

# R 语言简明教程

## 第一课、R 语言及其编程环境

张金龙

jinlongzhang01@gmail.com

2022-01-10

# 《R 语言简明教程》内容安排 (I)

- 第一课、R 语言及其编程环境
- 第二课、数据类型、结构和基础操作
- 第三课、用 graphics 包绘图
- 第四课、用 ggplot2 包绘图
- 第五课、tidydata 以及 tidyverse 系列程序包的使用
- 第六课、基础统计方法的实现
- 第七课、编写 R 函数
- 第八课、Rmarkdown 以及可重复性数据分析
- 第九课、git 与版本控制

# 《R 语言简明教程》内容安排 (II)

- 第十课、R 程序包的编写、检查与发表
- 第十一课、生物多样性分析常用程序包
- 第十二课、空间数据处理和地图绘制
- 第十三课、系统发育研究的常用程序包
- 第十四课、课程总结

# 内容提要

① 什么是 R?

② 编程环境

③ 熟悉 R 和 Rstudio

④ R 脚本 R script

⑤ 程序包

⑥ 怎么学习 R?

⑦ 讲了什么

1

## 什么是 R?

# 什么是 R?

- R 是数据处理、统计和绘图的软件，可以在 Windows、MacOS、Linux、UNIX 等多种操作系统运行
- R 既是软件，也是运行在该软件上的计算机语言

# 哪个软件跟 R 关系最近?

- MS Excel: 数据输入与分析
- SPSS: 统计、绘图
- SAS: 统计
- Matlab: 数学建模、仿真、绘图
- Origin: 绘图
- SigmaPlot: 绘图、基础统计

# 哪个语言跟 R 的关系最近?

- Perl: 脚本语言, 生物信息学分析
- Python: 脚本语言, 大数据、人工智能
- C: 编译语言, 编写应用程序
- C++: 编译语言, 编写应用程序
- Fortran: 编译语言, 高性能计算
- S: R 是 S 语言的一种“方言”(借鉴了 Scheme 的语法)

# R 语言简史

- 上世纪 70 年代，美国贝尔实验室的 John Chambers、Rick Becker、Trevor Hastie、Allan Wilks 等人研发了用于统计分析的 S 语言
- 1988 年，美国一间商业公司根据 S 语言的源代码开发了 Splus 商业软件。Splus 能更好地执行 S 语言，可同时进行统计分析和绘图，大受欢迎
- 90 年代初，新西兰奥克兰大学的 Robert Gentleman 和 Ross Ihaka 开发了一个软件，用于 S 语言的教学
- 1995 年，两人将该软件开源，并命名为 R
- 1997 年，R 由核心团队管理，同年，R 的镜像网站 CRAN 上线
- 2000 年，R1.0.0 发布

目前版本 R 4.1.2 (Bird Hippie)

# 近年来的重要进展

- 2007 年，绘图程序包 ggplot2 上线
- 2011 年，集成开发环境 Rstudio 发布
- 2014 年，rmarkdown 包发布
- 2016 年，tidyverse 系列程序包发布
- ..... .....

# 生物多样性研究

- 系统分类：描述、命名、分类修订
- 编目、评估、保护：如 IUCN 红色名录评估
- 标本数字化、标本馆管理：数据的清洁和标准化
- 物种潜在分布区预测与气候变化
- 群落生态学：多样性指数、空间分布格局、种面积曲线
- 群体遗传学：推断种群历史以及计算多种指标
- DNA 条形码、基因组、转录组分析：鉴定、推断系统发育关系
- 生成报告以及信息发布：论文撰写、微信公众号、博客，跟同行和公众交流

# 生物多样性研究的一些新变化

- 数据量巨大 (**big data**): 动辄几万、几十万条数据, 难以手动处理
- 复杂的工作流程 (**working flow**): 越来越依靠 R、Python、Julia、SQL 等计算机语言, 代码本身包括了工作流
- 学术界对研究可重复性 (**Reproducibility**) 的要求越来越高: 研究报告要提供数据收集的方法以及各种细节, 保证同行在看过材料与方法之后能将研究重复出来。不仅仅要提供数据, 还要说明用了什么软件、软件版本, 甚至是运行该软件的代码
- 开放科学 (**open science**): 开放获取文章、开放数据、开放代码、开放审稿意见

# 需要掌握的工具越来越多

- 数据输入、分析和绘图: Excel、SPSS、SAS、stata、minitab、PAST...
- 建立和美化进化树, 如 BEAST、MEGA、PAUP\*、MrBayes、RAxML、SplitsTree...
- 绘图与图像编辑软件: SigmaPlot、Excel、PhotoShop、Adobe Illustrator、Origin、FigTree、Geneious、paint.net、Inkscape
- 地理信息系统: ArcGIS、QGIS、SAGA-GIS
- 物种潜在分布区预测: Maxent、randomForest、OpenModeller...
- 系统发育比较分析: Mesquite、PAML...
- 编写植物名录: MS Access、SQLite、MySQL
- 数据清洁: OpenRefine
- 文献管理: Zotero、Mendeley、Endnote 等等

如果有工具能完成上述大部分工作就好了.....这个工具, 就是 R。

# R 提供了无限可能

“数据处理中的事，只要你能想得到，R 都能做。”

主要取决于：

- ① 你对 R 掌握的程度；
- ② 电脑是否有电（手动狗头）

# R 能做什么？

- 计算器，加减乘除
- 各种数据处理和统计分析 (*t* 检验、方差分析、回归、线性模型)
- 绘图（生成从简单到复杂的各种图形）
- 空间信息处理以及地图绘制（GIS）
- 处理文档，包括制作幻灯片、文档
- 开发自己的模型和算法
- 连接和驱动数据库（如 SQLite、MySQL）
- 机器学习与人工智能.....

# 生物多样性信息处理举例

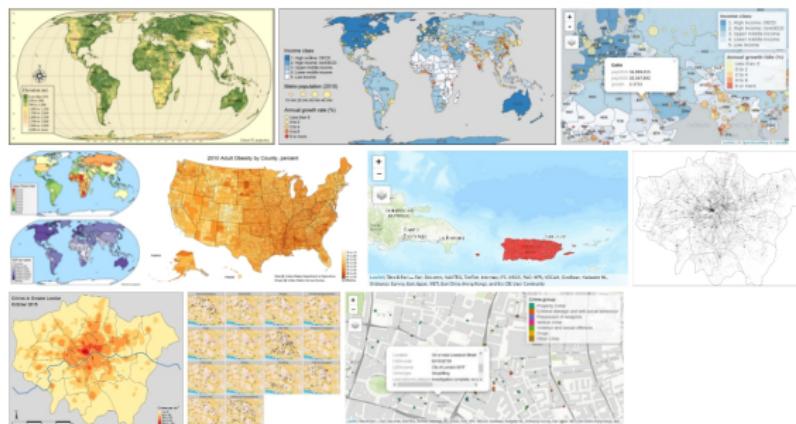
- 用 R 处理植物图库（按照科、属文件夹放置、提取 exif 信息、地标注化、添加水印、裁剪等等）
- 物种分布以及区系分区
- 进化树与系统发育分析
- Rmarkdown 与可重复研究，写论文，分享代码、写博客、写公众号、幻灯片.....

# 绘制地图 (sf + ggplot2 + tmap + raster + rgdal + spatial)

## tmap: thematic maps in R

[build](#) passing | License: GPL v3 | CRAN 3.3-2 | CRAN ERROR | downloads 24K/month

[tmap](#) is an actively maintained open-source R-library for drawing thematic maps. The API is based on [A Layered Grammar of Graphics](#) and resembles the syntax of [ggplot2](#), a popular R-library for drawing charts.



See [below](#) the source code for these images.

图 1: tmap 包绘制的地图

# 植物区系分区

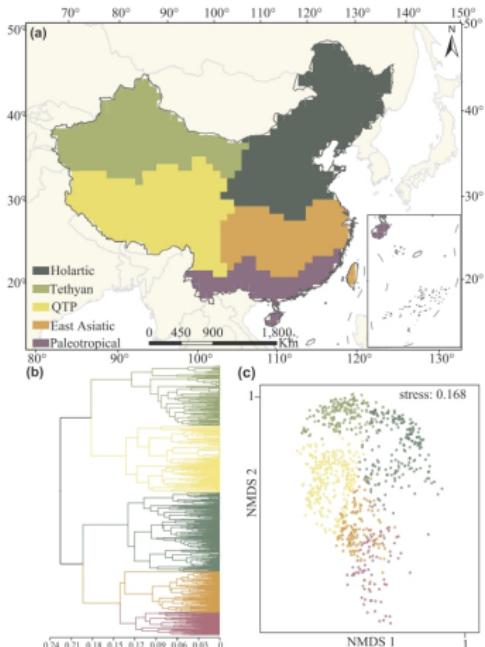


Fig. 3. Map (a) and dendrogram (b) based on UPGMA hierarchical clustering

图 2: 叶建飞等 (2019) 的中国植物区系分区方案

# 植物区系分区

## phyloregion



An R package for biogeographic regionalization (the classification of geographical areas in terms of their biotas) and macroecology. With macroecological datasets of ever increasing size and complexity, `phyloregion` offers the possibility of handling and executing very large scale analysis of biogeographic regionalization. It also allows fast and memory efficient analysis of standard conservation measures such as phylogenetic diversity, phylogenetic endemism, evolutionary distinctiveness and global endangerment. `phyloregion` can run on any operating system (Mac, Linux, Windows or even high performance computing cluster) with R 3.6.0 (or higher) installed.

