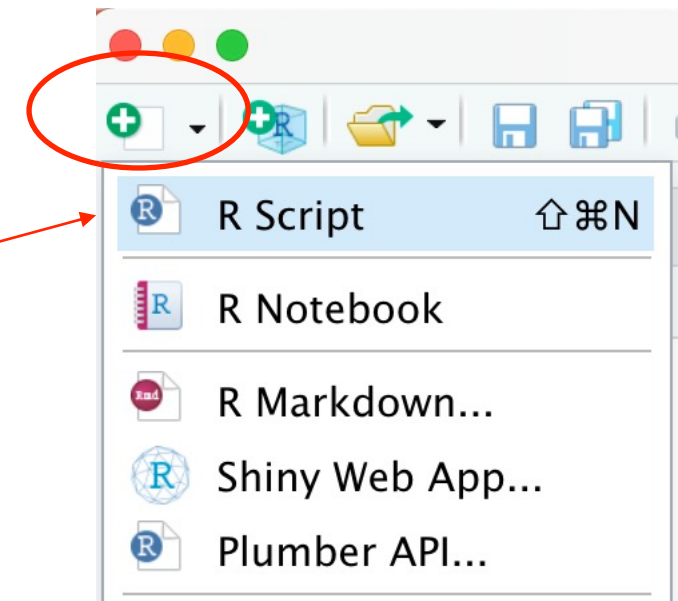
# 练习

1. 在console中按`alt+shift+k`,看看会是什么结果.
2. 依次尝试所有运算符和数学函数.
  - 计算 $\pi^{e+1}, e^\pi$
  - $\pi^e - e^\pi$
  - 计算 $2.3^8 + \log(7.5) - \cos(\pi/\sqrt{2})$
  - 计算 $3^5 > 5^3$
  - 不等式 $x^5 + e^{\frac{y}{2}} - 3xy - 213 > 0$ ,  $\begin{cases} x = 3 \\ y = 7 \end{cases}$ 和 $\begin{cases} x = 9 \\ y = 2 \end{cases}$ 是否是可行解。
  - 关于方程 $x^3 - 13y^3 - xy^2 = 27$ , 判断 $\begin{cases} x = 15 \\ y = 6 \end{cases}$ 是否是其解。

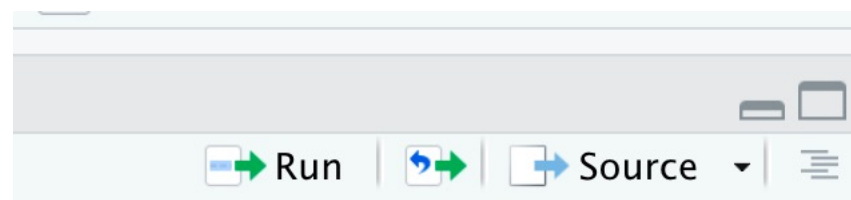
## 练习：脚本运行代码

创建脚本文件`compute.r`，将上页问题的代码写入文件中。

通过下拉菜单创建脚本文件(R Script)



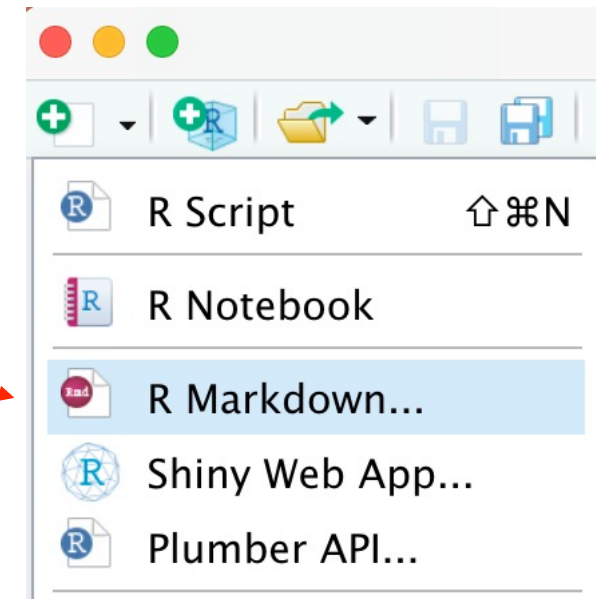
分别点击 `run` `source`看运行方式有什么区别？



# 练习：Rmarkdown运行环境

- 建立Rmarkdown文件，这是一种代码和文件混合在一起的方式。

第一次创建R Markdown文件  
会自动补充一系列支撑工具包



夹在上下````符号之间灰色字符被  
当作代码块，运行键在右侧

尝试在代码块中实现计算。

code chunks within the document. You can embed an R code chunk like this:

```
```${r}  
summary(cars)  
```
```

## Including Plots

