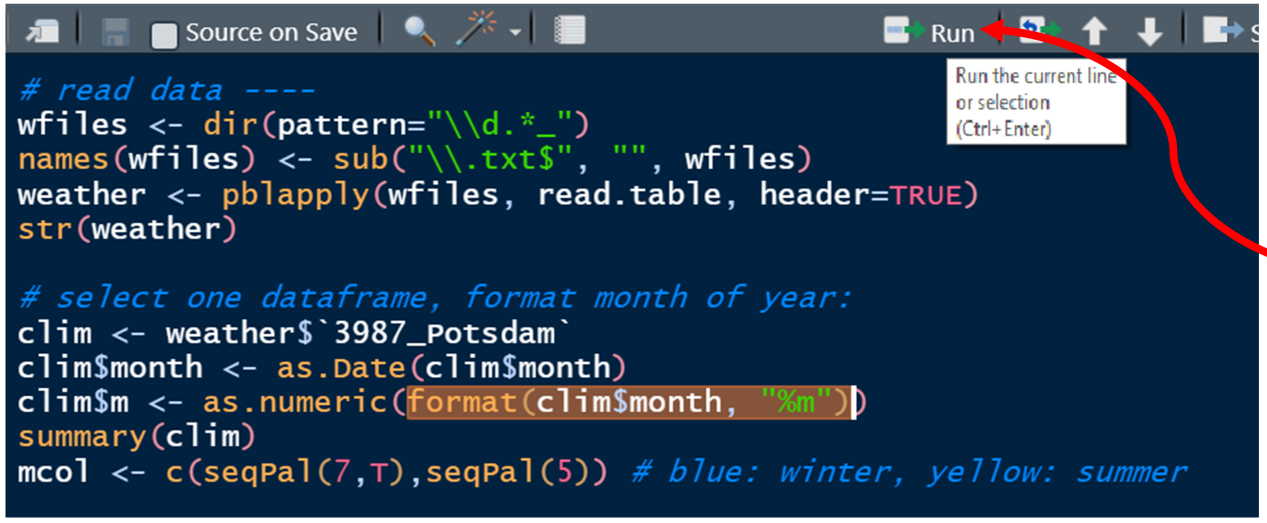
notebooks are awesome.
They are almost interactive!



统计软件: R 语言 101

安装, 第一行代码, Rmarkdown

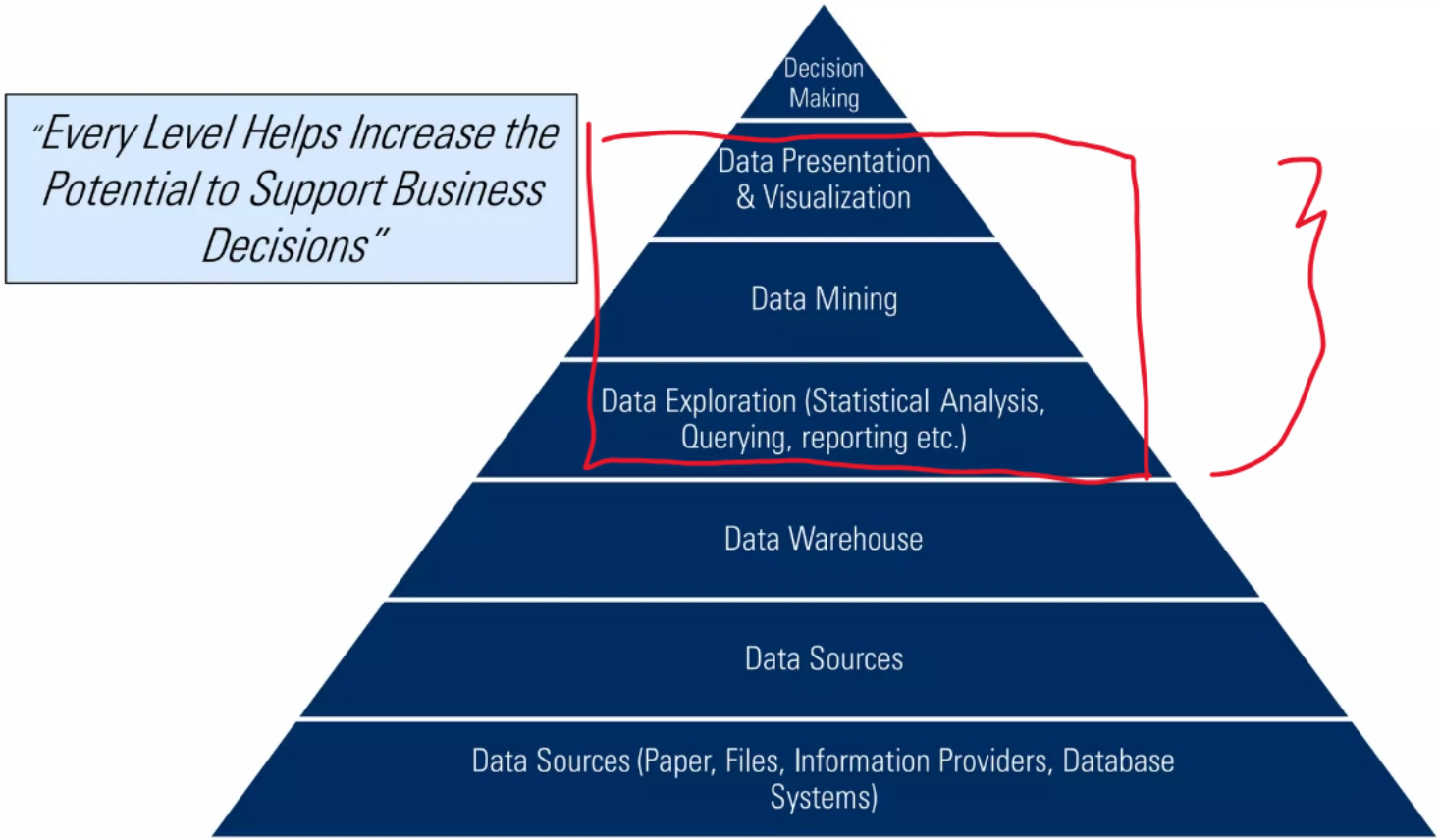
助教联系方式及课程群

- 姓名: 姜博恩
- 邮箱: bejiang24@m.fudan.edu.cn
- 答疑时间: 上机课后.

群組:统计软件25 春

此QR Code在7天内(2月27日前)有效,重新進入将更新。

为什么要学统计软件：多层模型



使用的工具

- R: 至少使用 3.6.0 版本以上, 目前最新版本为 4.4.2.

- Rstudio: R 的集成开发环境 (IDE), 你可以用它更方便地写 R, Python 等代码.

- 建议安装: Rtools, Rtools 是 R 的一个工具包, 用于编译 R 的包.

在 Windows x64 平台安装 R 和 Rtools

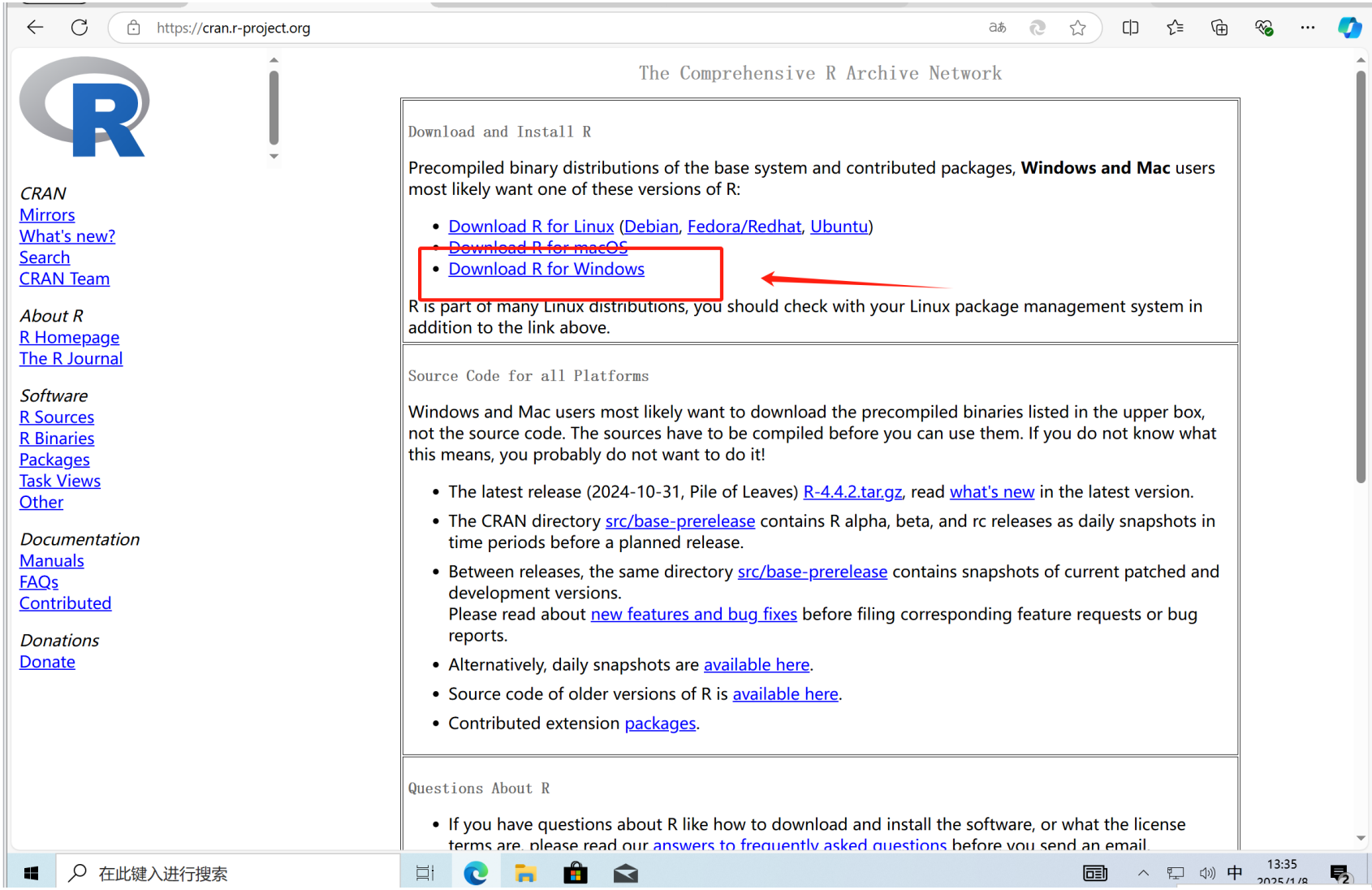
演示环境:

- Windows 10 x64,
- Mac 和 Linux 用户的安装方式也是类似的,
- Windows on ARM 需要至少 R 4.4.0 版本. 但是 Rstudio 目前不支持 ARM 平台.

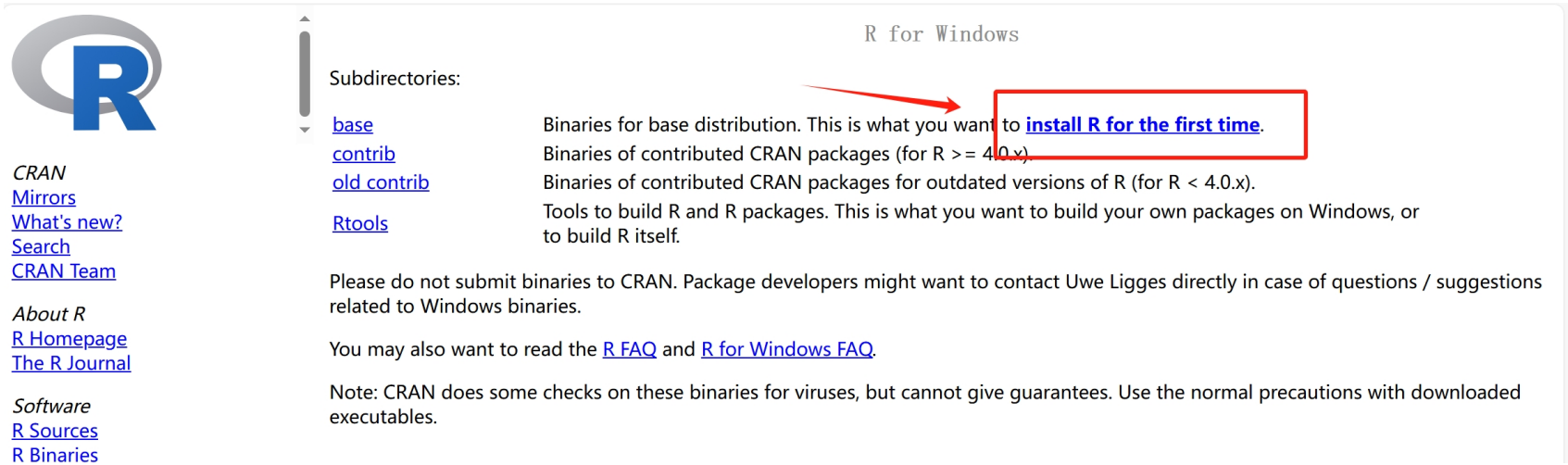
CRAN(Comprehensive R Archive Network)

- 包含最新和以前版本的 R 发行版, 文档和 R 包的存档,
- CRAN 由 Kurt Hornik 和 Friedrich Leisch 于 1997 年创建,
- `https://cran.r-project.org/`.
- 也可以通过 Rstudio 跳转到 CRAN, 注意在使用搜索引擎时注意下载的是否是广告, 可以使用 Bing 国际版等搜索引擎.