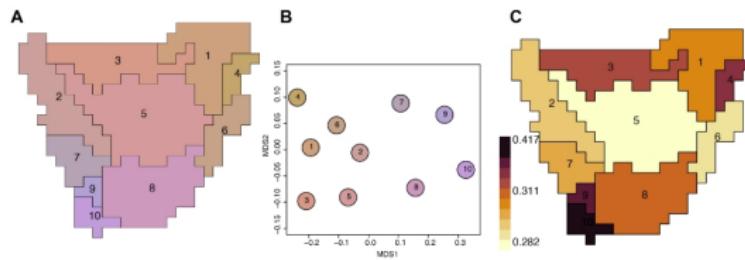


图 3: phyloregion 包

# 物种潜在分布区预测

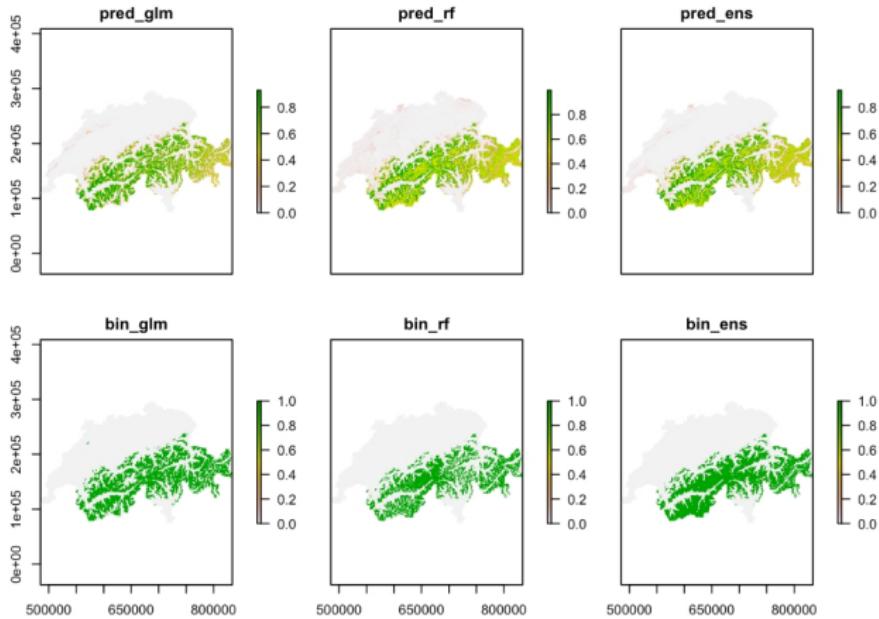
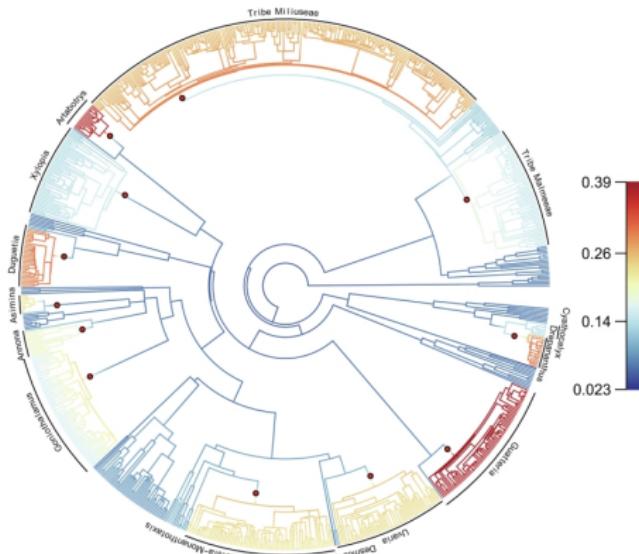


图 4: 物种潜在分布区预测 (<https://damariszurell.github.io/SDM-Intro/>)

## 绘制进化树 ape 包

B. Xue, et al.

*Molecular Phylogenetics and Evolution* 142 (2020) 106659



**Fig. 1.** Phylogenetic tree of Annonaceae with branches colored according to net diversification rate ( $\text{Myr}^{-1}$ ), resulting from BAMM analysis. Red dots indicate diversification rate shifts.

图 5: 薛彬娥等 (2020)

# 祖先性状重建 ape 包

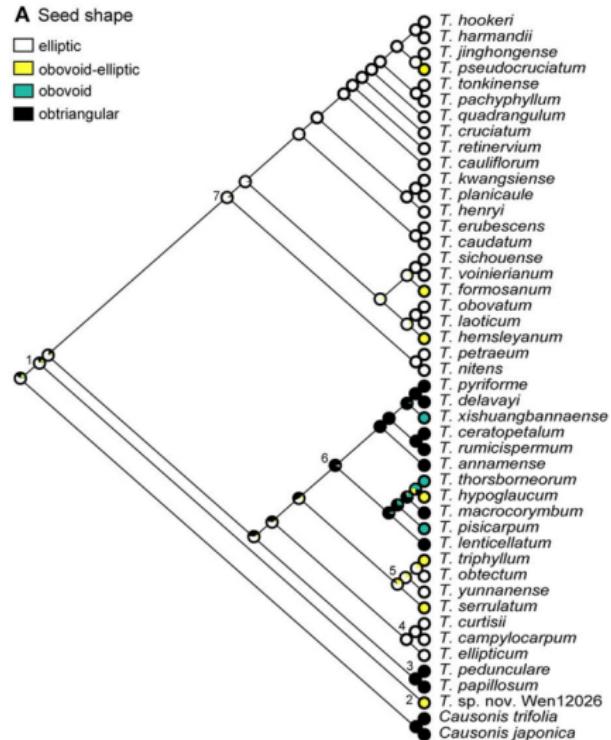
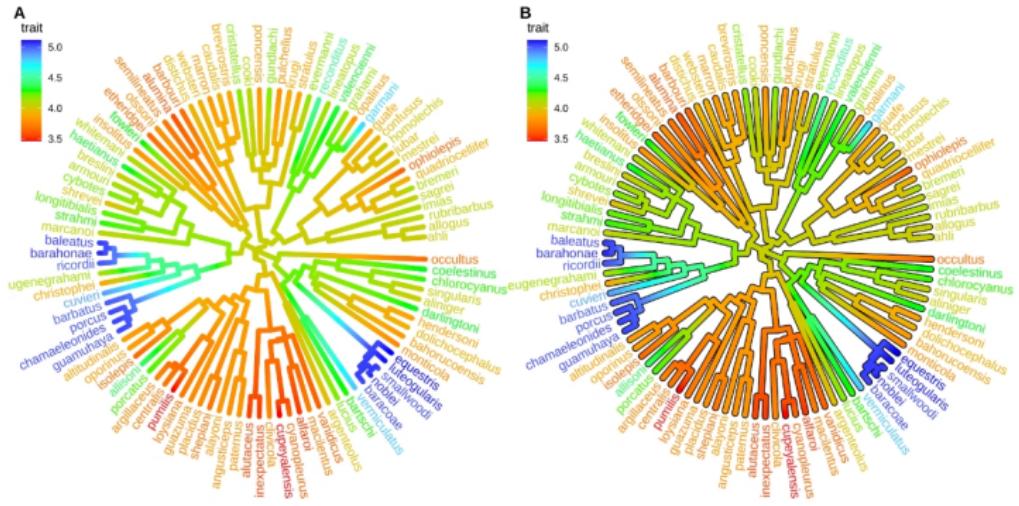


图 6: Habib et al (2017)

## 进化树与其他信息组合 ggTree 包



**Figure 4.11: Continuous state transition in edges.** Edges are colored by the values from ancestral trait to offspring.

图 7: Y 叔 (余光创教授) 神包 ggtree, 微信号 YuLabSMU

# 祖先分布区推断 BioGeoBEARS 脚本

BioGeoBEARS is short for “BioGeography with Bayesian (and likelihood) Evolutionary Analysis in R Scripts”

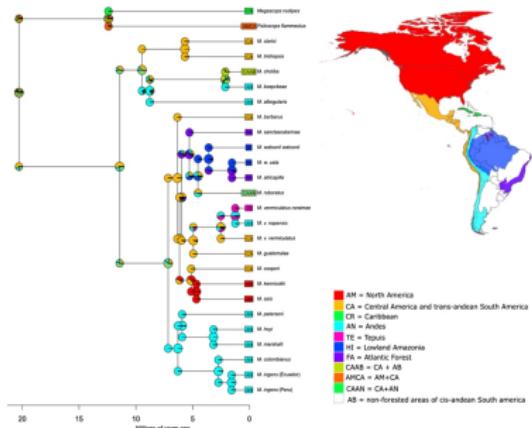


Fig. 3. Ancestral area reconstruction from BioGeoBEARS, derived from the Bayesian chronogram. The best-fit model was DIVIDER-3. Most likely biogeographic areas are shown in the circles, and the colors in the squares indicate the current species distribution. The map shows the location of the areas used. Combinations of two areas (for example CANAN) are not shown on the map. We considered Mexico as Central America for this study.

图 8: Dantas et al. 2015

# 关键性状与进化速率分析 geiger 包

B. Xue, et al.

Molecular Phylogenetics and Evolution 142 (2020) 106659

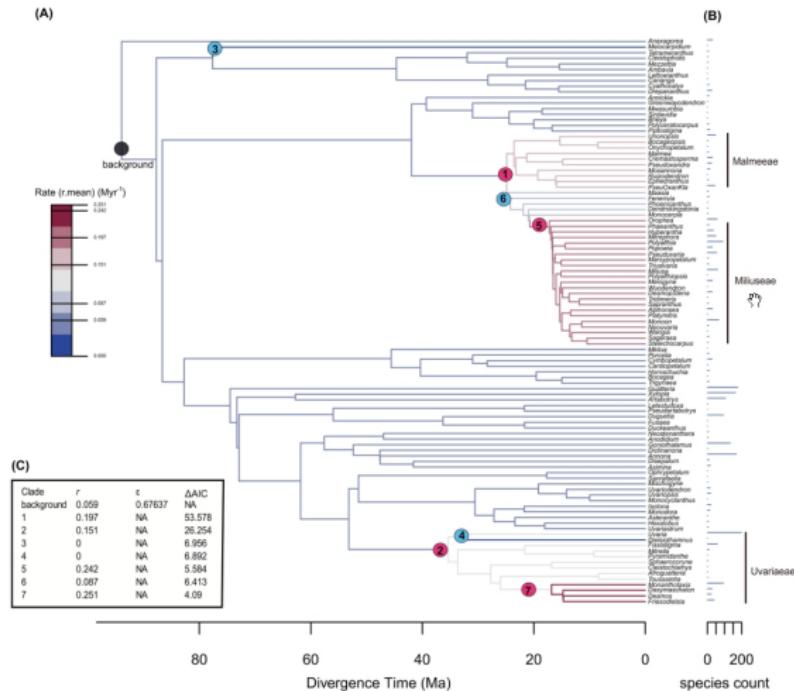


图 9: 薛彬娥等 (2020)

# 查询植物的科属和分类状态

- 批量查询科属：你熟悉 APG 系统吗？
- 根据中文名查询学名
- 根据拉丁名查询中文名和分布
- 根据拉丁名查询接受状态

相关程序包：`taxize`、`Taxonstand`、`lcvplants`、`plantlist` ...

# 2

## 编程环境

# 要安装的软件

- R (<https://www.r-project.org/>)
- Rstudio (<https://www.rstudio.com/>)
- Rtools (<https://cran.r-project.org/bin/windows/Rtools/>) (只适用于 Windows 用户)
- tinytex (<https://yihui.org/tinytex/>) 或 MiKTeX (<https://miktex.org>)
- git (<https://git-scm.com/>)
- 一些 R 程序包, 如 tidyverse 系列

注意, **Windows** 系统请将软件全部安装在 C 盘的默认位置, 否则程序将可能无法更新或运行。

# 编辑器、集成开发环境

任何文本编辑器都可以编辑 R 代码，但是目前一般使用：

- Rstudio (<https://www.rstudio.com/>)
- VScode (<https://code.visualstudio.com/download>)

注意：安装好 R 之后，再安装 Rstudio。如果是 Windows，想编写 R 程序包，还要将 R 添加到系统路径，以便从控制台（console）启动。

# 安装 Rtools (仅针对 Windows 用户)

- Rtools 是编译 R 程序包的一系列工具，如 `gcc`, `gfortran`, `Make` 等，有些 R 程序包含有 C、C++ 或 Fortran 编写的代码，这些代码需要用 rtools 编译为字节码才能运行 (.dll)。
- 下载 Rtools <https://cran.r-project.org/bin/windows/Rtools/>

# 将 Rtools 添加到系统路径

- 打开已经安装好的 R，在 Rconsole 中输入：

```
writeLines('PATH="${RTOOLS40_HOME}\\\usr\\\bin;${PATH}"', con = "~/.Renviron")
```

# 安装 TeX

TeX 是一套排版系统，主要用来生成 PDF 文件。但完整安装大约需要 10G 以上。

怎么办？

## 方案一：tinytex

Rstudio 公司的谢益辉编写了 tinytex 包（R 包），用来安装一些常用的 TeX 包。

安装方法，先在 Rconsole 中输入 `install.packages("tinytex")`，再输入 `tinytex::install_tinytex()` 即可自动下载和安装，大约 90M。

## 方案二：安装 MiKTeX

MiKTeX 是 Tex 包的集合，精选了一些基础包，大约 130M，需要时可自动从互联网下载和安装所需要的包。

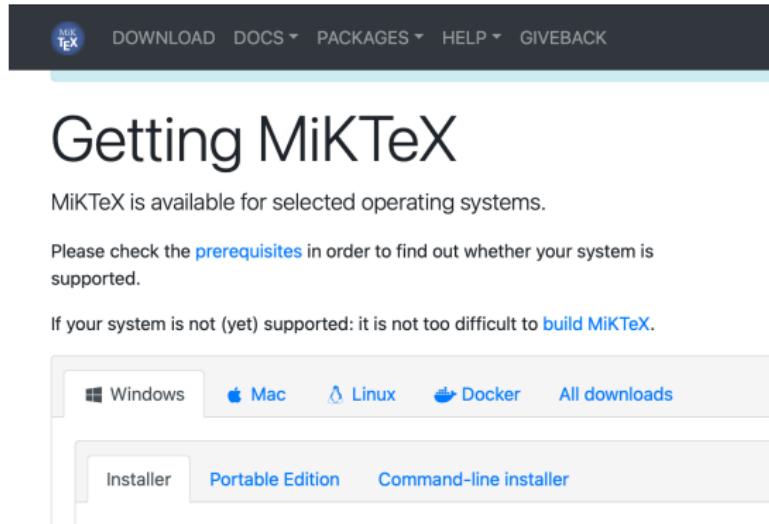


图 10: 下载 MIKTeX <https://miktex.org/download>

# 版本控制软件 git

git 是版本控制软件，可以记录文档（本课程中为 R 脚本、Rmarkdown 文档等）的修改，并且可以退回到之前的任何一个版本。

下载 git <https://git-scm.com>



图 11: 下载 git

# 3

## 熟悉 R 和 Rstudio

# 演示 RGUI

- 运行代码
- 加载程序包: `library(picante)`
- 安装程序包: `install.packages("picante")`
- 查询帮助: `?`

# 熟悉 Rstudio

- ① 创建新的 R 脚本: File > New File > R Script
- ② 若干工具 Tools
  - ① 安装程序包 Install packages
  - ② 更新程序包 Check for R package Updates
  - ③ 配置更改主题等 Global Options
- ③ 查询帮助 Help:
- ④ Rstudio 本身的更新: check for updates

# 4

## R 脚本 R script

# 啥是 R 脚本？

R 脚本是保存了 R 命令的文本文件，一般扩展名为.R 或.r。

```
a <- 1
b <- 3
c <- a + b

plot(1:10)
plot(1:20, pch = 1:20)
```

# R 脚本举例

```
# 华氏度转换为摄氏度
# by 张金龙
# 2021-9-20

setwd("/Users/jinlong/Desktop/temperature") # 设定工作路径
library(readxl)
library(lubridate)

## Conversion according to
## https://www.weather.gov/media/epz/wxcalc/tempConvert.pdf
convert_fahr_to_celsius <- function(temp) {
  celsius <- ((temp - 32) * (5 / 9))
  return(celsius)
}
dat1 <- read_excel("temperature1.xlsx", sheet = "Sheet1")
.....
```

# 完整 R 脚本的结构

- ① 说明此脚本的目的、作者以及日期、版本等
- ② 设定工作目录 `setwd()`
- ③ 加载程序包 `library()` 或者导入外部函数
- ④ 读取数据 `read.csv()`
- ⑤ 数据处理 `na.omit()`
- ⑥ 建立模型 `lm()`、`glm()`...
- ⑦ 绘图 `plot()`
- ⑧ 保存数据 `write.csv()`、`save.image()`

注意：R 脚本最好不需要输入任何数据，或者选择任何文件，就能运行。

# Rstudio 正确显示 R 脚本中的中文（UTF-8 编码）

中文纯文本文件，一般有 GB2312、GBK、GB18030、UTF-8 等编码方式。

Rstudio 的编码一般跟随操作系统，但最好设定编码方式为 UTF-8，这样，文本文件都以 UTF-8 编码保存，中文不会出现乱码。

UTF-8 编码，可从 Tools > Global Options 设定：

```
Options > Code > Saving >  
Default text Encoding> UTF-8
```