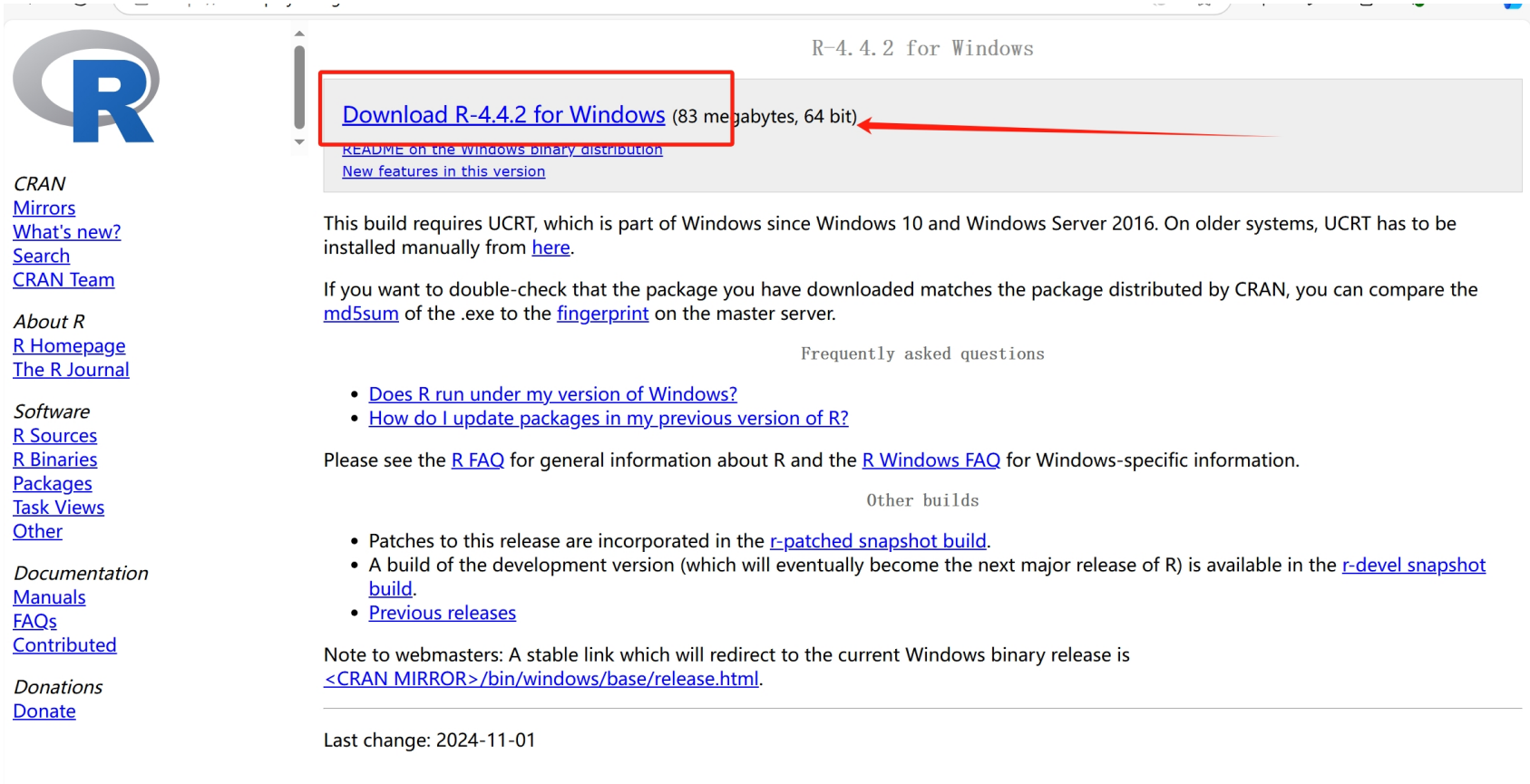
CRAN: 下载 R


https://cran.r-project.org
The Comprehensive R Archive Network
CRAN
Mirrors
What's new?
Search
CRAN Team
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Task Views
Other
Documentation
Manuals
FAQs
Contributed
Donations
Donate
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 (Debian, Fedora/Redhat, Ubuntu)
Download R for macOS
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 (2024-10-31, Pile of Leaves) R-4.4.2.tar.gz, read what's new in the latest version.
The CRAN directory src/base-prerelease contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
Between releases, the same directory src/base-prerelease contains snapshots of current patched and development versions.
Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
Alternatively, daily snapshots are available here.
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.
在此键入进行搜索
13:35

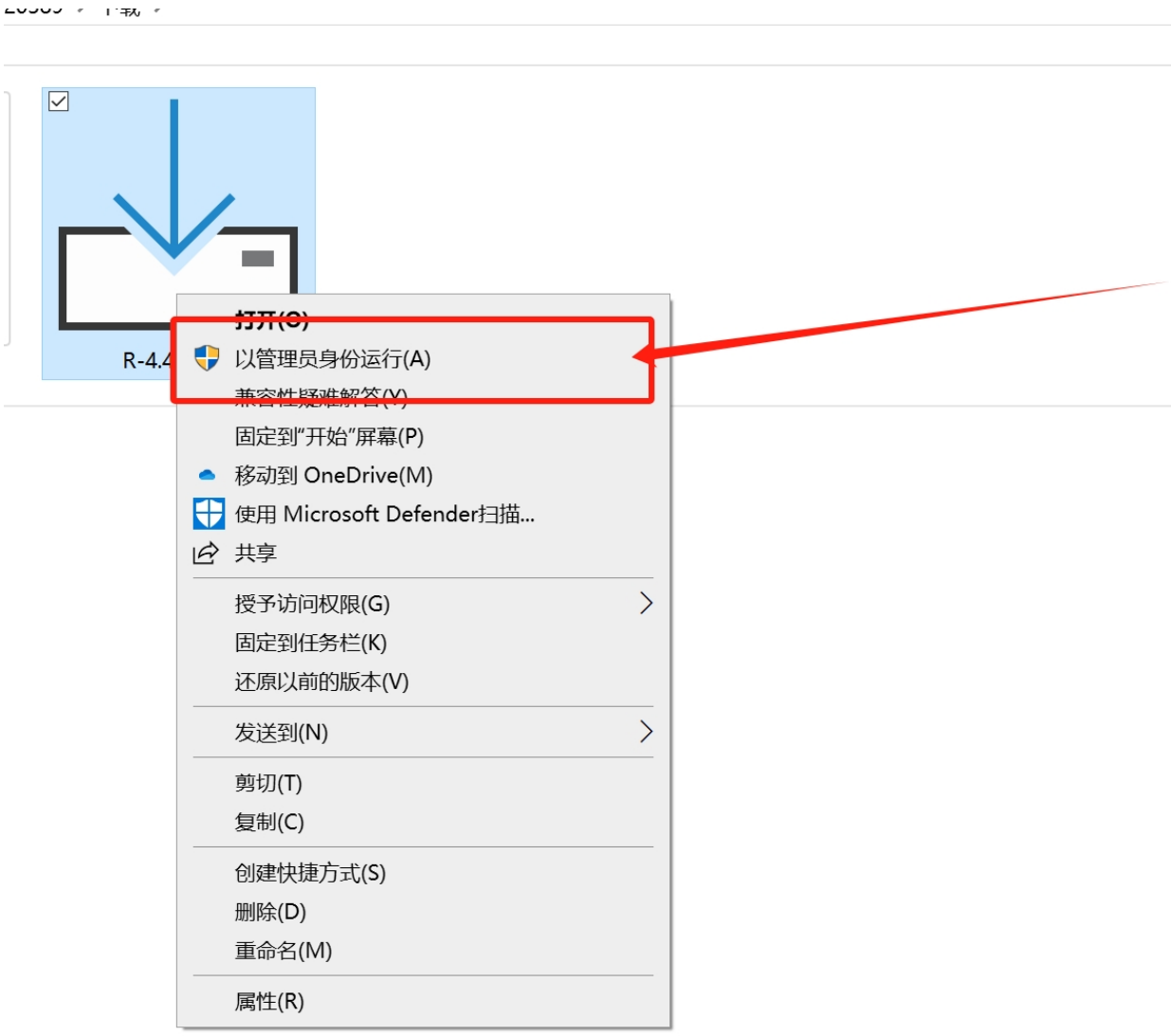
CRAN: 下载 R


CRAN
Mirrors
What's new?
Search
CRAN Team
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
R for Windows
Subdirectories:
base
Binaries for base distribution. This is what you want to install R for the first time.
contrib
Binaries of contributed CRAN packages (for R >= 4.0.x).
old contrib
Binaries of contributed CRAN packages for outdated versions of R (for R < 4.0.x).
Rtools
Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.
Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.
You may also want to read the R FAQ and R for Windows FAQ.
Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

CRAN: 下载 R


CRAN
Mirrors
What's new?
Search
CRAN Team
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Task Views
Other
Documentation
Manuals
FAQs
Contributed
Donations
Donate
R-4.4.2 for Windows
Download R-4.4.2 for Windows (83 megabytes, 64 bit)
README on the Windows binary distribution
New features in this version
This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from here.
If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server.
Frequently asked questions
Does R run under my version of Windows?
How do I update packages in my previous version of R?
Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.
Other builds
Patches to this release are incorporated in the r-patched snapshot build.
A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
Previous releases
Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.html.
Last change: 2024-11-01

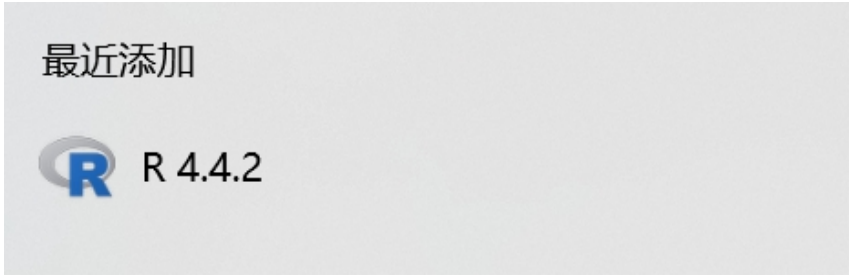
安装 R 4.4.2



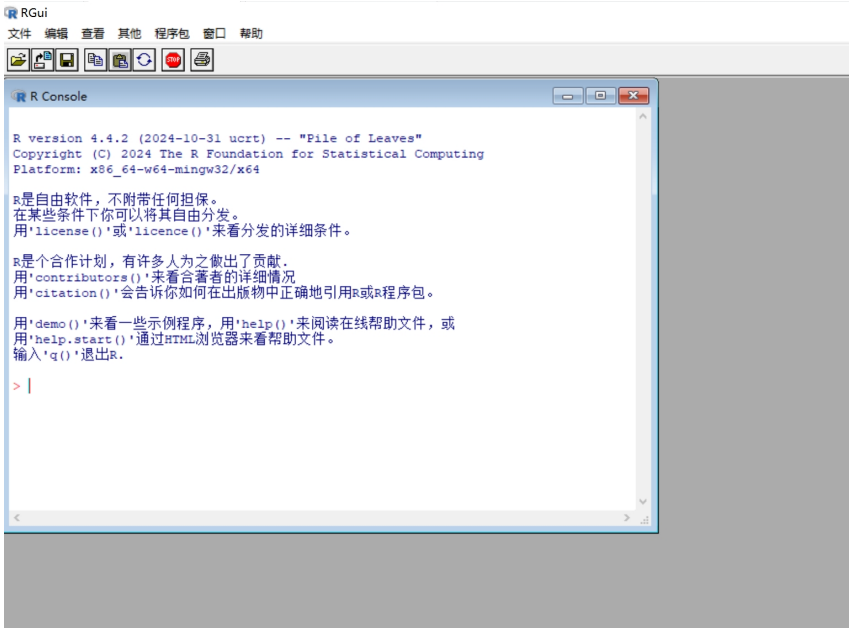
接下来是标准的软件安装流程，一路下一步即可．安装路径可以自己按照喜好选择.

运行 R 4.4.2

安装好后可以发现:

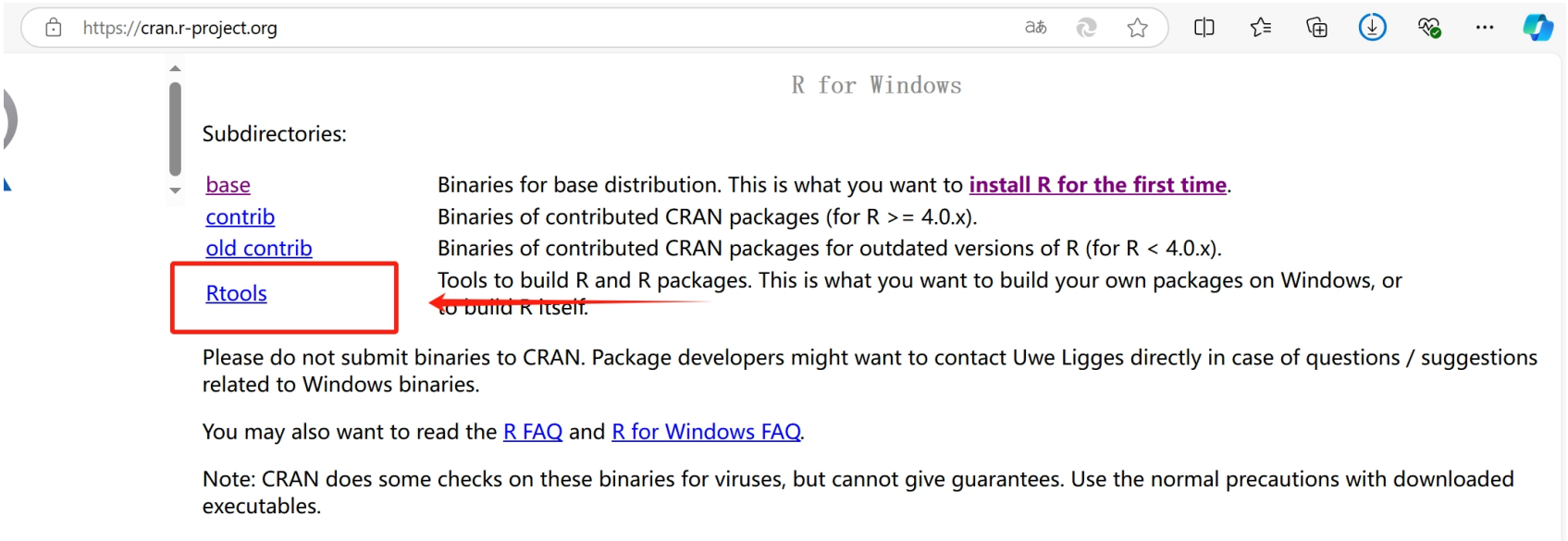


我们运行 R, 会看到如下界面:



就安装好了 R 语言本体.

CRAN: 下载 Rtools



选择自己合适的版本

CRAN: 下载 Rtools

Rtools44 for Windows

Rtools is a toolchain bundle used for building R packages from source (those that need compilation of C/C++ or Fortran code) and for building R itself. Rtools44 is currently used for R 4.4 and R-devel, the development version of R, to become R 4.5.0.

Rtools44 consists of Msys2 build tools, GCC 13/MinGW-w64 compiler toolchain, libraries built using the toolchain, and QPDF. Rtools44 supports 64-bit Windows and UCRT as the C runtime.

Compared to Rtools43, Rtools44 for 64-bit Intel machines has newer versions of three core components: GCC, MinGW-w64, and binutils. It is therefore recommended to re-compile all code with the new toolchain to avoid problems. The code compiled by Rtools older than Rtools42 is incompatible due to use of MSVCRT and has to be recompiled with Rtools44 for use in R packages.

Rtools44 is also available for 64-bit ARM machines (aarch64): it includes Msys2 build tools (64-bit Intel builds running via emulation) and aarch64 builds of LLVM 17/MinGW-w64 compiler toolchain, libraries built using the toolchain, and again QPDF. The 64-bit ARM version of Rtools44 is experimental: a number of CRAN packages don't work with it and the Fortran compiler (flang-new) is not yet able to compile Fortran code of all CRAN packages. A number of CRAN packages doesn't work because they require not-yet-available 64-bit ARM versions of external software.