# 在 Rstudio 新建 R 脚本

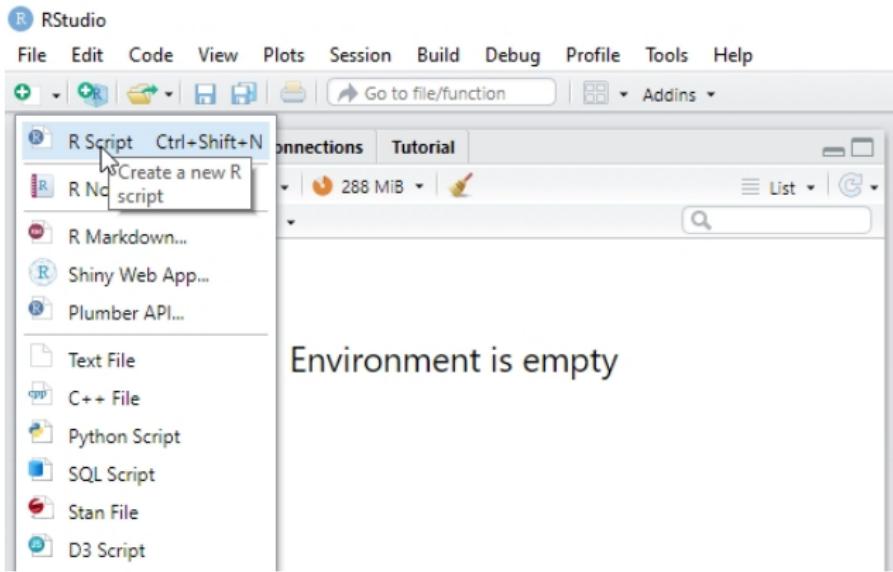
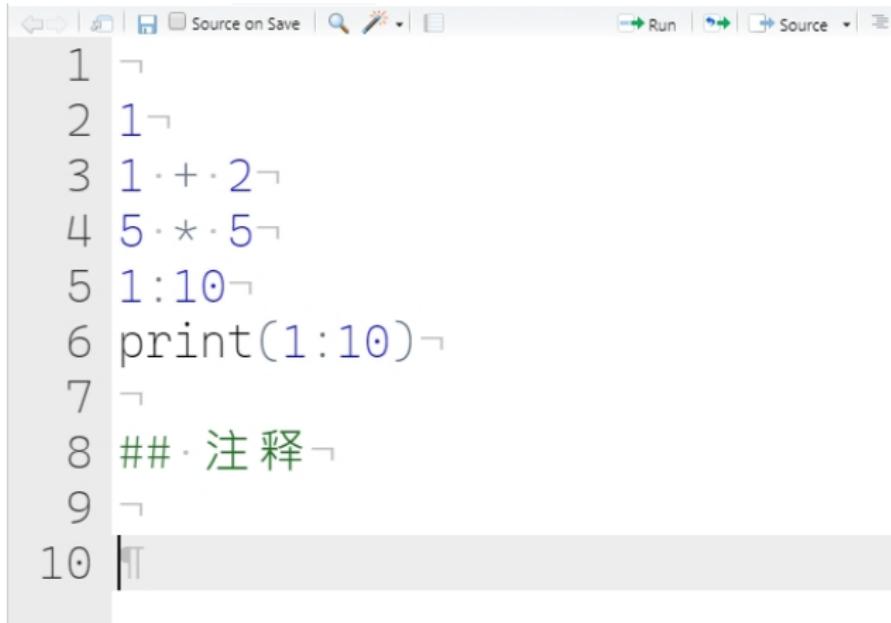


图 12: 新建 R 脚本

# 怎么运行 R 脚本?

- 在 Rstudio 的 Source 窗口打开，发送到 Console 运行
- 或者复制 + 粘贴到 Console 运行

# 运行 R 脚本



```
1 #  
2 1  
3 1 + 2  
4 5 * 5  
5 1:10  
6 print(1:10)  
7 #  
8 ## · 注释  
9 #  
10 #
```

图 13: 点击右上角的 run 运行本行, 点击 source 运行整个脚本

# 查询 R 帮助

- `? 函数名`: 查询函数的帮助文档
- `help(函数名)`: 查询函数的帮助文档
- `help.start()`: 打开帮助页面
- `apropos()`: 已知部分函数名, 列出完整的函数名
- `str()`: 查看数据的基本结构
- `head()`: 查看数据的前几行或者前几个元素

# 工作路径 working directory

- `setwd()`: 设定工作路径, 注意, Windows 要改为\\或者/
- `getwd()`: 查询当前的工作路径
- `list.files()`: 列出工作文件夹下所有文件的名称

# 工作空间 Session

`sessionInfo()`, 查询 R 的版本和加载程序包的版本

- 打开 R, 就打开一个 `Session`
- 关闭 R, 就关闭了这个 `Session`

关闭了 R 之后, `session` 中原有的对象, 如运算结果, 就全部消失了。

# 工作空间有关的函数

要将输入的 R 命令，计算得到的结果，保存起来，就要用到管理工作空间的函数。

```
ls()      # 查看工作空间的对象  
data()    # 加载数据  
save()    # 保存数据成为二进制文件，扩展名为 .Rda  
save.image() # 保存工作空间为二进制文件，扩展名为 .Rdata  
load()    # 加载二进制文件，还原 R 工作空间  
savehistory(file = ".Rhistory") # 保存 R 历史记录  
rm()      # 删除对象
```

5

## 程序包

# 什么是 R 程序包?

- 如果把若干 R 函数、帮助文档、数据以及相应说明文档按照一定的方式组织并压缩，就组成了一个 R 程序包（package）。
- 一个 R 包执行一个或者多个特定的功能。
- Windows、MacOS 平台下都有自己对应的程序包，都是由源代码编译而成。
- 注意 zip 文件不要去手动解压缩。

# 默认安装的程序包

安装 R 后，只安装 “stats, graphics, grDevices, utils, datasets, methods, base” 等包，其他程序包需要再手动安装：

```
install.packages("ape")
install.packages("ggplot2")
```

某些程序包依赖（depending）于另外一些程序包，在安装时能自动安装。

# 通用 R 包和专用 R 包

- **base 系列程序包**: R 的基础程序包，可以实现基本的统计、绘图、文字处理、文件操作等
- **tidyverse 系列程序包**: 主要用于数据科学，如数据整理和绘图等
- **某一领域专用的程序包**: 如系统发育分析、生物多样性分析、空间分析、绘图等

# 常用 R 包

- `testthat`、`devtools`、`remotes`: R 程序包的安装和测试
- `ggplot2` 绘图、`RColorBrewer` 配色, `cowplot`、`patchwork` 等拼图
- `Rsqlite`、`dbplyr`: 操作数据库
- `rmarkdown`、`knitr`、`bookdown`、`blogdown`、`shiny` 等: 文档排版,  
可编译 markdown 和 rmarkdown 文档, 生成 word、html、pdf 等

# 生态学常用 R 包

- **vegan**: 排序、生物多样性指数、群落特征等计算
- **ape**、**ggtree**、**phytools**: 进化分析和绘图
- **sf**、**raster**、**tmap**、**spatial**、**terra**: 空间数据处理以及绘制地图
- **dismo**、**ENMeval**、**wallace**、**biomod**: 物种潜在分布区预测
- **metafor**、**meta**、**forestplot**: 荟萃分析
- **lme4**、**r2glmm**、**lmerTest**、**MuMIn**: 混合模型
- **rstan**: 贝叶斯层次模型

# tidyverse 的 8 个核心程序包

- **ggplot2**: 图层叠加式高级绘图
- **dplyr**: 提供数据的链式转换，或者称为”管道”
- **tidyverse**: 包括 `pivot_longer()`、`pivot_wider()` 等函数，主要用于长宽数据格式之间的转换
- **readr**: 读取各种常用数据格式
- **purrr**: 减少循环、`apply` 的使用
- **tibble**: 优化过的数据表结构
- **stringr**: 字符串处理
- **forcats**: 处理因子（factor）类型的向量

# tidyverse 系列程序包扩展

- DBI、RSQLite、RMariaDB、RPostgres、odbc：连接数据库
- haven：读取 SPSS、Stata、SAS 数据
- httr：网站 APIs
- readxl：读取 Excel 文件
- googlesheets4、googledrive：读取 Google 文档
- rvest：爬取网页
- jsonlite：读取 JSON
- xml2：读取 XML

# 在哪里找程序包?

- CRAN 网站: 各种类别的程序包, 参考 CRAN task views, 用 `install.packages()` 安装
- Bioconductor 网站: 基因组学和生物信息学方向
- github 网站: 多为开发阶段的程序包, 用 `devtools`、`remotes` 安装
- Rforge 网站: 多为开发阶段的程序包, 现在已不太活跃

# R 包的搜索引擎

- <https://www.r-pkg.org>
- <https://rdrr.io>
- <https://www.rdocumentation.org> (仅限 CRAN 和 Bioconductor)

# CRAN, R 的镜像网站

## CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: [main page](#), [windows release](#), [windows old release](#).