Installing Rtools44

Rtools is only needed for installation of R packages from source (those that need compilation of C/C++ of Fortran code) or building R from source. R can be installed from the R binary installer and by default will install binary versions of CRAN packages, which does not require Rtools44.

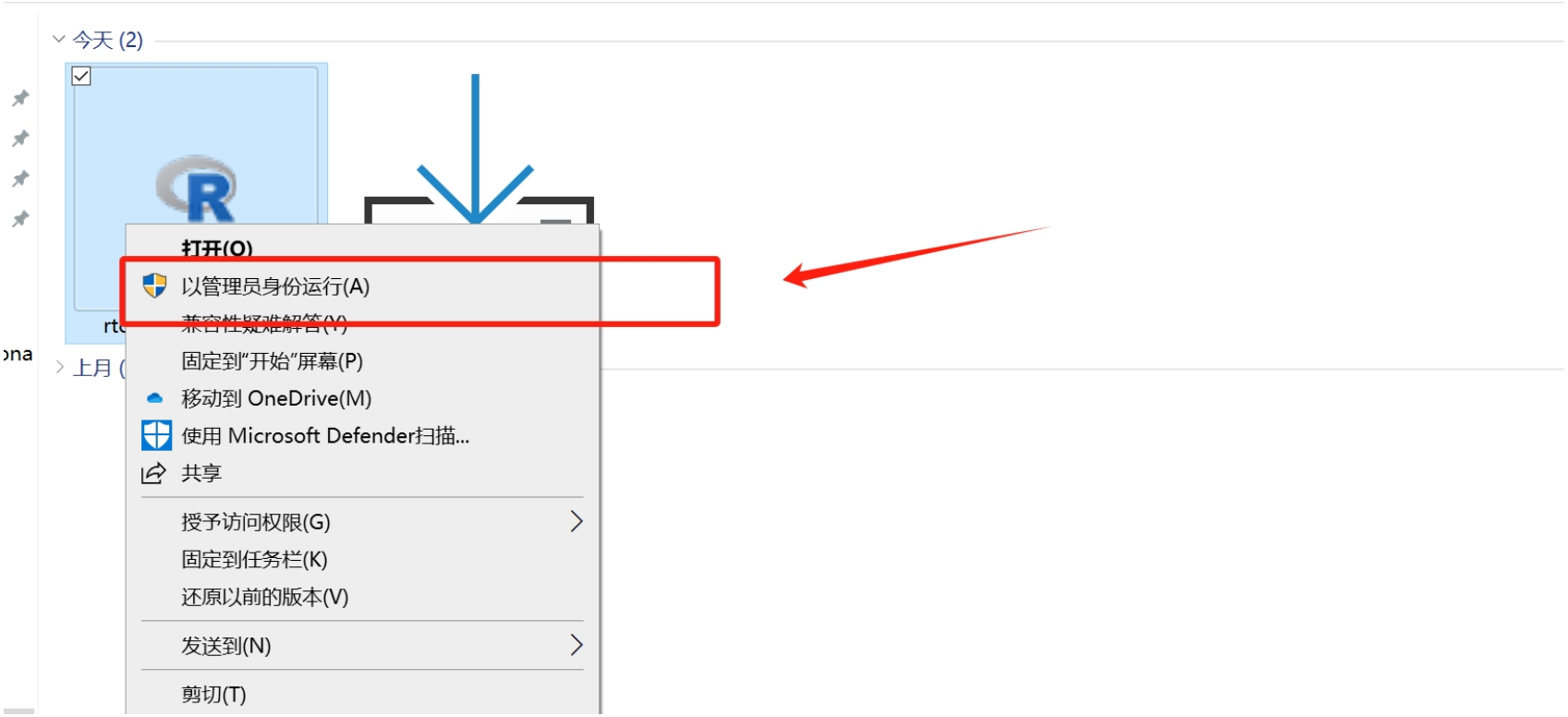
Moreover, online build services are available to check and build R packages for Windows, for which again one does not need to install Rtools44 locally. The Winbuilder check service uses identical setup as the CRAN incoming packages checks and has already all CRAN and Bioconductor packages pre-installed.

Rtools44 may be installed from the Rtools44 installer or 64-bit ARM Rtools44 installer. It is recommended to use the defaults, including the default installation location of `C:\rtools44`.

When using R installed by the installer, no further setup is necessary after installing Rtools44 to build R packages from source. When using the default installation location, R and Rtools44 may be installed in any order and Rtools44 may be installed when R is already running.

On ARM, binary versions of packages are currently not available from CRAN, so Rtools44 is required to install any package that needs compilation.

安装 Rtools

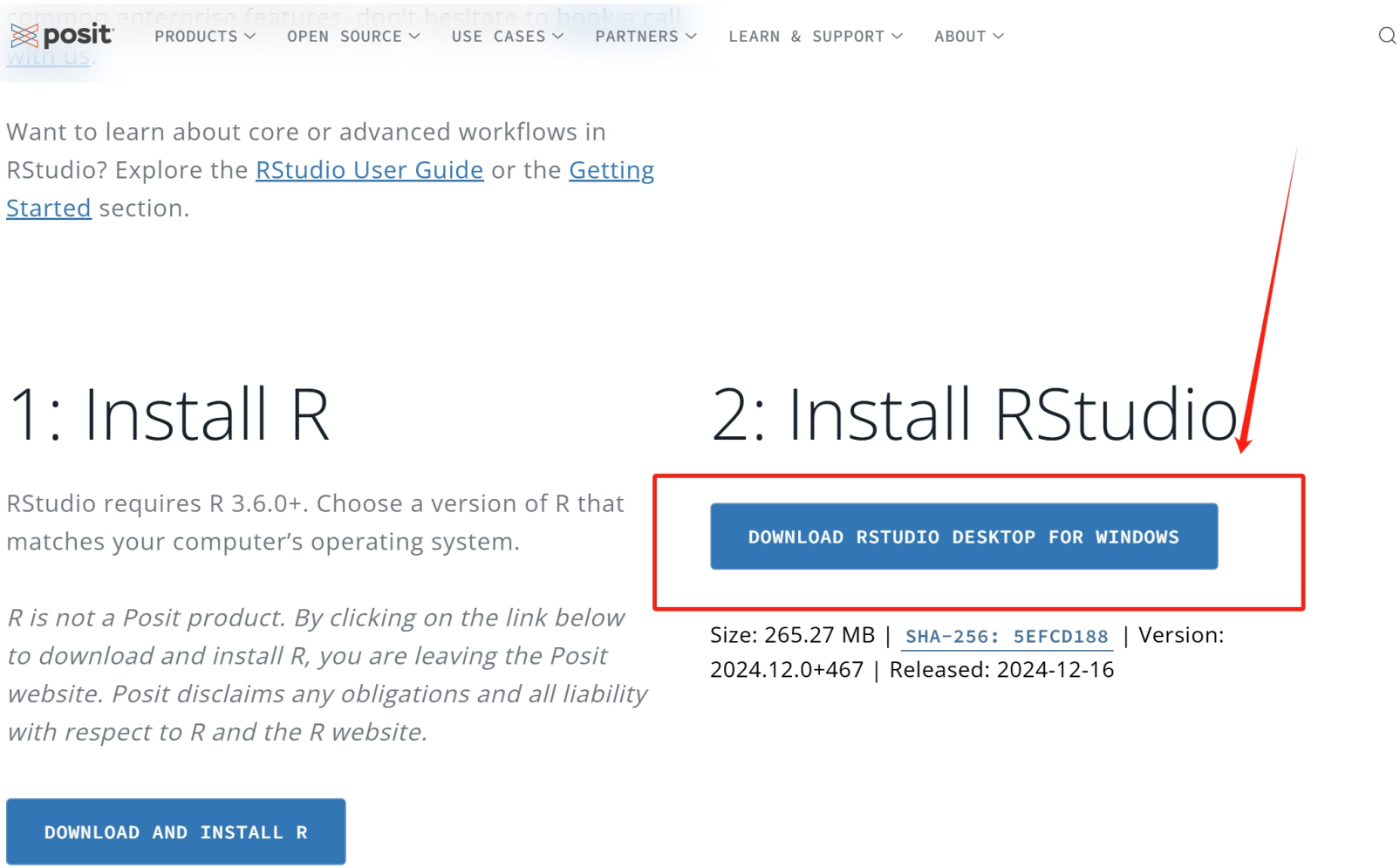


接下来是标准的软件安装流程，一路下一步即可．安装路径可以自己按照喜好选择．

Posit: 下载 Rstudio

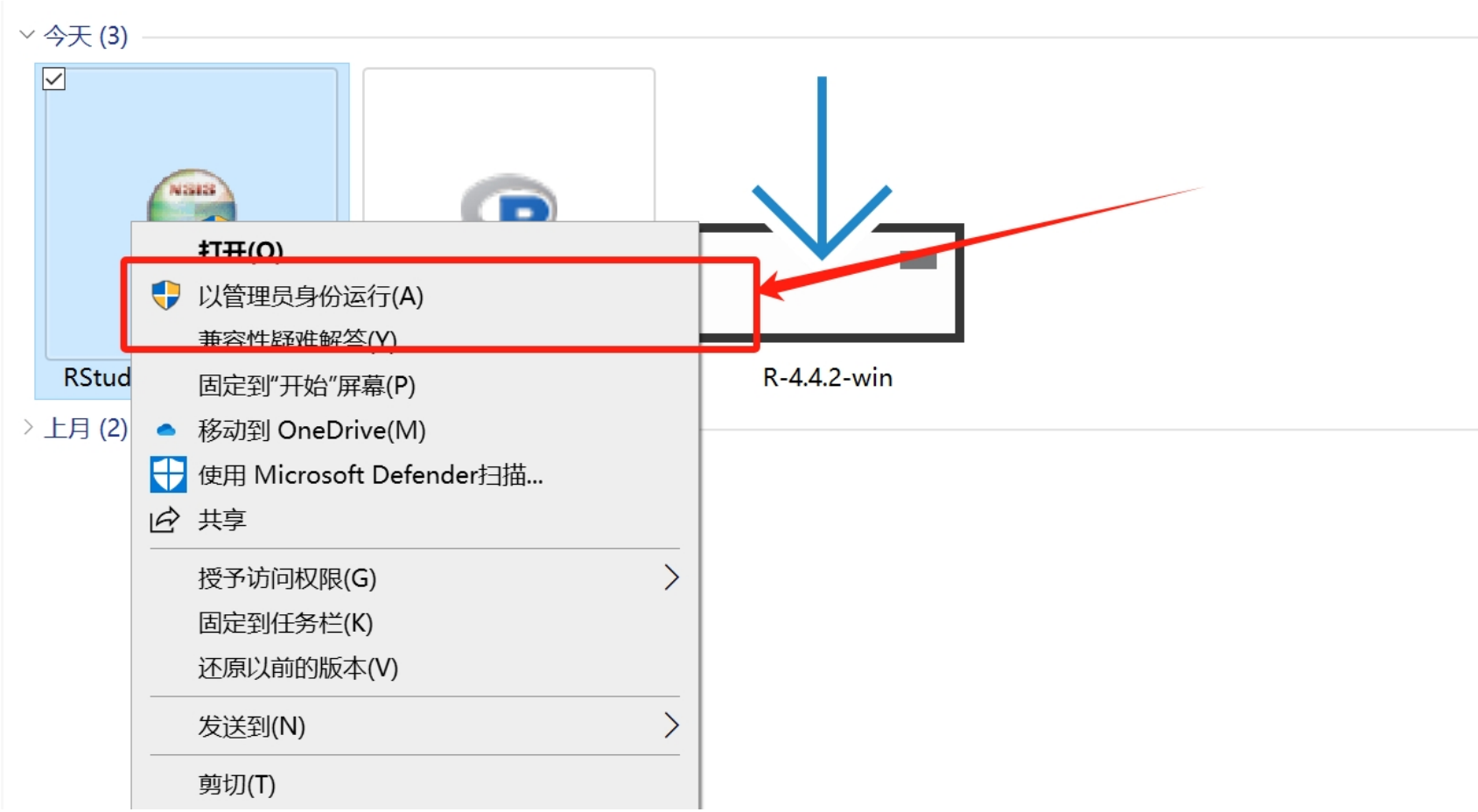
- Rstudio 在 2022 年将公司名从 Rstudio Inc. 改为 Posit PBC.
- 下载 Rstudio desktop:
  `https://posit.co/download/rstudio-desktop/`

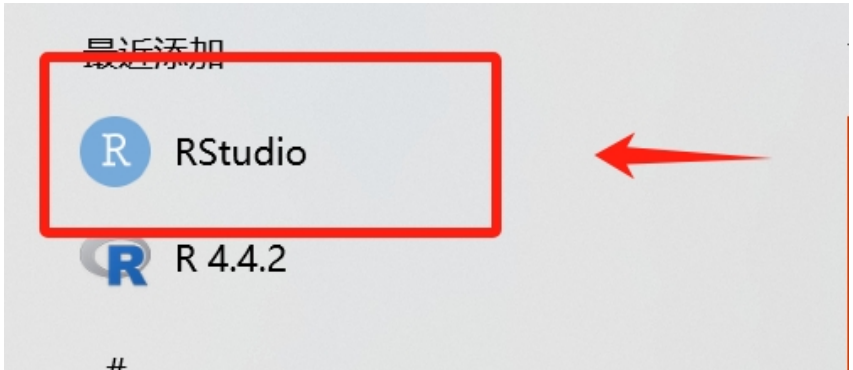
下载 Rstudio



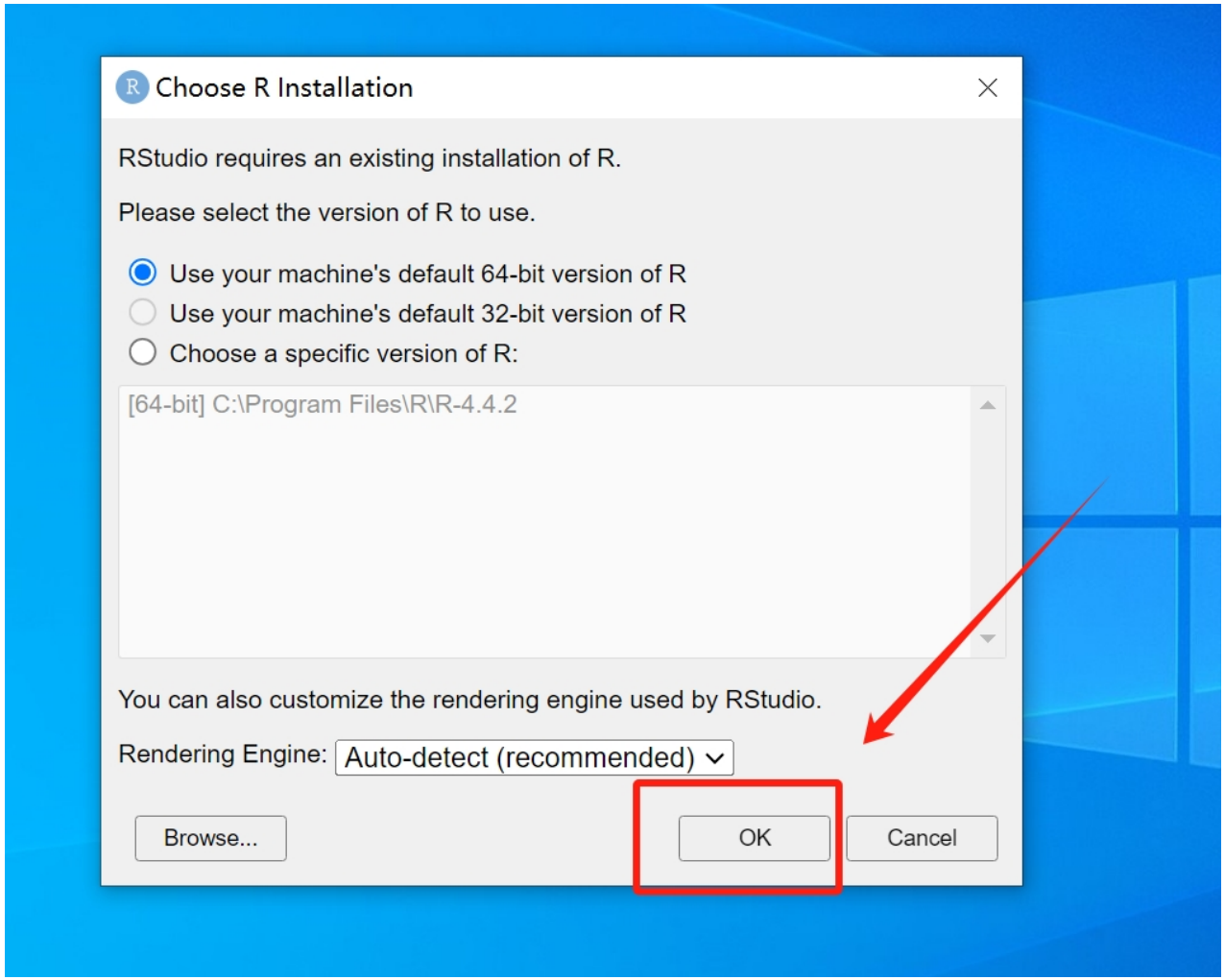
可以看到之前我们已经完成了第一步，这里我们直接下载 Rstudio Desktop 即可.

安装 Rstudio


今天 (3)
RStud
R-4.4.2-win
上月 (2)
打开(O)
以管理员身份运行(A)
兼容性疑难解答(Y)
固定到"开始"屏幕(P)
移动到 OneDrive(M)
使用 Microsoft Defender扫描...
共享
授予访问权限(G)
固定到任务栏(K)
还原以前的版本(V)
发送到(N)
剪切(T)

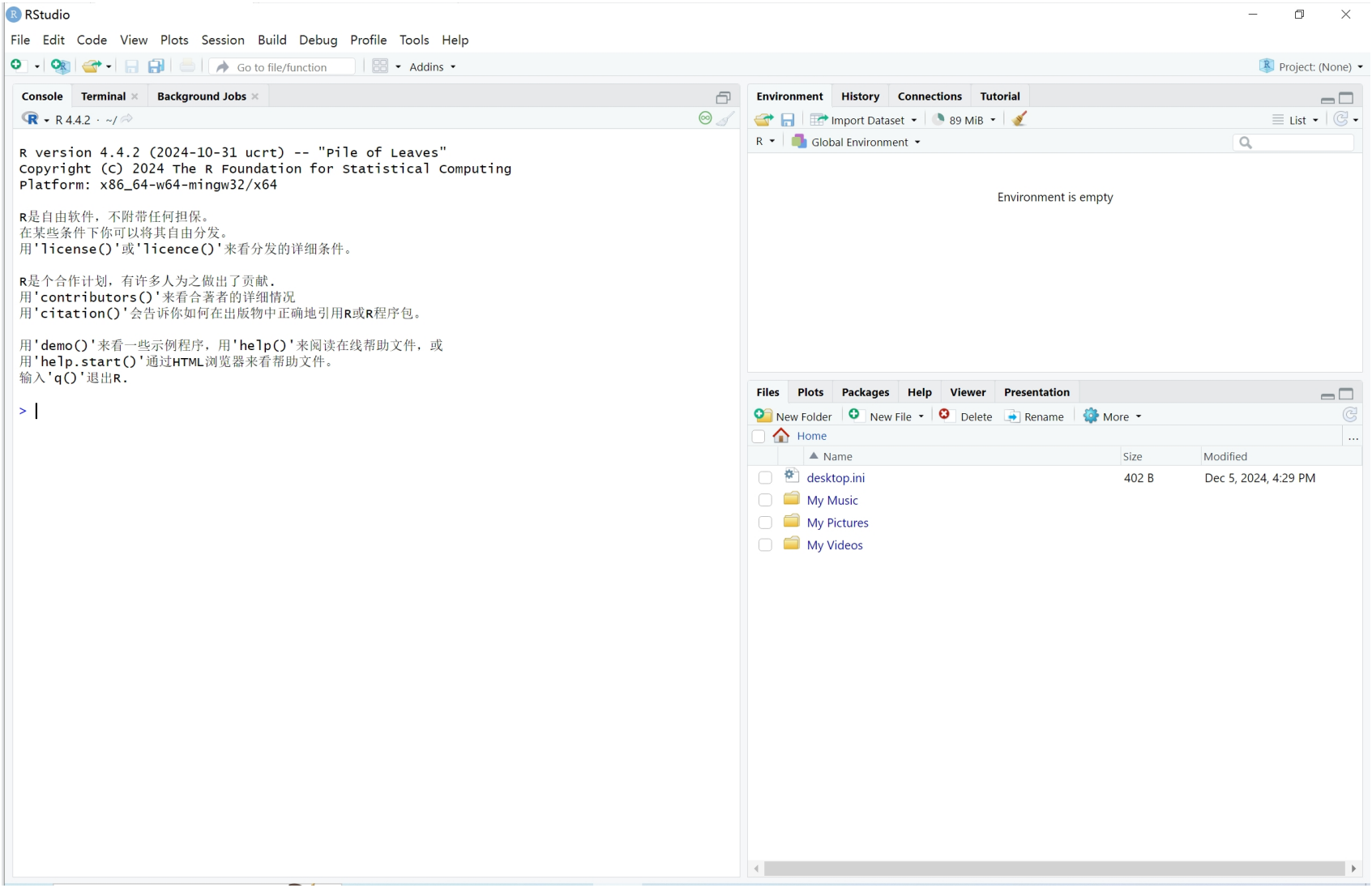

RStudio
R 4.4.2

运行 Rstudio



安装部分至此结束.

Rstudio 界面



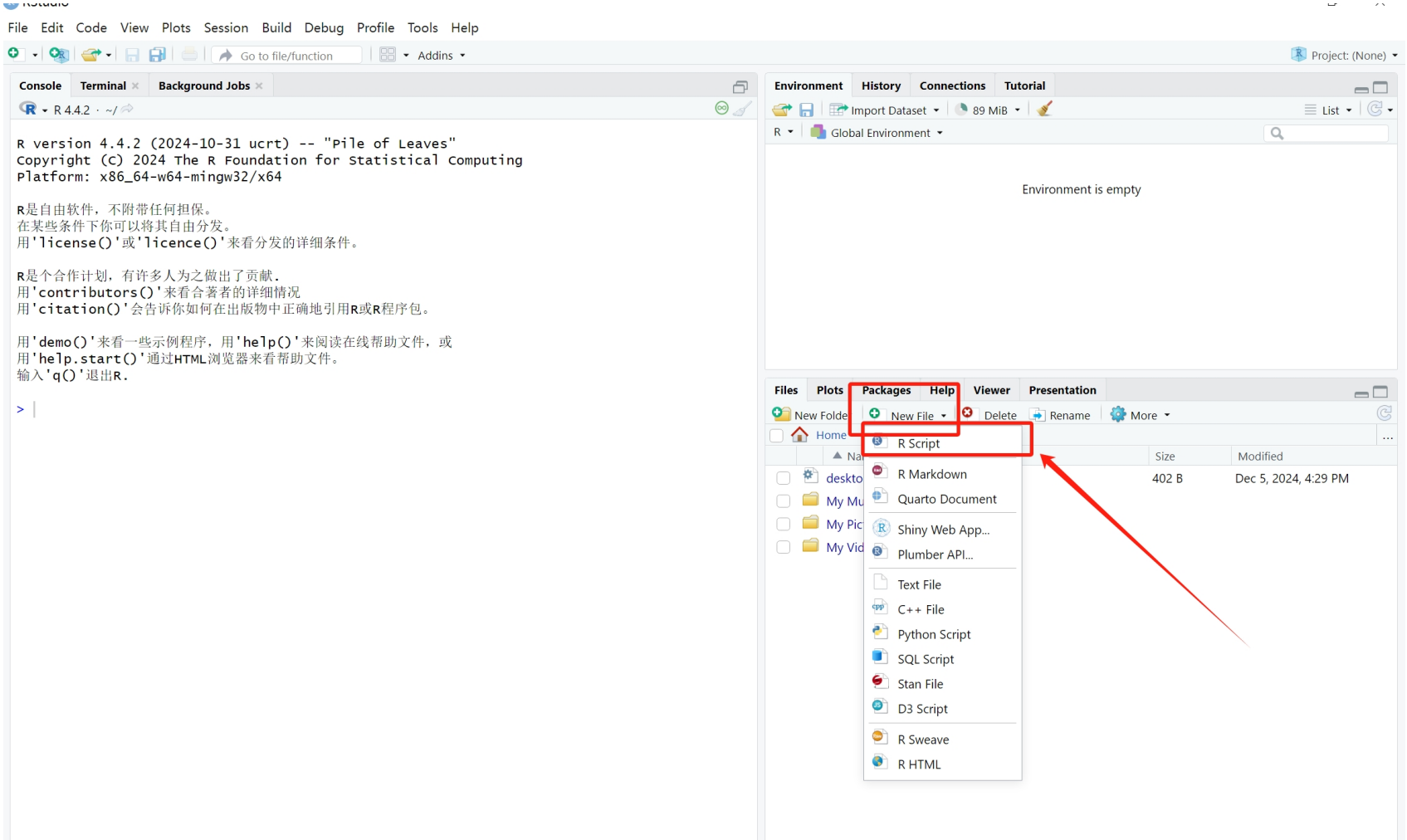
- "Pile of Leaves": R 语言的版本昵称 (nickname).
- 这个昵称没有规律: 来源于 Peanuts cartoon.

Rstudio 界面

常用的几个区域:

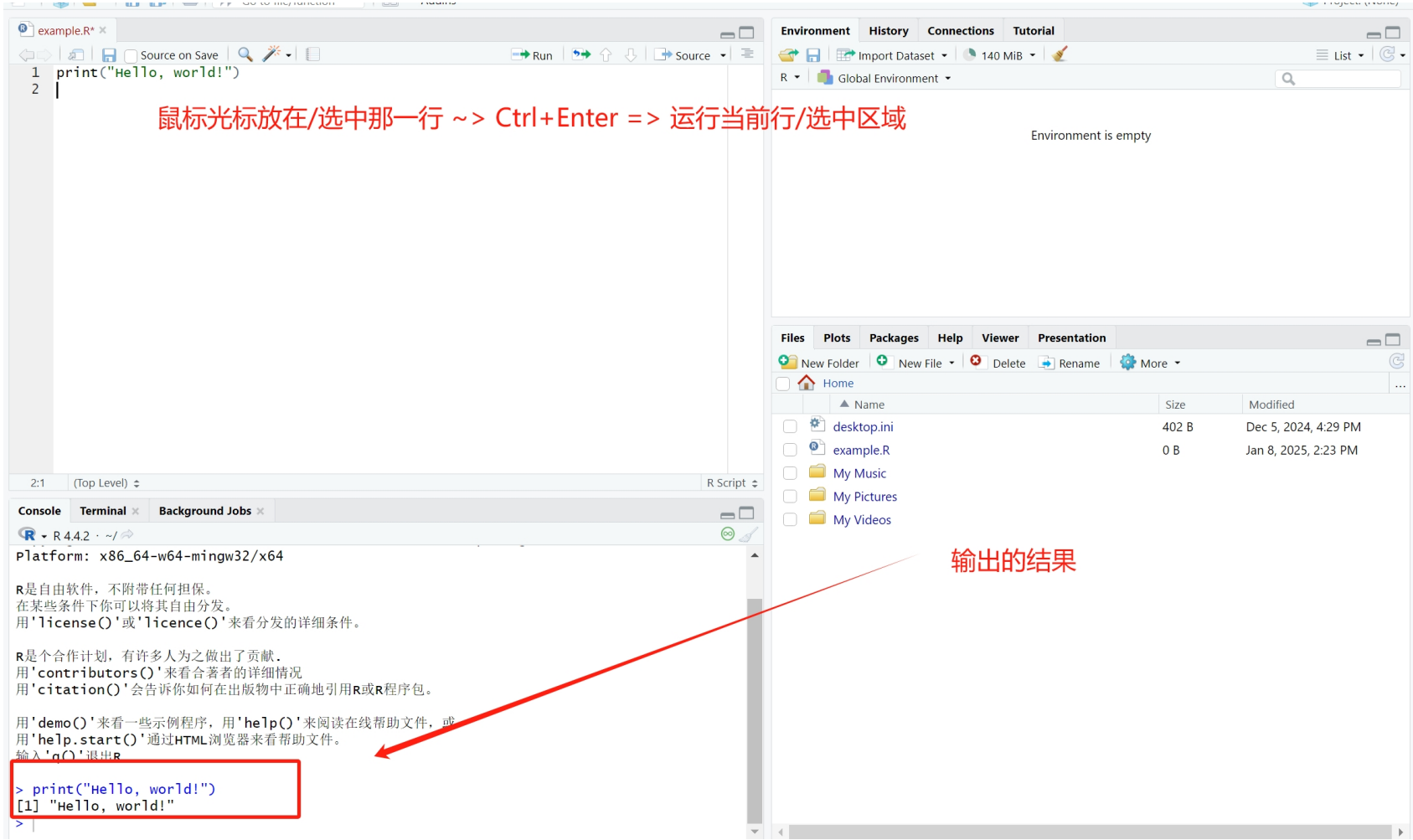
- Console: 控制台. 交互式界面, 可以直接输入 R 代码.
- Teminal: 终端. 可以直接输入系统命令, 等同于 windows 的 cmd.
- Environment: 环境. 显示当前环境的变量, 函数等.
- Files: 文件. 显示当前工作目录的文件.
- Plots: 图形. 显示绘制的图形.
- Packages: 包. 显示已经安装的包.