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

0-Cloud

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

Automatic redirection to  
servers worldwide,  
currently sponsored by  
Rstudio

图 14: CRAN 镜像

# 国内的 CRAN 镜像

China

<a href="https://mirrors.tuna.tsinghua.edu.cn/CRAN/">https://mirrors.tuna.tsinghua.edu.cn/CRAN/</a>	TUNA Team, Tsinghua University
<a href="https://mirrors.bfsu.edu.cn/CRAN/">https://mirrors.bfsu.edu.cn/CRAN/</a>	Beijing Foreign Studies University
<a href="https://mirrors.ustc.edu.cn/CRAN/">https://mirrors.ustc.edu.cn/CRAN/</a>	University of Science and Technology of China
<a href="https://mirror-hk.koddos.net/CRAN/">https://mirror-hk.koddos.net/CRAN/</a>	KoDDoS in Hong Kong
<a href="https://mirrors.e-ducation.cn/CRAN/">https://mirrors.e-ducation.cn/CRAN/</a>	Elite Education
<a href="https://mirror.lzu.edu.cn/CRAN/">https://mirror.lzu.edu.cn/CRAN/</a>	Lanzhou University Open Source Society
<a href="https://mirrors.nju.edu.cn/CRAN/">https://mirrors.nju.edu.cn/CRAN/</a>	eScience Center, Nanjing University
<a href="https://mirrors.tongji.edu.cn/CRAN/">https://mirrors.tongji.edu.cn/CRAN/</a>	Tongji University
<a href="https://mirrors.sjtu.edu.cn/cran/">https://mirrors.sjtu.edu.cn/cran/</a>	Shanghai Jiao Tong University
<a href="https://mirrors.sustech.edu.cn/CRAN/">https://mirrors.sustech.edu.cn/CRAN/</a>	Southern University of Science and Technology (SUSTech)

图 15: CRAN 镜像

# 从 github 安装程序包

The screenshot shows Jinlong Zhang's GitHub profile at <https://github.com/helixcn>. The profile features a circular background image divided into nine colored quadrants (blue, green, yellow, red, purple, grey). Below the profile picture, the user's name is listed as Jinlong Zhang and their GitHub handle as helixcn. A bio states: "Ecologist working on species distribution". There is a button to "Edit profile". At the bottom, it shows 133 followers, 305 following, and 694 commits. The GitHub header includes links for Pull requests, Issues, Marketplace, and Explore. The "Overview" tab is selected. Below the tabs, there is a section titled "Pinned" containing three package cards:

- EcoEvoLinks**: Links for Ecology and Evolution. Rating: 19 stars, 14 forks. Languages: R.
- phylotools**: Phylogenetic tools for Eco-phylogenetics. Rating: 4 stars, 5 forks. Languages: R.
- R\_phylo\_refcard**: PDF Reference Card for Phylogenetic Comparative Analysis in R. Rating: 4 stars, 5 forks. Languages: TeX.

图 16: github 个人网页

# 安装 CRAN 上的程序包

在 Rconsole 中输入，R 即可完成程序包的下载和安装

```
install.packages("ape")
# 安装一个

install.packages(c("ape", "vegan"))
# 安装多个

install.packages("tidyverse",
  dependencies = TRUE)
# 安装一系列，并安装所有依赖的程序包
```

# 安装 Github 上的程序包

举例：从源代码安装 github 上的 `phylotools` 包

```
library(devtools)
install_github("helixcn/phylotools",
  build_vignettes = TRUE)
```

如果还没有安装 `devtools`, 则输入

```
install.packages("devtools")
```

# 安装 Bioconductor 上的程序包

```
if (!requireNamespace("BiocManager",
  quietly = TRUE))
install.packages("BiocManager")

BiocManager::install("ggtree")
```

# 有多少程序包?

答案 (2022 年 1 月):

- CRAN 上, 18668
- Bioconductor 上, 2083
- Rforge 上, 2149
- Github 上, .....

# R 程序包的分类介绍 CRAN Task Views

<a href="#"><u>Optimization</u></a>	Optimization and Mathematical Programming
<a href="#"><u>Pharmacokinetics</u></a>	Analysis of Pharmacokinetic Data
<a href="#"><u>Phylogenetics</u></a>	<b>Phylogenetics, Especially Comparative Methods</b>
<a href="#"><u>Psychometrics</u></a>	Psychometric Models and Methods
<a href="#"><u>ReproducibleResearch</u></a>	Reproducible Research
<a href="#"><u>Robust</u></a>	Robust Statistical Methods
<a href="#"><u>SocialSciences</u></a>	Statistics for the Social Sciences
<a href="#"><u>Spatial</u></a>	Analysis of Spatial Data
<a href="#"><u>SpatioTemporal</u></a>	Handling and Analyzing Spatio-Temporal Data
<a href="#"><u>Survival</u></a>	Survival Analysis
<a href="#"><u>TeachingStatistics</u></a>	Teaching Statistics

图 17: CRAN 包的分类介绍

# CRAN Task View: Phylogenetics 网页

## CRAN Task View: Phylogenetics, Especially Comparative Methods

**Maintainer:** Brian O'Meara

**Contact:** omeara.brian at gmail.com

**Version:** 2021-02-25

**URL:** <https://CRAN.R-project.org/view=Phylogenetics>

The history of life unfolds within a phylogenetic context. Comparative phylogenetic met  
analyzing historical patterns along phylogenetic trees. This task view describes R packag  
comparative phylogenetic methods. This is an active research area and much of the infor  
note is that many important packages are not on CRAN: either they were formerly on CR  
example, if they failed to incorporate necessary changes as R is updated) or they are deve  
on CRAN yet. Such packages may be found on GitHub, R-Forge, or authors' websites.

图 18: CRAN Task Views 关于系统发育相关程序包的介绍

# 安装某个 CRAN Task Views 的全部程序包

```
install.packages("ctv") # 只需要安装一次
library(ctv)
install.views("Phylogenetics")
install.views("Environmetrics")
```

# 每个上传到 CRAN 的程序包都有页面

## **picante: Integrating Phylogenies and Ecology**

Functions for phylocom integration, community analyses, null-models, traits and evolution. Implements numerous ecophylogenetic approaches including measures of community phylogenetic and trait diversity, phylogenetic signal, estimation of trait values for unobserved taxa, null models for community and phylogeny randomizations, and utility functions for data input/output and phylogeny plotting. A full description of package functionality and methods are provided by Kembel et al. (2010)

[<doi:10.1093/bioinformatics/btq166>](https://doi.org/10.1093/bioinformatics/btq166).

Version:	1.8.2
Depends:	<a href="#">ape</a> , <a href="#">vegan</a> , <a href="#">nlme</a> , methods
Suggests:	<a href="#">brglm</a> , <a href="#">circular</a> , <a href="#">corpcor</a> , <a href="#">quantreg</a>
Published:	2020-06-10
Author:	Steven W. Kembel, David D. Ackerly, Simon P. Blomberg, Will K. Cornwell, Peter D. Cowan, Matthew R. Helmus, Helene Morlon, Campbell O. Webb
Maintainer:	Steven W. Kembel <steve.kembel@gmail.com>
License:	<a href="#">GPL-2</a>
NeedsCompilation:	yes
Citation:	<a href="#">picante citation info</a>
Materials:	<a href="#">README</a>
In views:	<a href="#">Phylogenetics</a>
CRAN checks:	<a href="#">picante results</a>

图 19: picante 程序包的网页

# CRAN 上 picante 包的页面

Downloads:

Reference manual: [picante.pdf](#)

Vignettes: [Data and analysis in picante](#)

Package source: [picante\\_1.8.2.tar.gz](#)

Windows binaries: r-devel: [picante\\_1.8.2.zip](#), r-devel-UCRT: [picante\\_1.8.2.zip](#), r-release: [picante\\_1.8.2.zip](#), r-oldrel: [picante\\_1.8.2.zip](#)

macOS binaries: r-release (arm64): [picante\\_1.8.2.tgz](#), r-release (x86\_64): [picante\\_1.8.2.tgz](#), r-oldrel: [picante\\_1.8.2.tgz](#)

Old sources: [picante archive](#)

Reverse dependencies:

Reverse depends: [CommEcol](#), [idar](#), [RPANDA](#)

Reverse imports: [betapart](#), [CountClust](#), [DAMOCLES](#), [nodiv](#), [PCPS](#), [pez](#), [recluster](#)

Reverse suggests: [cati](#), [EcoIndR](#), [ecolottery](#), [enveomics.R](#), [fishtree](#), [microeco](#), [phyr](#), [Rarefy](#), [RRphylo](#)

图 20: 每个 R 程序包都尽量提供多种操作系统的版本

# 加载程序包

```
> library(tidyverse)
-- Attaching packages --
v ggplot2 3.3.5     v purrr   0.3.4
v tibble   3.1.2     v dplyr    1.0.7
v tidyr    1.1.3     v stringr  1.4.0
v readr    1.4.0     vforcats  0.5.1
-- Conflicts --
x dplyr::collapse() masks nlme::collapse()
x dplyr::filter()   masks stats::filter()
x dplyr::lag()      masks stats::lag()
> |
```

图 21: 用 `library()` 函数加载程序包

# 正确引用程序包

如何正确引用 `ape` 程序包?

```
citation("ape")
```

To cite `ape` in a publication please use:

Paradis E. & Schliep K. 2019. `ape` 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.

# 正确引用 R

```
citation()
```

To cite R in publications use:

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

# 6

怎么学习 R?