创建一个 R 语言脚本



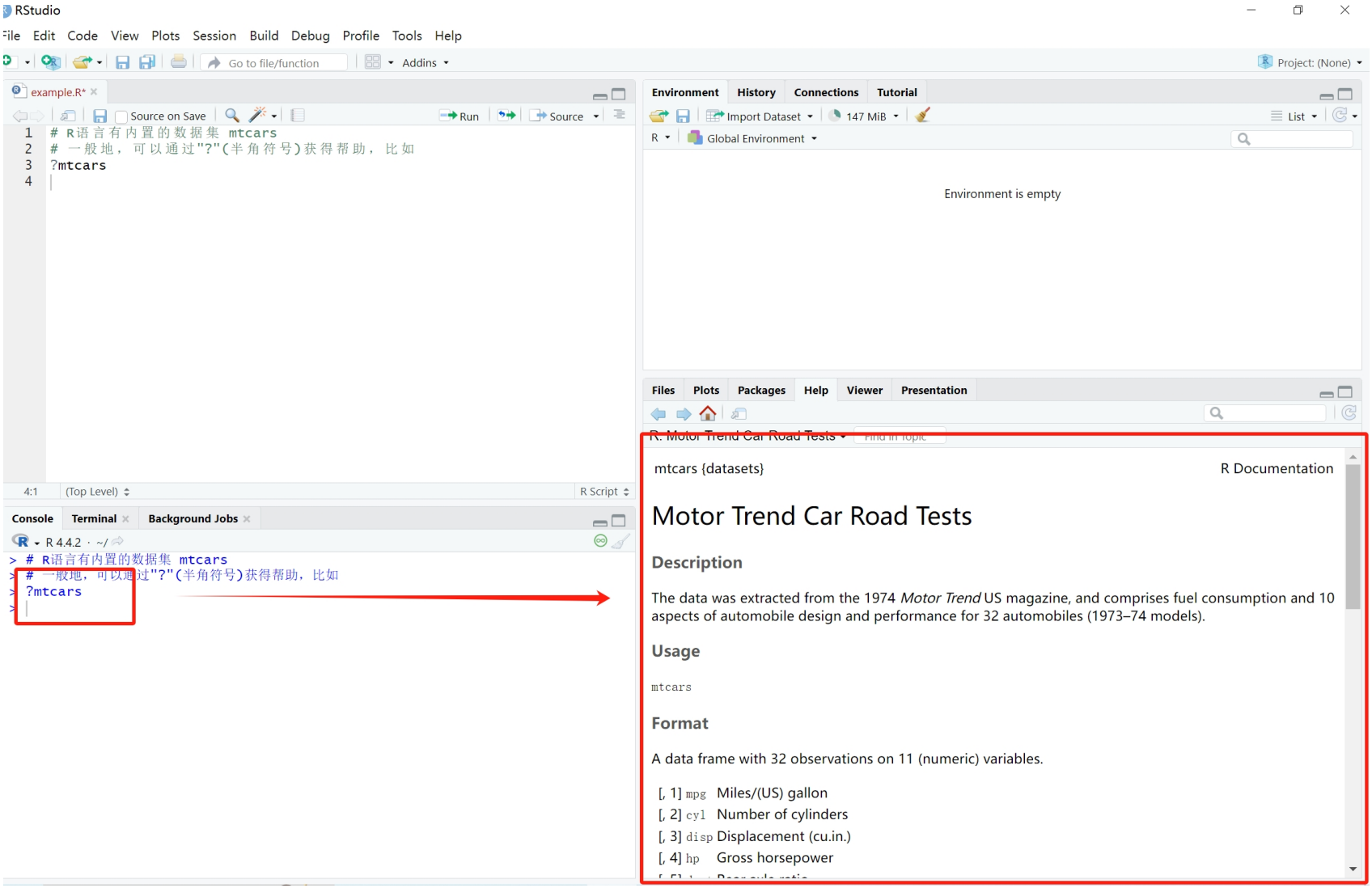
- 可以看到，除了 R 脚本，还有 Rmarkdown，Python，SQL，C++ 等.

Hello, World!



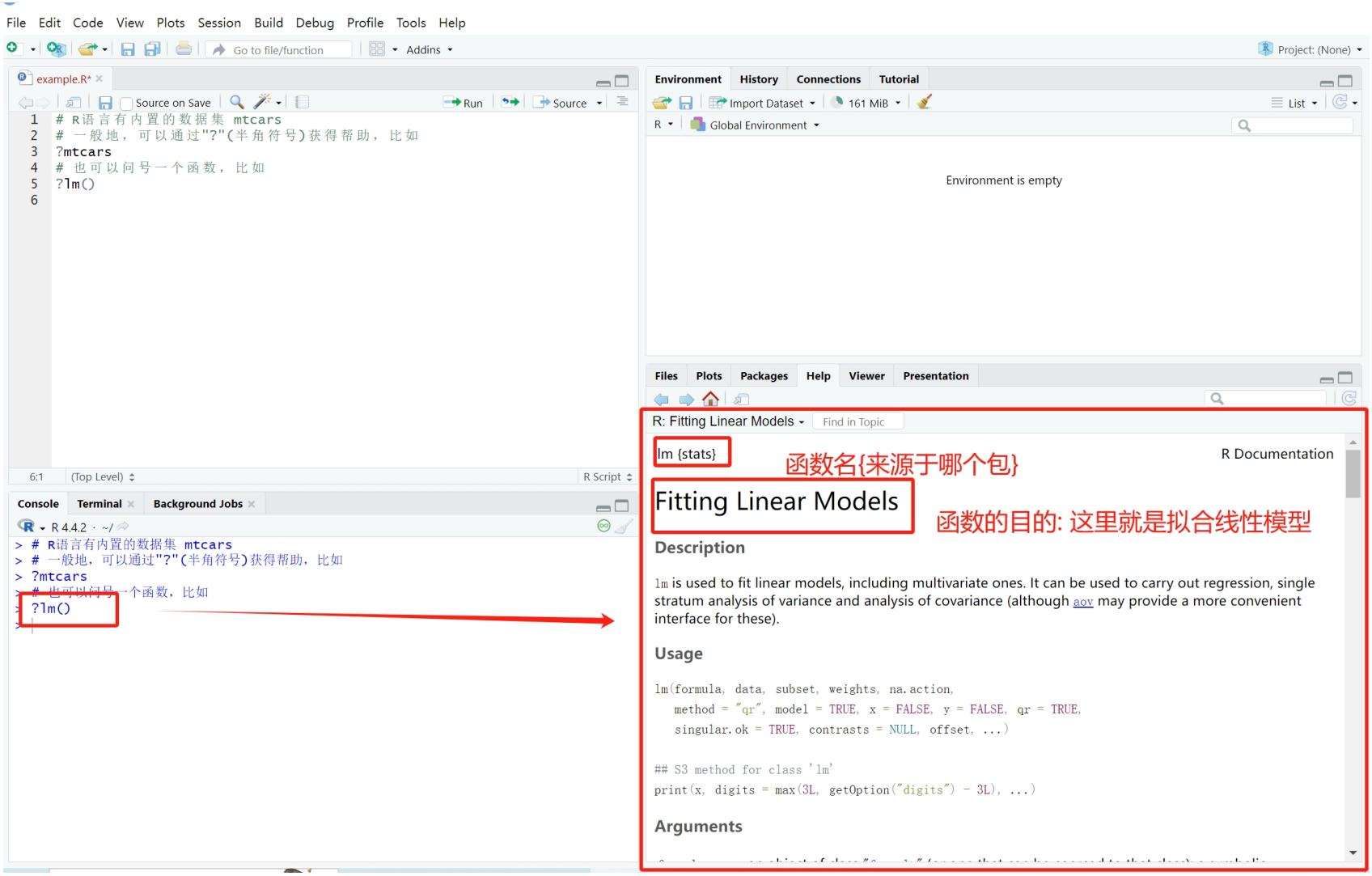
- 能不能直接在 console 里面输入？试试看.
- R 语言是大小写敏感的. 试试看 `print("Hello, World!")` 和 `Print("Hello, World!")` 有什么区别?

如何获得帮助



- 可以看到 # 号后面的内容是注释，不会被执行.
- 试一试 `help(mtcars)`, `help.search(mtcars)`, `help.search("mtcars")`, `?"mtcars"` 哪些可以执行?

如何获得帮助



- 试试看不加括号.
- 要善于使用 Tab 键自动补全.

一个线性模型

example.R*

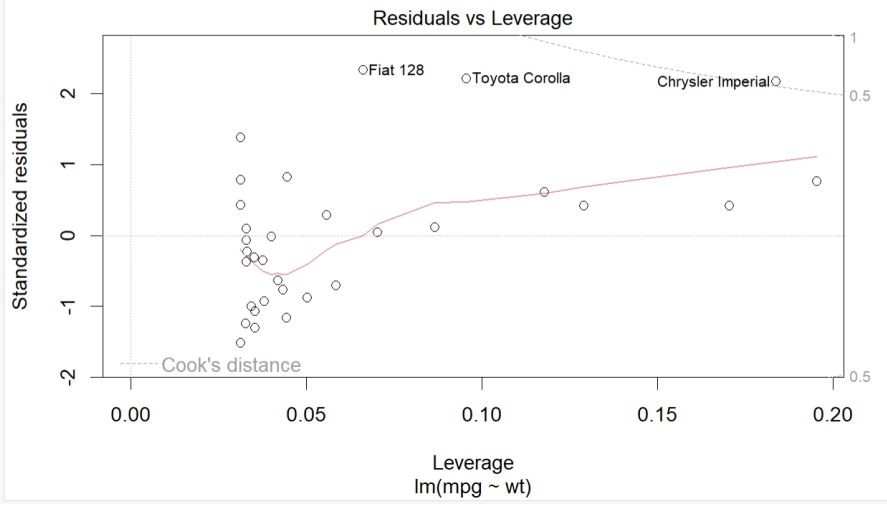
```
1  # R语言有内置的数据集 mtcars
2  # 一般地，可以通过"?"(半角符号)获得帮助，比如
3  ?mtcars
4  # 也可以问号一个函数，比如
5  ?lm()
6  # 线性模型
7  lmfit = lm(mpg~wt, data = mtcars)
8  plot(lmfit)
9
```

Environment | History | Connections | Tutorial

Data

lmfit | List of 12

Files | Plots | Packages | Help | Viewer | Presentation



Console | Terminal | Background Jobs

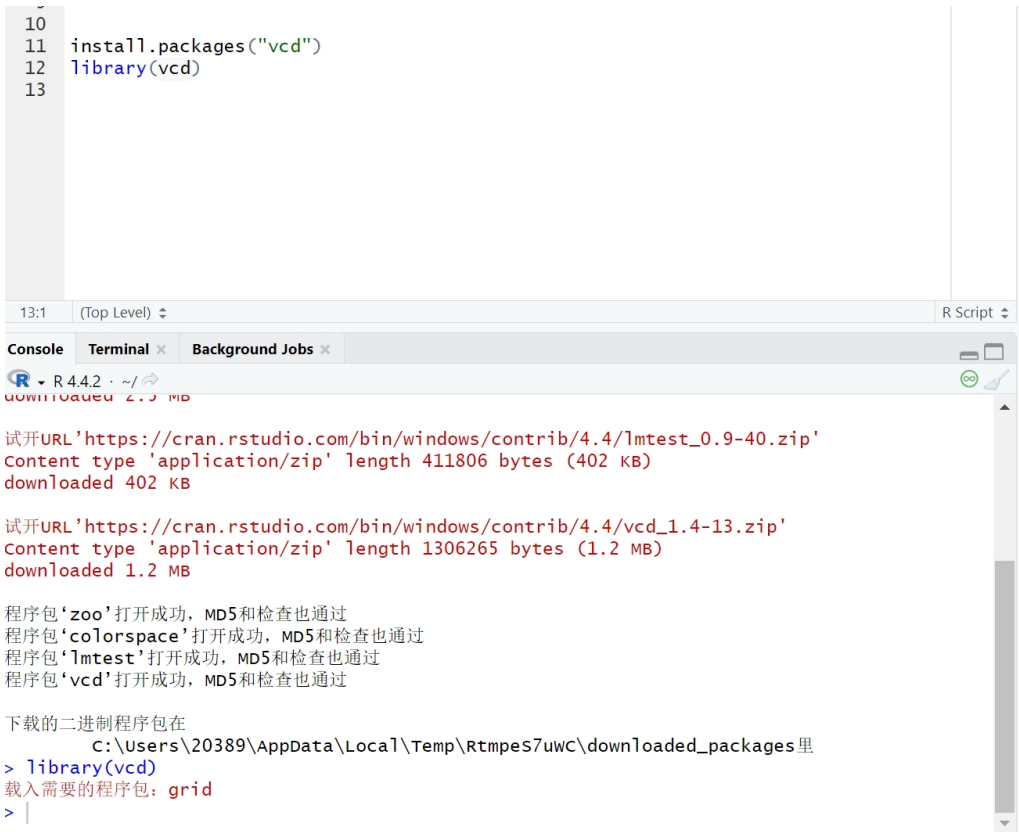
R 4.4.2 · ~/

```
> # R语言有内置的数据集 mtcars
> # 一般地，可以通过"?"(半角符号)获得帮助，比如
> ?mtcars
> # 也可以问号一个函数，比如
> ?lm()
> # 线性模型
> lmfit = lm(mpg~wt, data = mtcars)
> plot(lmfit)
按<Return>键来看下一个图:
按<Return>键来看下一个图:
按<Return>键来看下一个图:
按<Return>键来看下一个图:
>
```

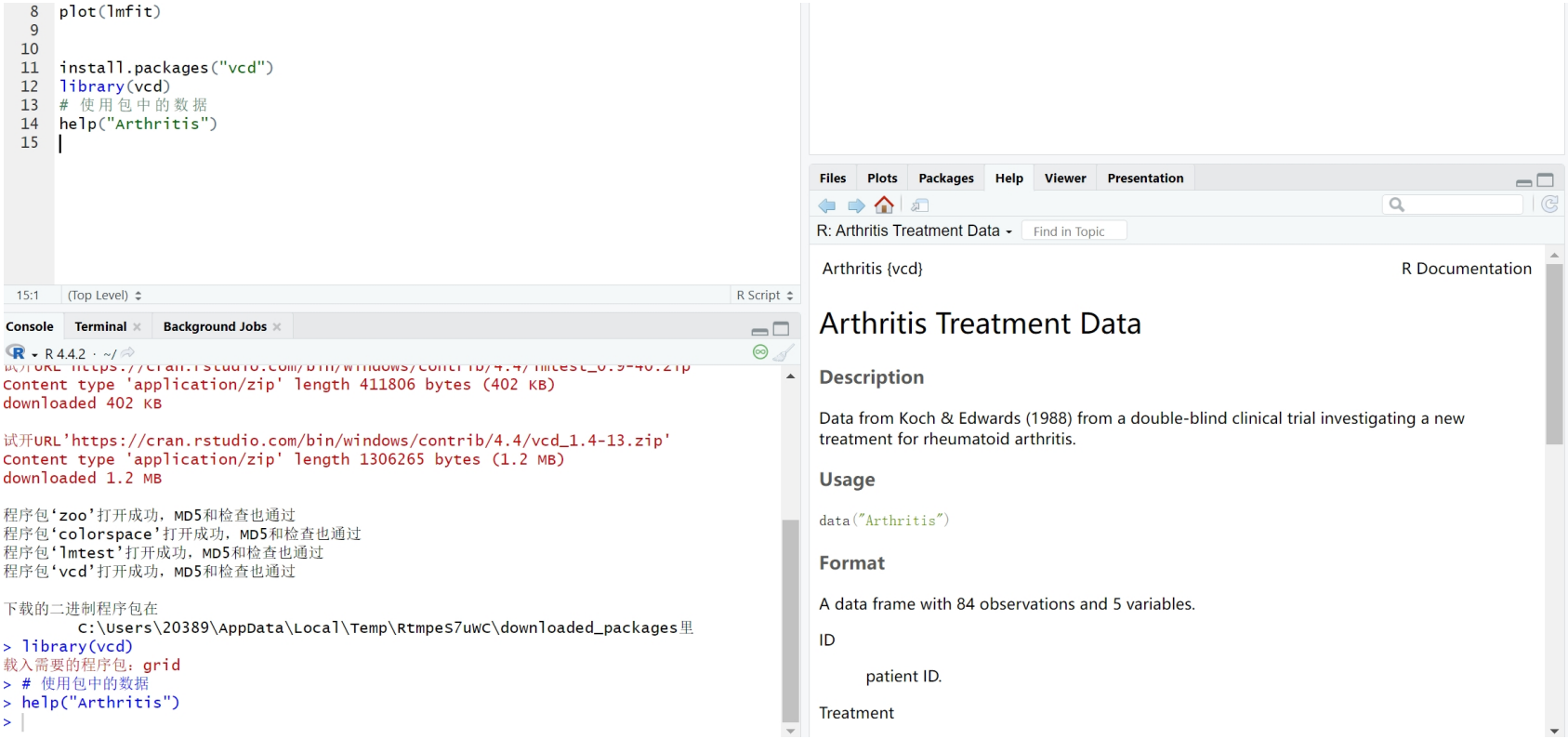
- 试试看把“=”改成“<-”，能不能运行?
- 试试 `lm(mpg wt, data = mtcars) -> lmfit`，能运行吗?

使用一个包

- 类似于 Python 的 pip install 和 from import.
- 试试看 `install.packages("vcd")`, 然后 `library(vcd)`.
- `install.packages(vcd)` 可以吗? `library("vcd")` 可以吗?

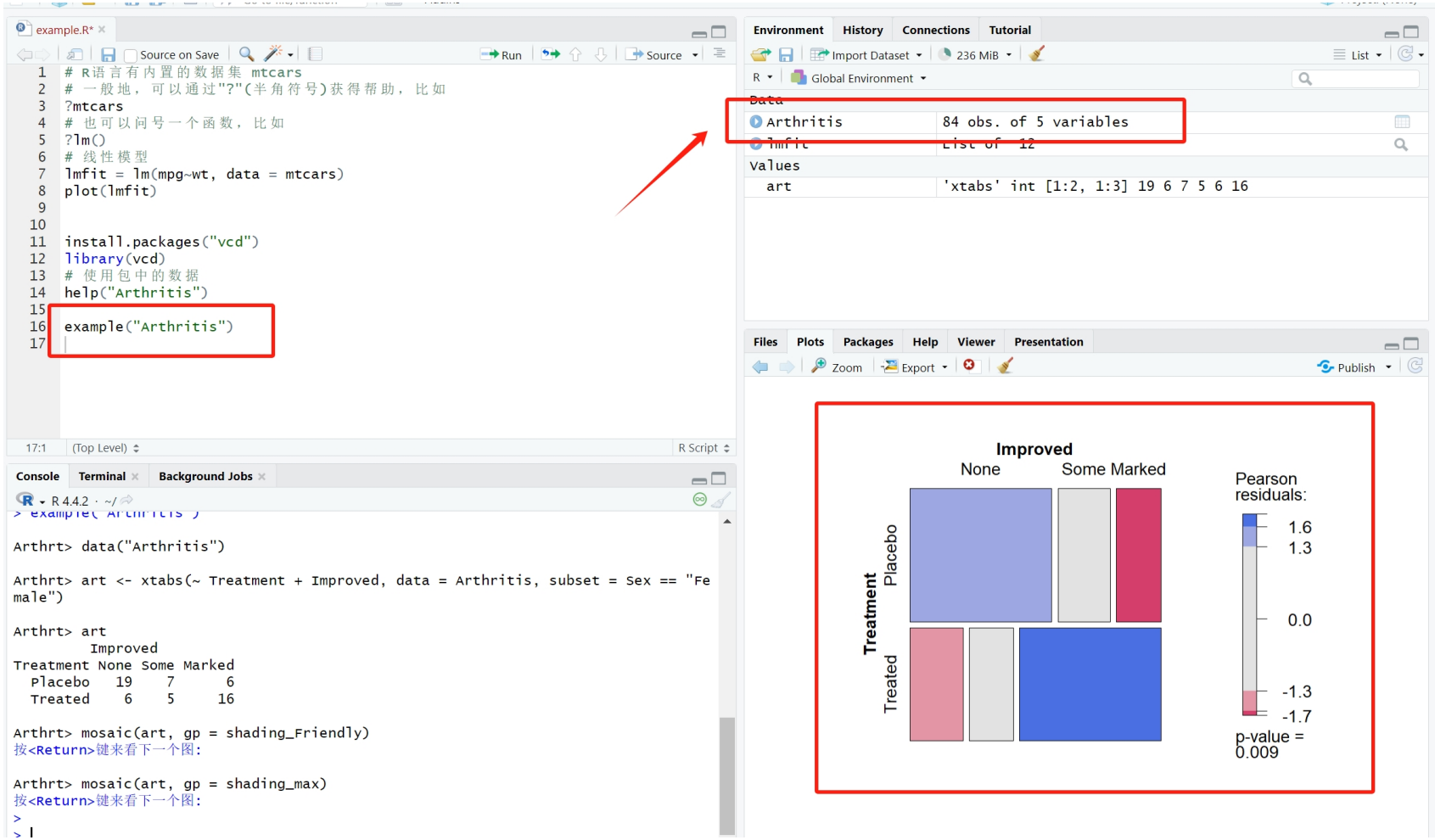


使用包中的数据



- 能不能查看一下数据的结构？试试看 `str(Arthritis)`, `attridutes(Arthritis)`.

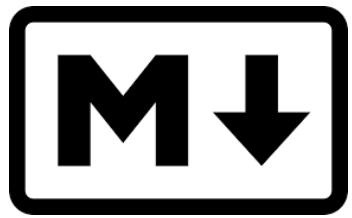
使用包中的数据



- 试试点击环境中带三角符号的变量，能发现什么？
- 试试 Plot 中的 Export 按钮，有什么用？能不能拉伸一下图片？

Rmarkdown 介绍

- Markdown 是一种轻量级标记语言, 经常用于写作文档. 几乎没有困难的语法, 你可以将它视为 txt 文档的升级版.



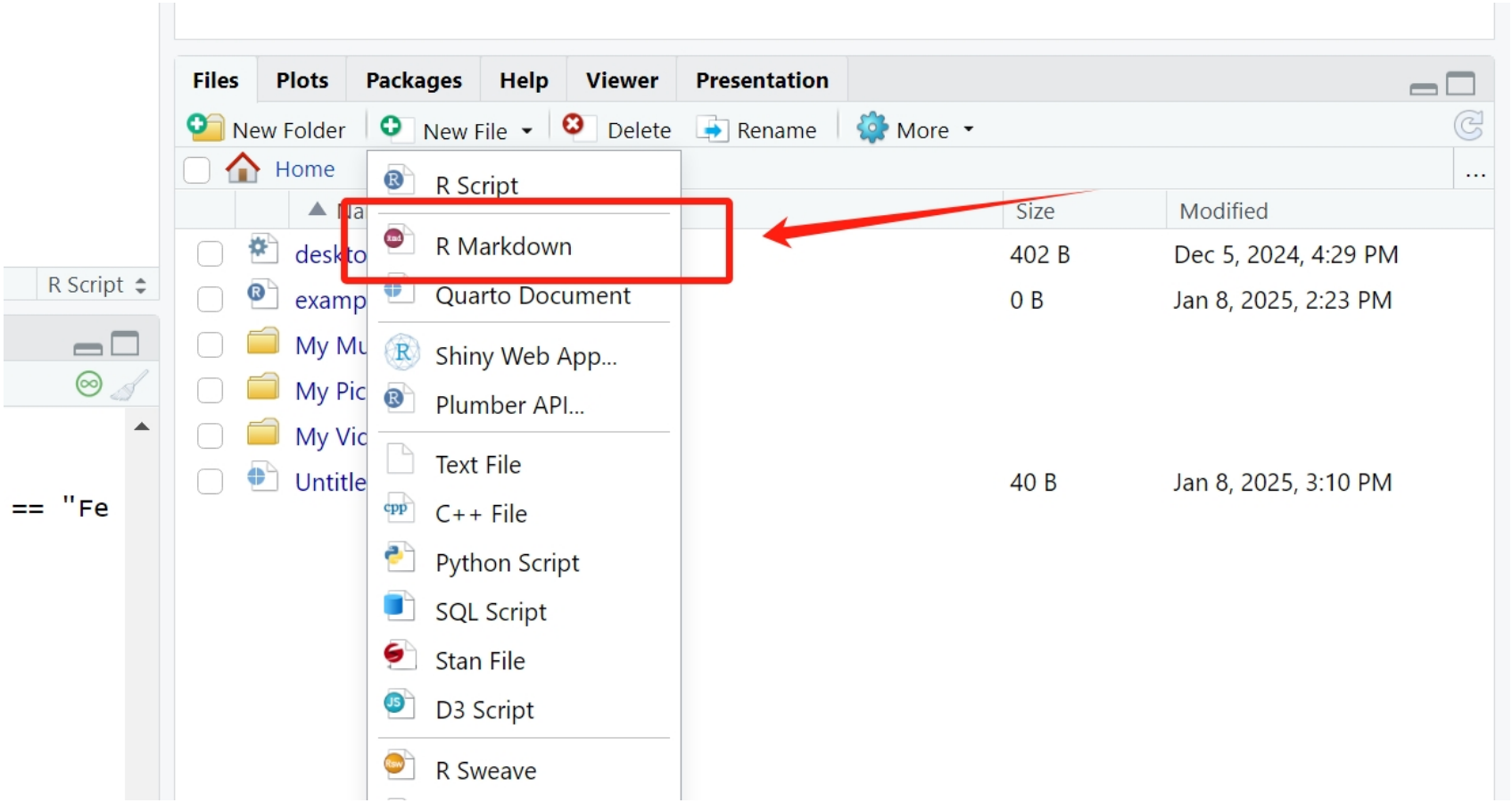
- Rmarkdown 是 R 语言的一个扩展, 可以在文档中直接运行 R 代码, 类似于 Python 的 Jupyter.

- 当然, 你甚至可以在 Jupyter 中运行 R 代码. 也可以在 Rmarkdown 中运行 Python 代码!

Rmarkdown

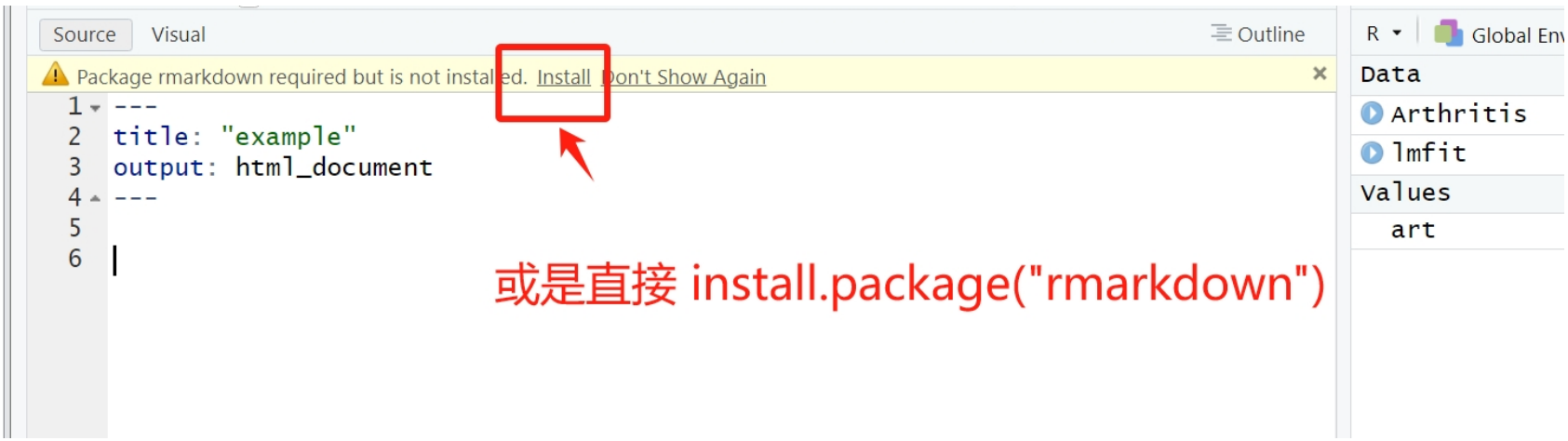

R Markdown
HTML
PDF
Adobe
W
TUFTE
Shiny
Spark
C++
Connect to data.
Run code.
Generate reports.