# 我能学好 R 吗?

问：如果没有很好的数学和计算机基础，能学好 R 吗？

答：能。会用 Excel，就能很快学会 R 的操作

最重要的，是要清楚自己要解决的问题本质是什么，并将其分解成一步一步的程序化操作，这样就能知道寻找什么程序包和函数，甚至自己开发函数和程序包。

# 几个“小目标”

- 把 R-base cheatsheet 背熟。
- 看几本经典的统计学入门书籍，掌握常见统计检验的原理，知道什么情况用什么统计检验，重复书中的例子
- 熟悉 `base`、`stats`、`graphics`、`ggplot2`、`dplyr` 等核心程序包的常用函数。
- 遇到不熟悉的函数，及时阅读帮助文档，通过运行例子去理解文档
- 多阅读别人的代码，包括程序包的源代码，熟悉工作流程
- 多写多练，注意遵循代码编写的规范

# R 是一种工具

- R 本身是工具
- R 能帮我们获得想要的结果，但更重要还是理解结果的成因及意义

# R 的一些学习资源

- R 的一些学习资源  
<https://github.com/helixcn/awosome-R-ecology-evolution>
- Youtube R 语言编程视频合集  
[https://pan.baidu.com/s/1\\_wyj1sTeb7wp4cWFwk6Vzw](https://pan.baidu.com/s/1_wyj1sTeb7wp4cWFwk6Vzw) 提取码:  
6pfj
- 微信订阅: #r 语言主题
- 微信公众号: **ecoinformatics**、生态学文献分享、小明的数据分析笔记本、R 语言与藻类生态学、Y 叔 (YuLabSMU)、生信小白鱼、listenlii、科白君的土壤世界、庄闪闪的 R 语言手册.....

# 图书资料

- Kabacoff R I. *R in action: data analysis and graphics with R.* Simon and Schuster, 2015.
- *R for data science* <https://r4ds.had.co.nz>
- Douglas A. et al. *An Introduction to R* <https://intro2r.com>
- Paradis E, *R for beginners*  
[https://cran.r-project.org/doc/contrib/Paradis-rdebut\\_en.pdf](https://cran.r-project.org/doc/contrib/Paradis-rdebut_en.pdf)
- 李东风的《R 语言教程》[https://www.math.pku.edu.cn/teachers/lidif/docs/Rbook/html/\\_Rbook/Rbook.pdf](https://www.math.pku.edu.cn/teachers/lidif/docs/Rbook/html/_Rbook/Rbook.pdf)
- R cheat sheets (<https://www.rstudio.com/resources/cheatsheets/>)

# 我能学好 R 吗?

- 多读资料、多钻研：书读百遍，其义自见
- 熟悉以后，自然就有意思了：Nothing is fun until you're good at it. -- Amy Chua
- 记笔记，整理清楚：If you're thinking without writing, you only think you're thinking. -- Leslie Lamport

# 7

讲了什么

# 讲了什么

- ① R 是一种用于数据处理、统计和绘图的脚本语言
- ② 安装和配置 R、Rstudio、Rtools、tinytex、MiKTeX、git
- ③ R 脚本的结构和运行
- ④ 常用 R 程序包
- ⑤ 怎么学习 R

敬请关注《R 语言简明教程》第二课：R 数据类型及基本操作。详情请留意 ecoinformatics 的通知。

# 课程信息

- 本课程是在马克平研究员的鼓励和支持下录制的，由陈建平老师提供技术支持，全部视频和代码都在 BioONE—生物多样性与生态安全大数据网络平台发布。

完整课程视频和代码敬请关注：

- <https://www.bio-one.org.cn>
- 微信公众号： **ecoinformatics**
- 课程材料和源代码：<https://github.com>Hello-BD/R-in-brief>

# 致谢

感谢提出意见的各位老师、同学（以下排名不分次序），特别是：  
陈建平、赖江山、赵宇莹、刘璐妹、余光创、向小果、胡文浩、安瑞志、  
杨拓、朱慧玲、乔慧捷、文香英、葛学军、曾思金、刘景欣、杨雪飞、  
胡小丽、李勤、崔煜文等。

# 信息反馈

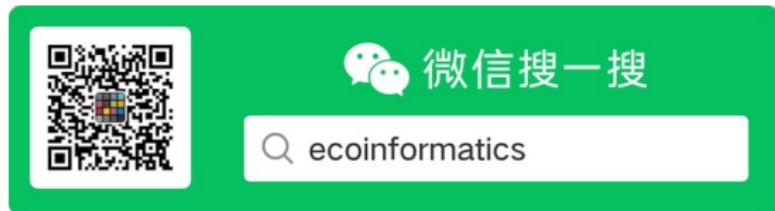


图 22: 敬请关注 `ecoinformatics`

- 如有任何问题、意见或者建议，欢迎联系本人
- Email: `jinlongzhang01@gmail.com`

## 附录：将 R 添加到启动路径（Windows）

目的：从 CMD 命令行使用 R 和 Rcmd 等

- R 默认安装在（请注意 R 的版本）：

C:\Program Files\R\R-4.x.x\bin\i386

方法：请参考本幻灯片的附录

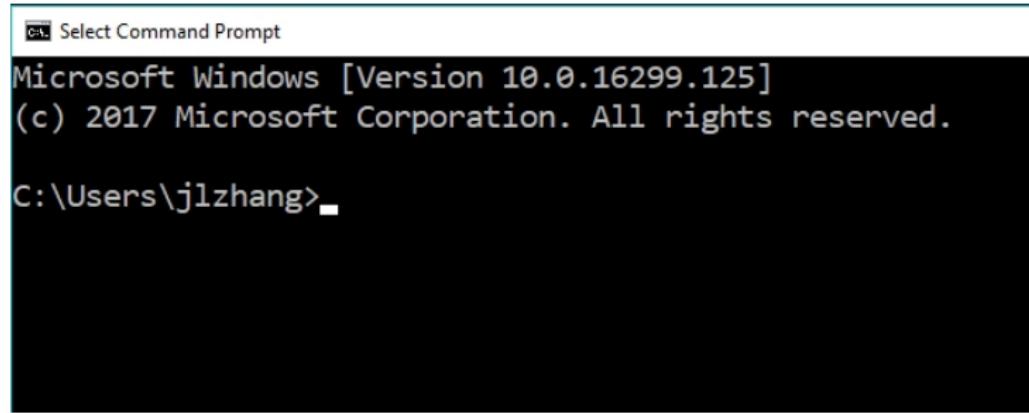


图 23: 打开 Windows CMD

# 附录：将 R 添加到系统路径 1

我的电脑 > 右键 > 属性

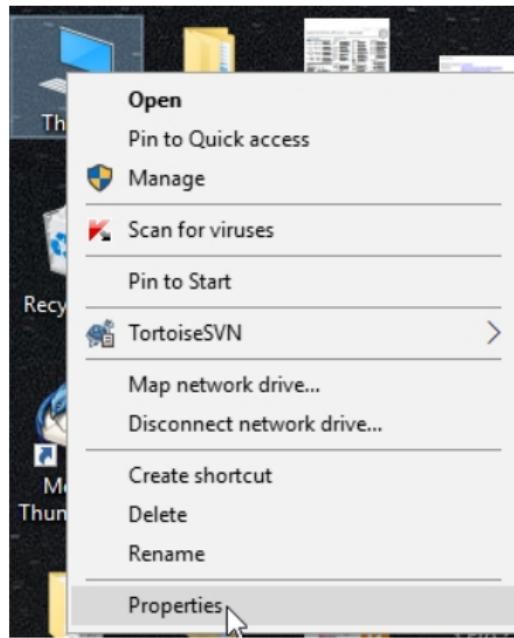


图 24: 桌面 > 我的电脑 > 属性

# 附录：将 R 添加到系统路径 2

导航栏 > 我的电脑 > 右键 > 属性



图 25: 导航栏 > 我的电脑 > 属性

# 附录：将 R 添加到系统路径 3

Windows10 系统页面，点击左侧的 Advanced system settings

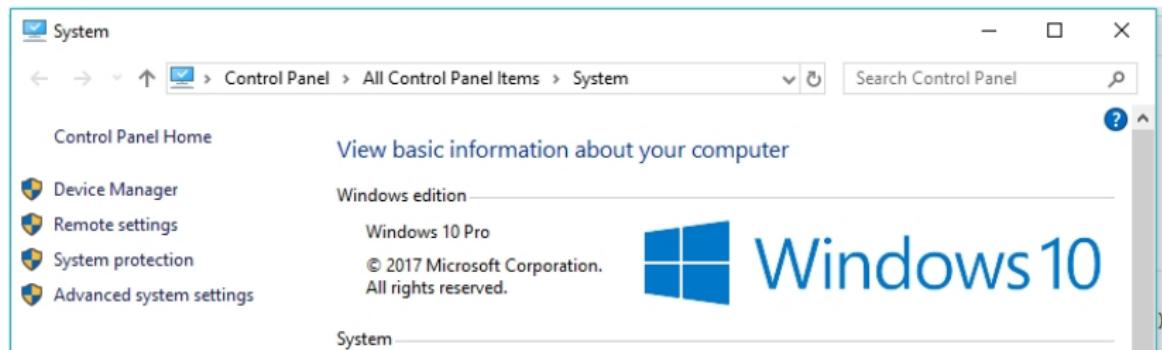


图 26: Window10 系统信息

# 附录：将 R 添加到系统路径 4

系统属性窗口，点击环境变量（Environment Variables）按钮

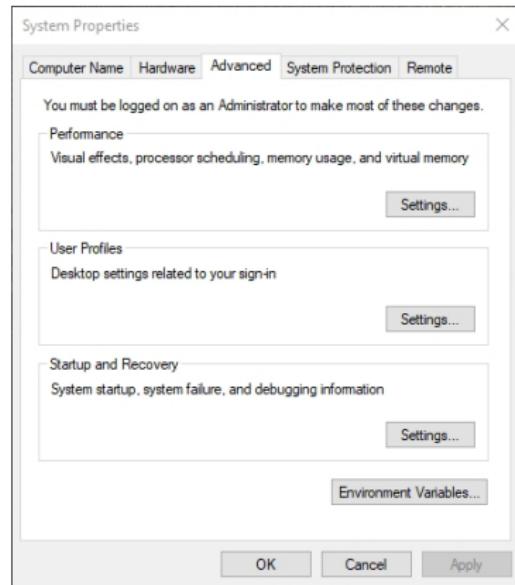


图 27: 高级 > 环境变量

# 附录：将 R 添加到系统路径 5

在系统变量 (System variables) 一栏，找到 PATH 变量，点击”编辑 Edit”