创建 rmd 文件



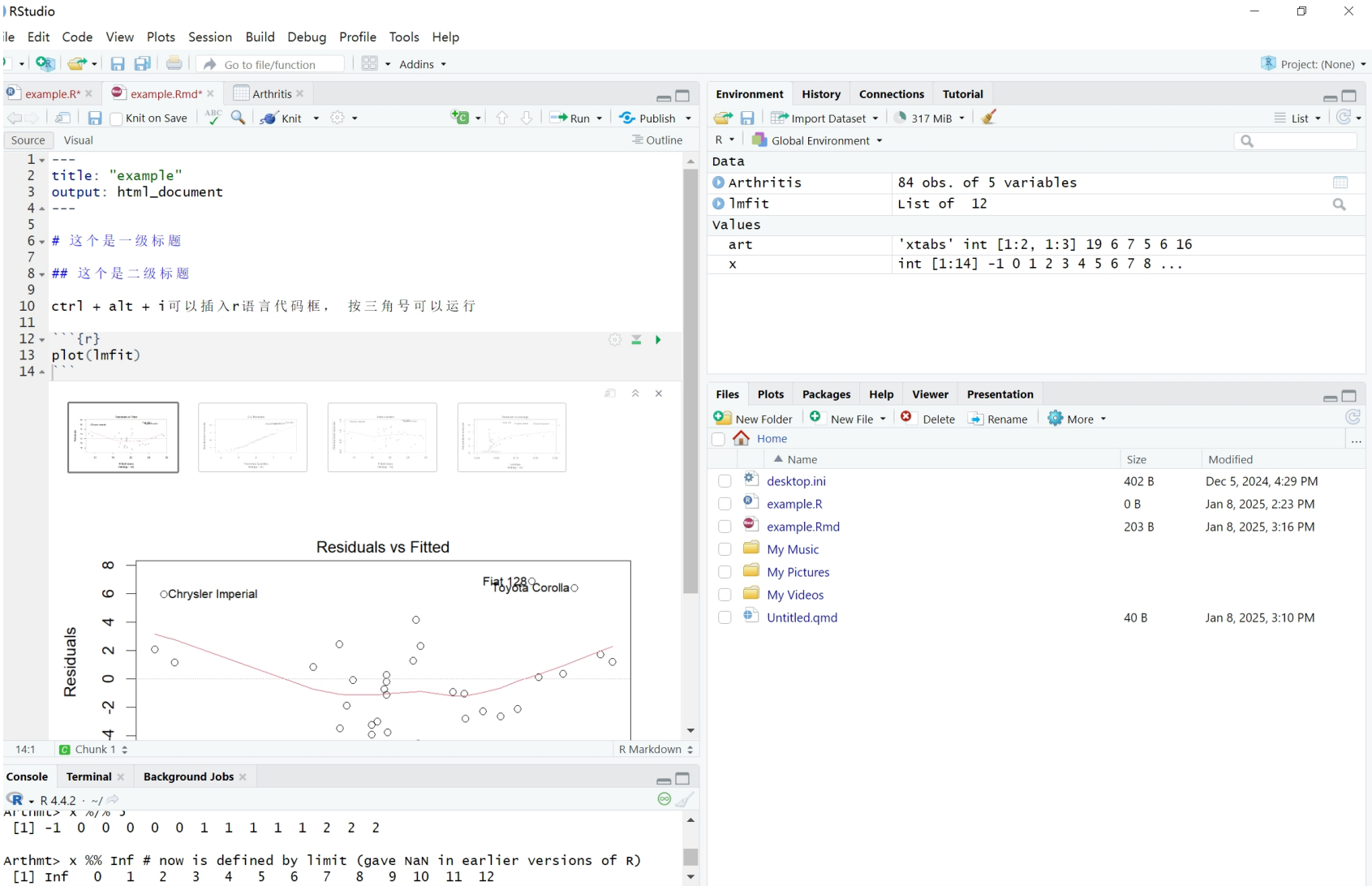
- 还能在哪里创建？试试看左上角.

安装包



- install.packages
- 也可以在右下角的 packages 栏中进行安装.

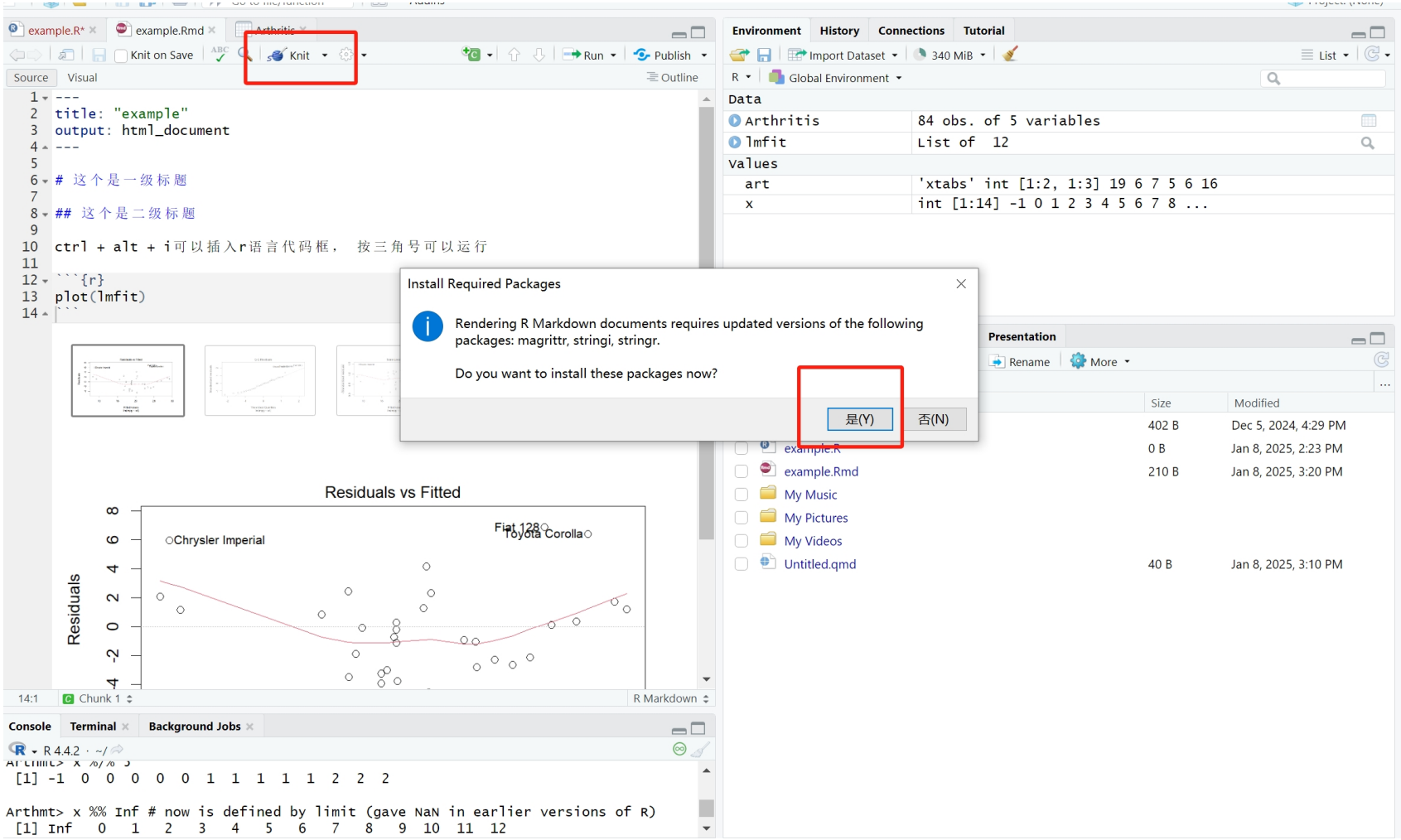
Rmarkdown



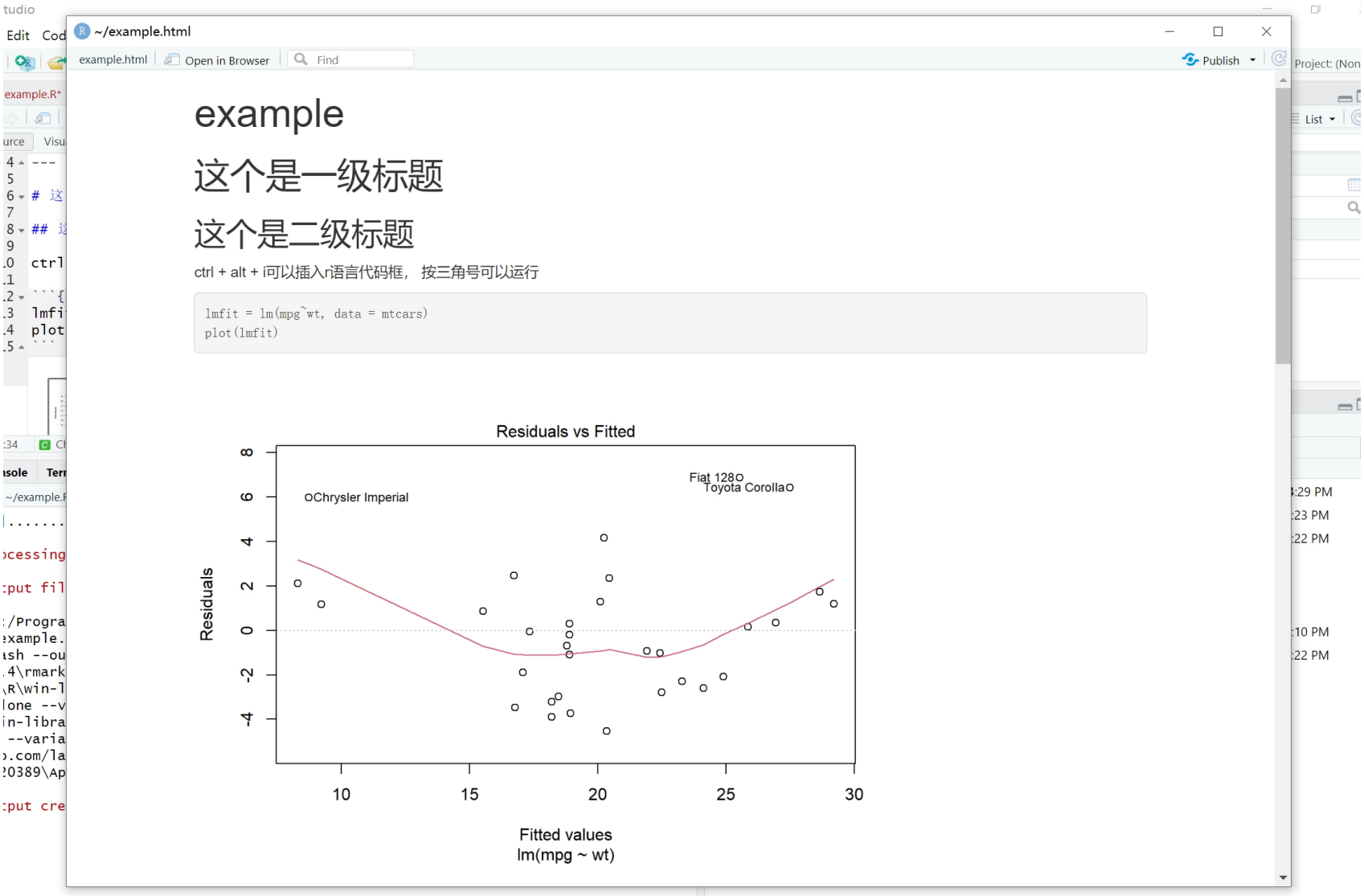
- 最常用的符号就是 # 号，用于表示标题；* 号改变字体格式；- 号表示列表等.
- 怎么改变字体的颜色？

Render/Knit

- 试试看 Knit 按钮，会发生什么？
- 试试看 Knit 按钮旁边的下拉菜单，有什么选项？选择到 html
- 是否会有报错？为什么？

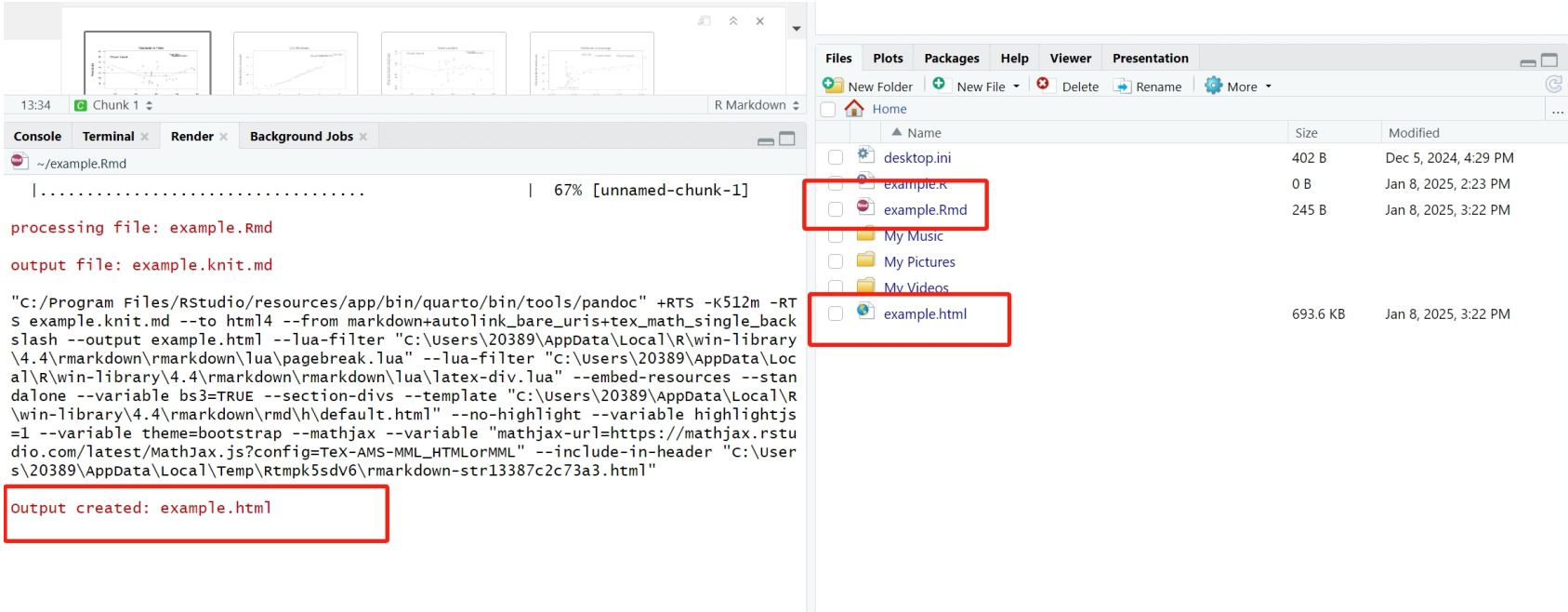


Render/Knit



- vignettes: 一般是一个包的使用说明书, 也是 Rmarkdown 的一个应用.