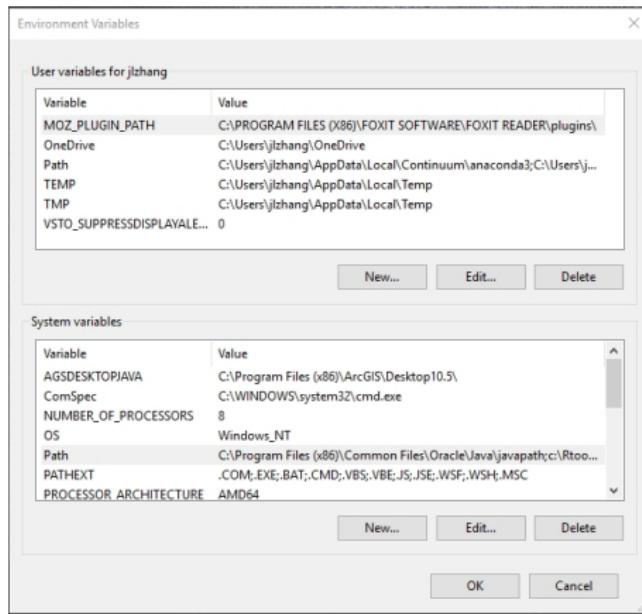


图 28: 系统变量 >Path> 编辑

# 附录：将 R 添加到系统路径 6

编辑环境变量窗口，点击新建（New）

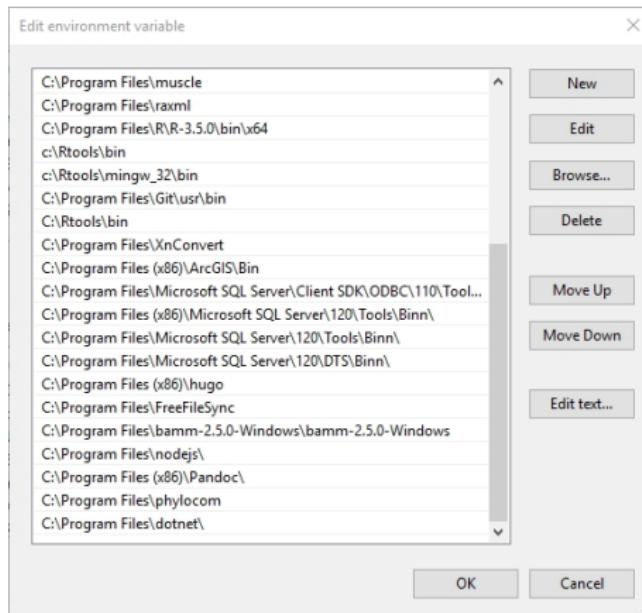
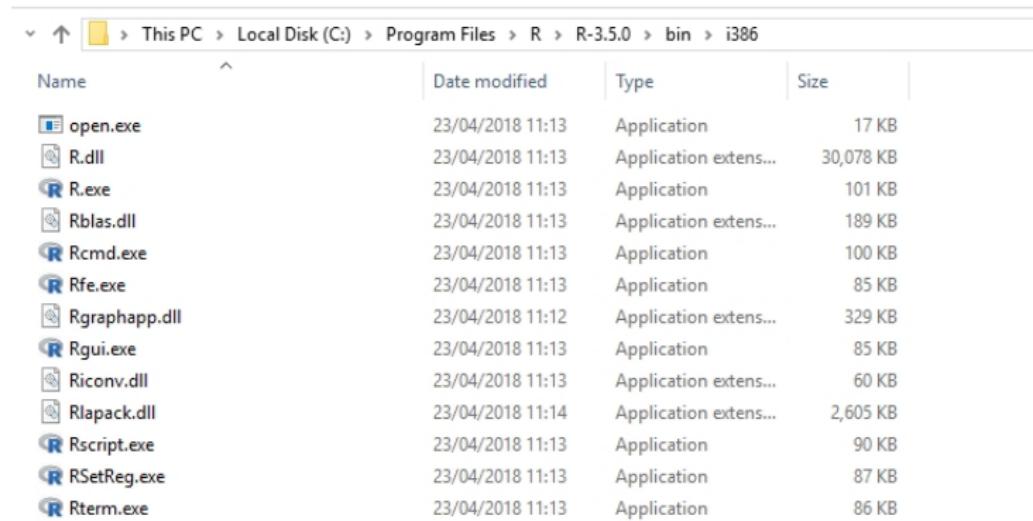


图 29: 编辑环境变量 > 新建

# 附录：将 R 添加到系统路径 7

找到 R 安装路径（32 位 R 的文件夹为 i386; 64 位为 x64, 请根据需要拷贝）



Name	Date modified	Type	Size
open.exe	23/04/2018 11:13	Application	17 KB
R.dll	23/04/2018 11:13	Application extens...	30,078 KB
R.exe	23/04/2018 11:13	Application	101 KB
Rblas.dll	23/04/2018 11:13	Application extens...	189 KB
Rcmd.exe	23/04/2018 11:13	Application	100 KB
Rfe.exe	23/04/2018 11:13	Application	85 KB
Rgraphapp.dll	23/04/2018 11:12	Application extens...	329 KB
Rgui.exe	23/04/2018 11:13	Application	85 KB
Riconv.dll	23/04/2018 11:13	Application extens...	60 KB
Rlapack.dll	23/04/2018 11:14	Application extens...	2,605 KB
Rscript.exe	23/04/2018 11:13	Application	90 KB
RSetReg.exe	23/04/2018 11:13	Application	87 KB
Rterm.exe	23/04/2018 11:13	Application	86 KB