Render/Knit



- 推荐 render 到 html, 也可以 render 到 pdf, word 等.
- 试一试 render 到 pdf, 会发生什么?

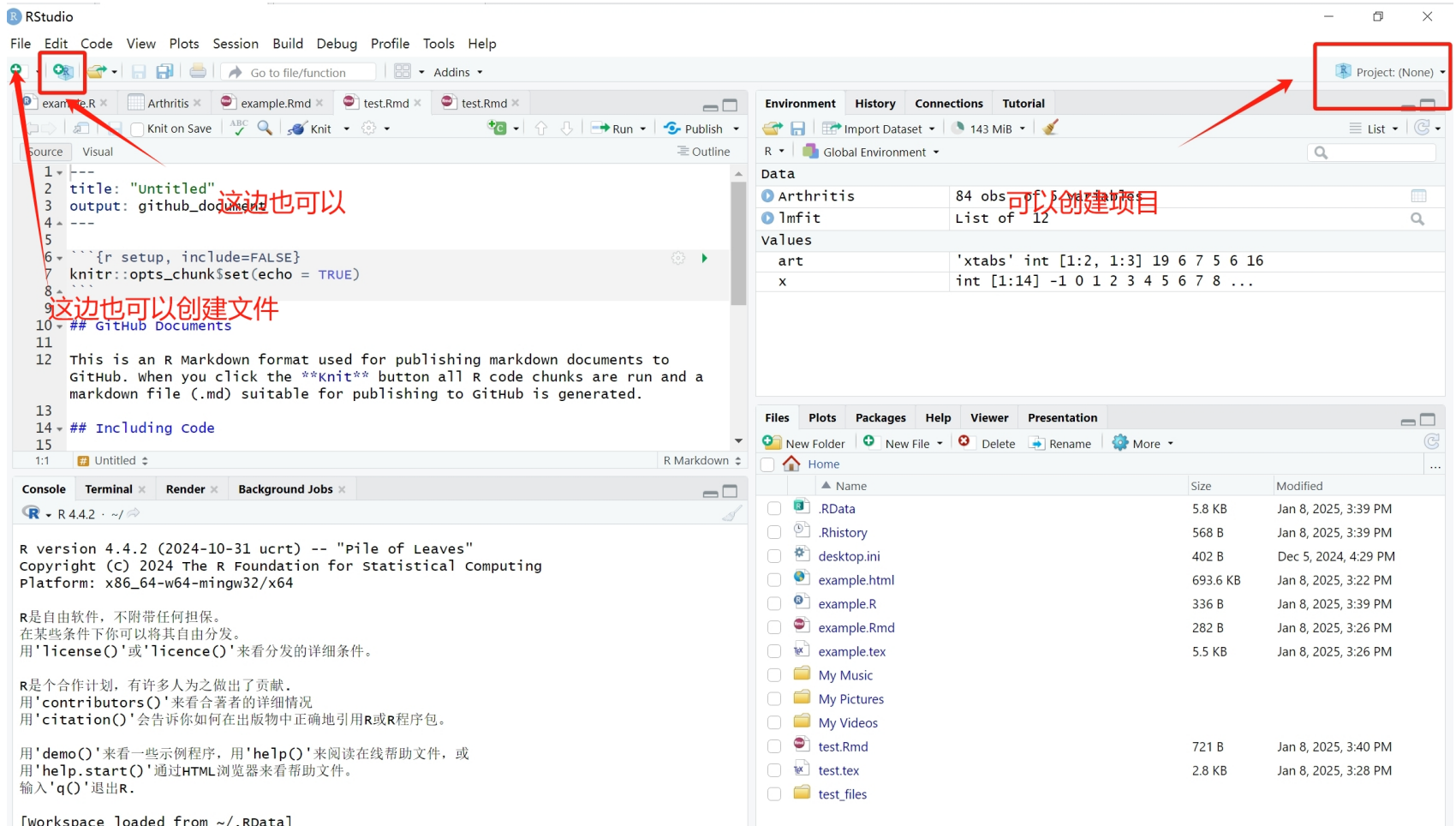
Rmarkdown 中文乱码问题

1. 本地没有 LaTeX 发行版,
   - 安装 TinyTeX, 一个精简的 LaTeX 发行版.
   - `install.packages("tinytex")`
   - `tinytex::install_tinytex()`
2. 本地已经有 TexLive, MikTex 等 LaTeX 发行版,
   - `install.packages("rticles")`
   - 选择适当的模板, 比如 ctex template
3. 多多查找互联网上的解决方案, 比如知乎等平台一定会有相应解决方法分享.

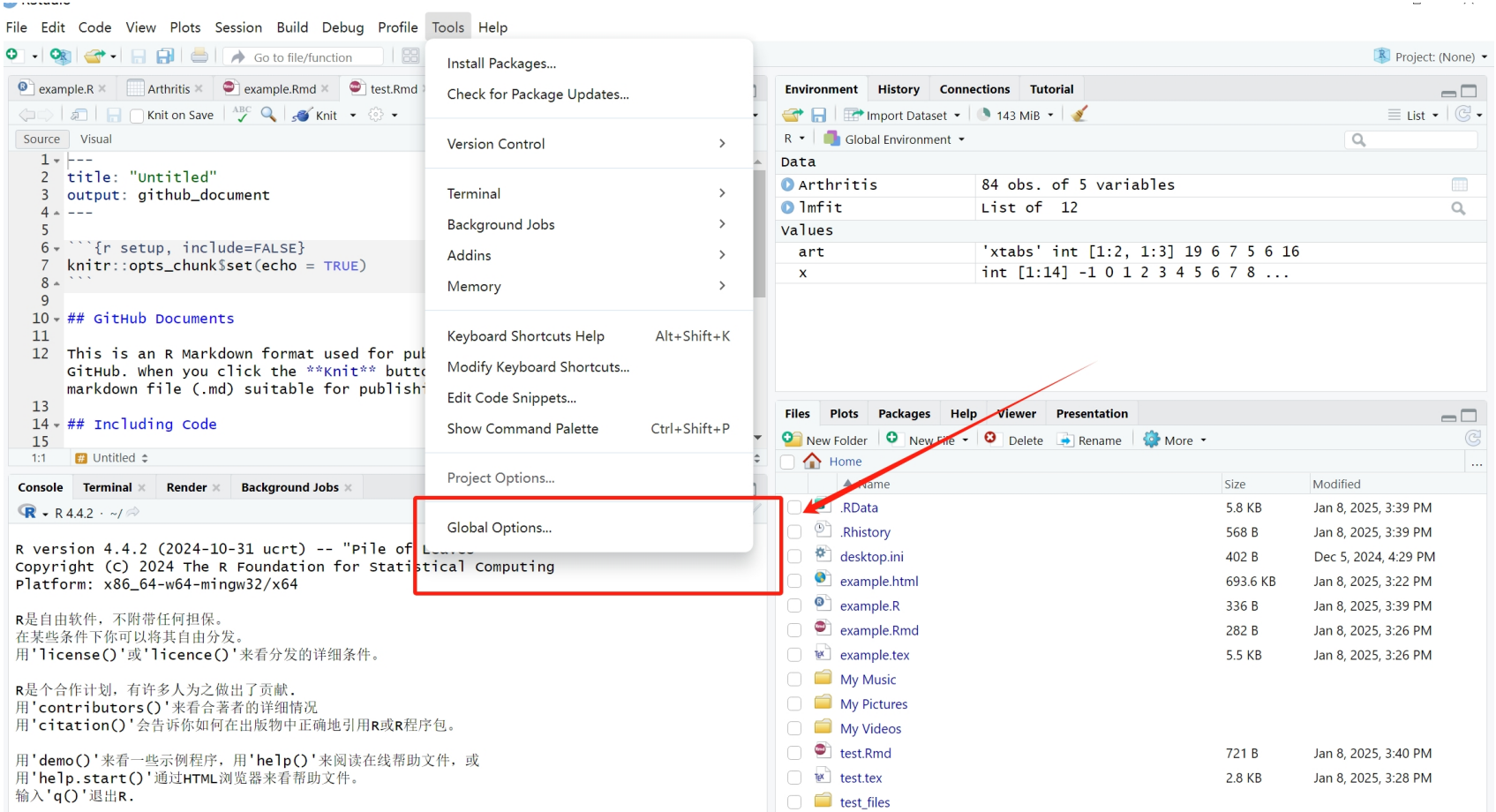
   `https://zhuanlan.zhihu.com/p/96993644`

创建 project

- 注意到，我们当前的工作路径是 home，也就是用户目录.
- 一般来说，我们不希望这种路径中的文件一团糟，往往自己会创建一个文件夹，用于存放项目文件.
- 另一方面，创建一个 project，可以方便的管理文件，也可以方便的分享文件.



设置 Rstudio



- 把界面设置为不同的主题？字体大小？
- 设置代码自动补全？缩进格式？
- 选择合适的 CRAN 镜像？

一些公开数据来源

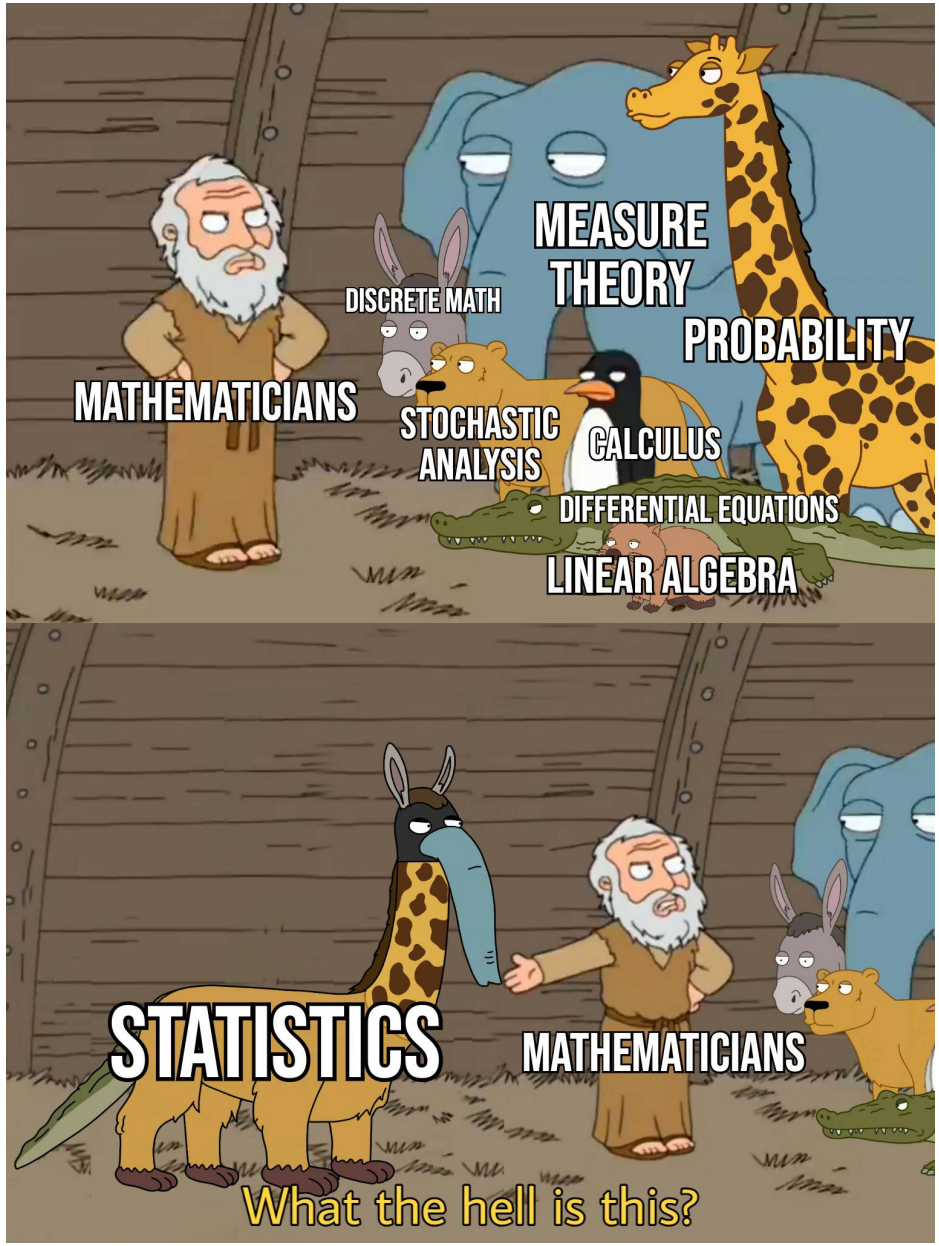
以下公开数据集来源供各位同学完成报告作业时参考，也可以使用其他来源的数据集，但请在小组展示的幻灯片和数据分析报告中注明数据源.

- General Statistical Learning Data
  - UC Irvine Machine Learning Repository: `https://archive.ics.uci.edu/ml/index.php`
  - Kaggle: `https://www.kaggle.com/datasets`
  - Data collections for psychological science research, by American Psychological Association: `https://www.apa.org/research-practice/conduct-research/data-links`
- Economics and Finance
  - Global Financial Development Database by The World Bank: `https://datacatalog.worldbank.org/search/dataset/0038648`
  - Yahoo Finance: `https://finance.yahoo.com/`
  - Finance Data Directory by U.S. Department of The Treasury: `https://home.treasury.gov/`

一些公开数据来源

- Biomedical Data
  - NIH Data Sharing Repositories: `https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html`
  - Gene Expression Omnibus (GEO), a genomics data repository maintained by The National Center for Biotechnology Information (NCBI), U.S.: `https://www.ncbi.nlm.nih.gov/geo/`
  - The Cancer Genome Atlas (TCGA) by National Cancer Institute (NCI), U.S.: `https://www.cancer.gov/ccg/research/genome-sequencing/tcga`
- Governmental Data
  - 国家统计局: `https://www.stats.gov.cn/sj/`
  - 上海公共数据开放平台: `https://data.sh.gov.cn/`
  - U.S. Population Data from U.S. Census Bureau: `https://www.census.gov/data.html`

学习统计学


MEASURE THEORY
DISCRETE MATH
PROBABILITY
MATHEMATICIANS
STOCHASTIC ANALYSIS
CALCULUS
DIFFERENTIAL EQUATIONS
LINEAR ALGEBRA
STATISTICS
MATHEMATICIANS
What the hell is this?

总结

看了本教程, 你应该会:

- 安装 R, Rstudio, Rtools.
- 创建一个 R 脚本, 运行一些代码.
- 安装一个包, 使用包中的数据.
- 创建一个 Rmarkdown, 运行一些代码.
- 创建一个 project.
- 设置 Rstudio.

更深入地学习统计学和统计软件, 不但要在课上认真思考, 还要多看书, 多学习别人的代码, 希望大家在学习 R 语言的过程中有所收获.