图 30: R 软件默认安装路径（注意 R 的版本不同，路径也不同）

# 附录：将 R 添加到系统路径 8

## 拷贝 R 安装路径 2

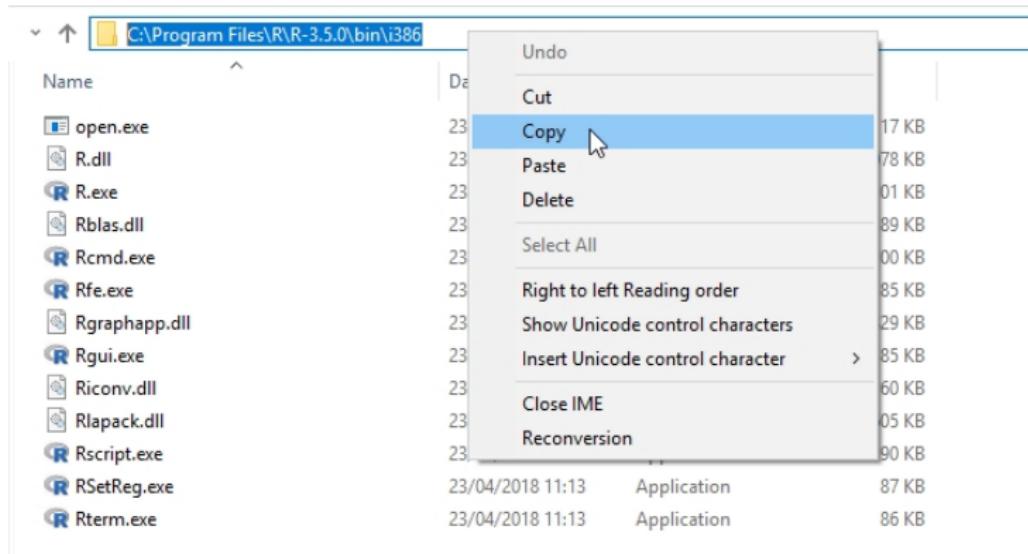


图 31: 拷贝 R 安装路径（注意 R 的版本不同，路径也不同）

# 附录：将 R 添加到系统路径 9

添加到 Path 之后，点击确定，关闭所有窗口

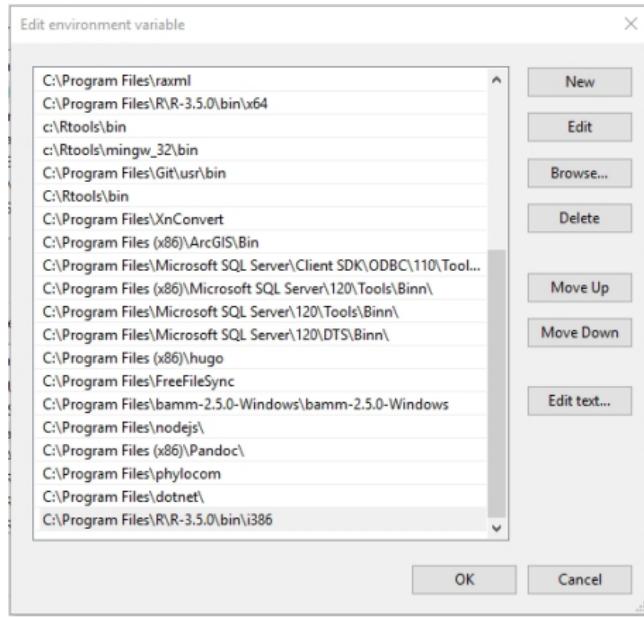


图 32: 将 R 添加到系统路径（注意 R 的版本不同，路径也不同）

## 附录：软件安装是否成功？

Windows 打开 cmd， MacOS 和 Linux 打开 Terminal  
在搜索 > 输入 cmd> 提示 Command Prompt

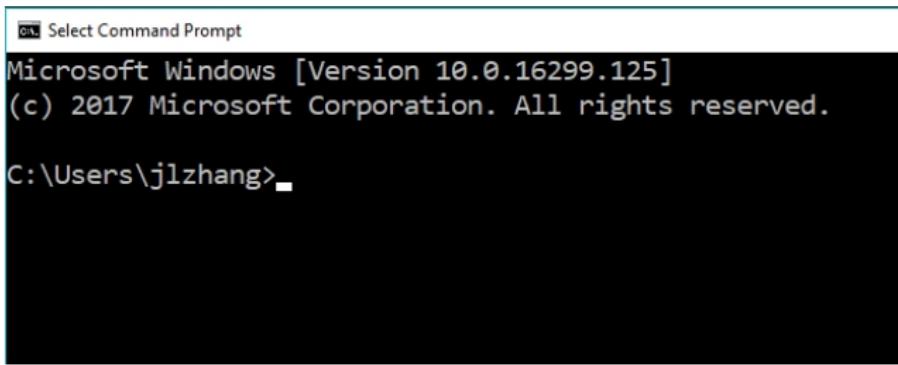
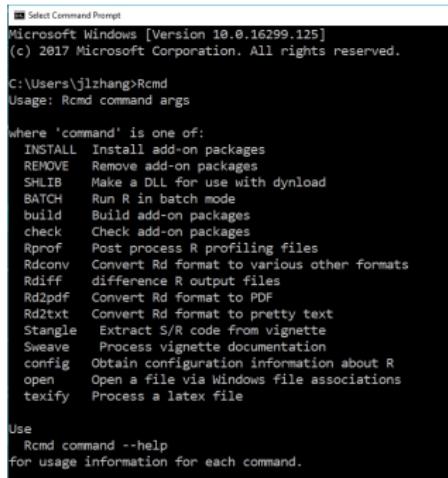


图 33: 打开 cmd

# 附录：检查 R 软件路径的配置是否成功：Rcmd

- Rcmd 主要用来安装和检查 R 程序包
- 在 command prompt 中输入 Rcmd
- Linux 和 MacOS 需要在 Terminal 中输入 R CMD



```
Microsoft Windows [Version 10.0.16299.125]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\jlzhang>Rcmd
Usage: Rcmd command args

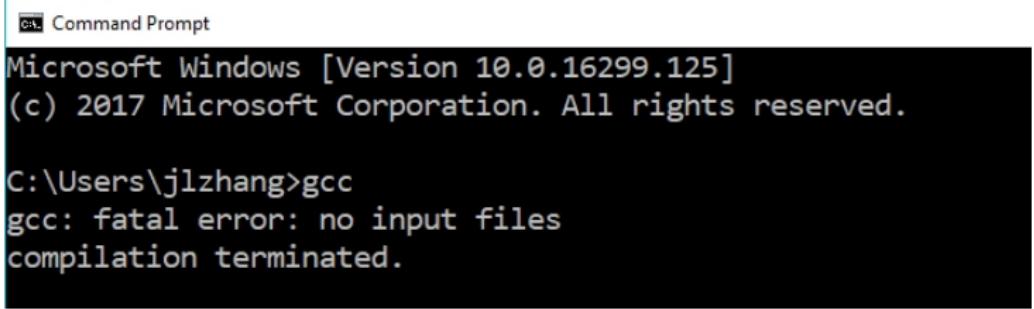
where 'command' is one of:
  INSTALL  Install add-on packages
  REMOVE   Remove add-on packages
  SHLIB    Make a DLL for use with dynload
  BATCH    Run R in batch mode
  build    Build add-on packages
  check    Check add-on packages
  Rprof    Post process R profiling files
  Rdconv   Convert Rd format to various other formats
  Rdiff    difference R output files
  Rd2pdf   Convert Rd format to PDF
  Rd2txt   Convert Rd format to pretty text
  Stangle   Extract S/R code from vignette
  Sweave   Process vignette documentation
  config   Obtain configuration information about R
  open     Open a file via Windows file associations
  texify   Process a latex file

Use
  Rcmd command --help
for usage information for each command.
```

图 34: 在 command prompt 中输入 Rcmd

## 附录：检查软件安装：Rtools

若 Rtools 已经安装好，则可在 command prompt 输入 `gcc` 检查



```
cmd Command Prompt
Microsoft Windows [Version 10.0.16299.125]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\jlzhang>gcc
gcc: fatal error: no input files
compilation terminated.
```

图 35: 在 command prompt 中输入 `gcc`

# 附录：检查软件安装：TinyTeX

若 TinyTeX 已经安装好，可在 command prompt 输入 pdflatex 检查 TeX 是否安装成功

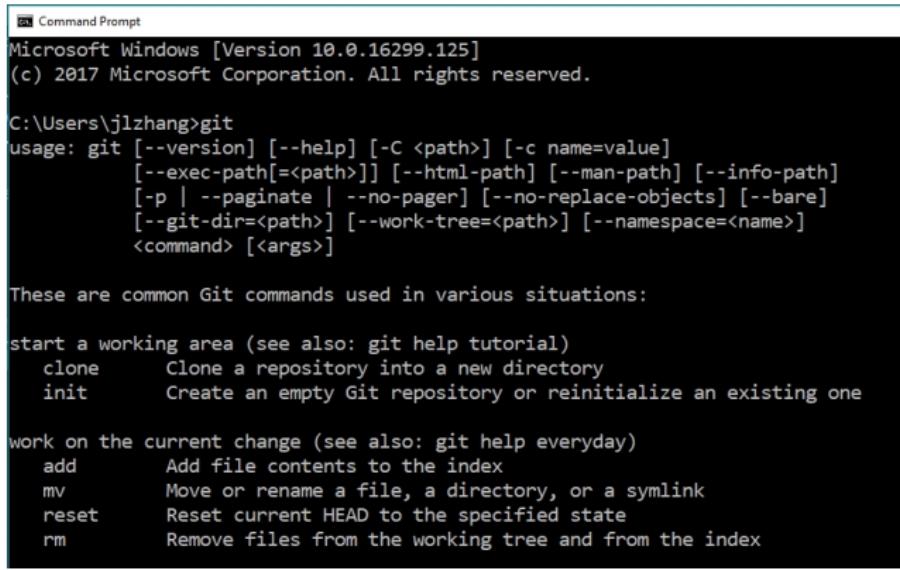


```
(base) Jinlongs-MBP:~ jinlong$ pdflatex
This is pdfTeX, Version 3.141592653-2.6-1.40.23 (TeX Live 2022/dev)
preloaded format=pdflatex
restricted \write18 enabled. 则应在command prompt输入pdflatex检查
** stalls
custom LaTeX
based on TeX
130
131 ! [在command prompt中输入pdflatex](images/)](width=80%)
```

图 36: 在 command prompt 中输入 pdflatex

# 附录：检查 git 是否已经安装好

可在 command prompt 输入 git 检查安装是否成功



```
Microsoft Windows [Version 10.0.16299.125]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\jlzhang>git
usage: git [--version] [--help] [-C <path>] [-c name=value]
           [--exec-path[=<path>]] [--html-path] [--man-path] [--info-path]
           [-p | --paginate | --no-pager] [--no-replace-objects] [--bare]
           [--git-dir=<path>] [--work-tree=<path>] [--namespace=<name>]
           <command> [<args>]

These are common Git commands used in various situations:

start a working area (see also: git help tutorial)
    clone      Clone a repository into a new directory
    init       Create an empty Git repository or reinitialize an existing one

work on the current change (see also: git help everyday)
    add        Add file contents to the index
    mv        Move or rename a file, a directory, or a symlink
    reset     Reset current HEAD to the specified state
    rm        Remove files from the working tree and from the index
```

图 37: 在 command prompt 中输入 git

# 附录：检查 Rstudio 是否安装成功？

- Windows: 开始
- MacOS: Launchpad