

Multi-omics data analysis and visualisation, #1

HUST Bioinformatics course series

Wei-Hua Chen (CC BY-NC 4.0)

02 September, 2024

section 1: TOC

Course contents

- ① 开发平台相关软件安装
- ② R 基础知识
- ③ 数据处理
- ④ 做图
- ⑤ 常见组学数据和分析、可视化方法
- ⑥ 机器学习在组学分析当中的应用

机器学习（重点新增内容）

内容

- 基础知识
- 应用示例

现有工作介绍

- 宁康：人工智能赋能菌群大数据挖掘
- 薛宇：机器学习在新冠标志物和图像识别当中的应用
- 陈卫华：机器学习在肠道菌群为基础的疾病标志物挖掘当中的应用

宁康



Figure 1: 宁康教授在热心肠先生的主题报告

SOHU 页面

宁康, cont.

Briefings in Bioinformatics

Issues Advance articles Submit ▾ Purchase Alerts About ▾ Briefings in Bioinf



Volume 24, Issue 3
May 2023

JOURNAL ARTICLE

Artificial intelligence-enabled microbiome-based diagnosis models for a broad spectrum of cancer types [Get access ▾](#)

Wei Xu, Teng Wang, Nan Wang, Haohong Zhang, Yuguo Zha, Lei Ji, Yuwen Chu, Kang Ning 

Briefings in Bioinformatics, Volume 24, Issue 3, May 2023, bbad178,
<https://doi.org/10.1093/bib/bbad178>

Published: 04 May 2023 Article history ▾

< Previous Next >

“ Cite  Permissions  Share ▾

Figure 2: *Briefings in Bioinformatics* (2023 IF=9.5)

基于肿瘤菌的多种肿瘤诊断模型

薛宇

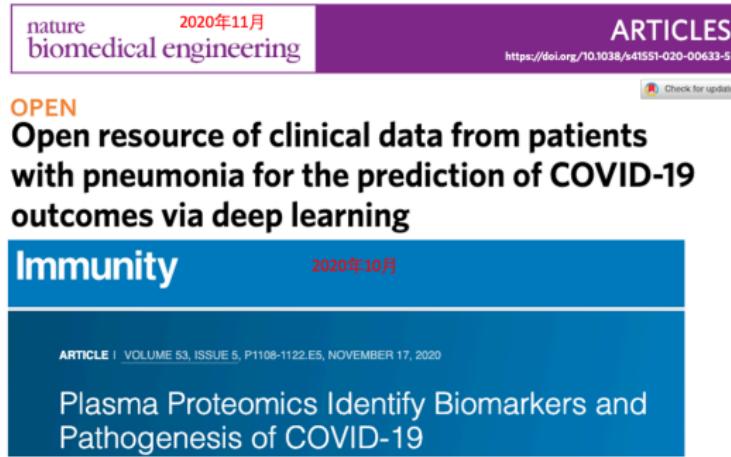


Figure 3: Nature BME & Immunity (2023 IF=29)

机器学习在新冠标志物和图像识别当中的应用

薛宇, cont.

nature communications

Article <https://doi.org/10.1038/s41467-023-38414-8>

Small-sample learning reveals propionylation in determining global protein homeostasis

Received: 18 July 2022
Accepted: 28 April 2023
Published online: 17 May 2023

Ke Shui^{①,6}, Chenwei Wang^{①,6}, Xuedi Zhang^②, Shanshan Ma¹, Qinyu Li¹,
Wanshan Ning^③, Weizhi Zhang^④, MiaoMiao Chen¹, Di Peng^⑤, Hui Hu^⑥,
Zheng Fang^{⑦,8}, Anyuan Guo^⑨, Guanjun Gao^⑩, Minglang Ye^⑪,
Luoying Zhang^{⑫,4} & Yu Xue^{⑬,5}

Check for updates

Figure 4: Nature communications (2023 IF=16.6)

小样本机器学习

陈卫华

Cell Metabolism

2021年3月

CellPress

Letter

Treatment regimens may compromise gut-microbiome-derived signatures for liver cirrhosis

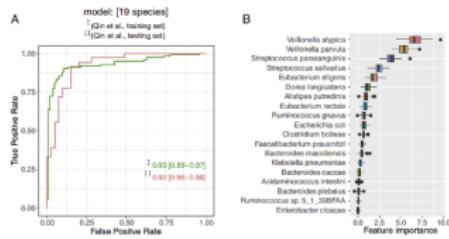
Siheng Wu,^{1,6} Puzi Jiang,^{1,6} Xing-Ming Zhao,^{2,3,4,*} and Wei-Hua Chen^{1,5,*}

Figure 5: Cell Metabolism (2023) IF:29

机器学习在肠道菌群为基础的疾病标志物挖掘当中的应用

基于肠道菌群的疾病诊断

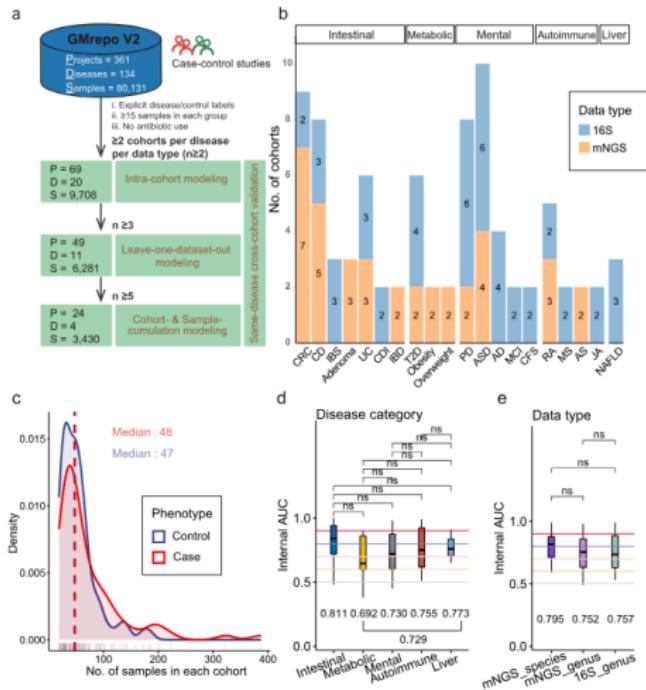


Figure 6: Gut Microbes (2023 IF: 12.2)

基于肠道菌群的预后分析

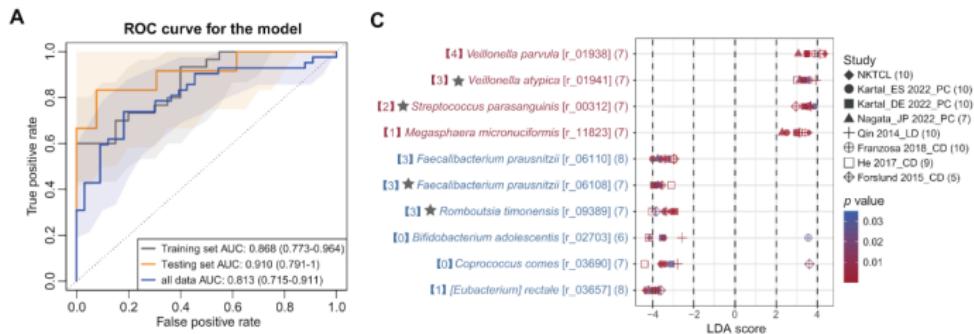


Figure 7: Gut (2023) IF: 24)

肠道菌群可预示 NKTCL 的预后

其它动物的肠道菌群

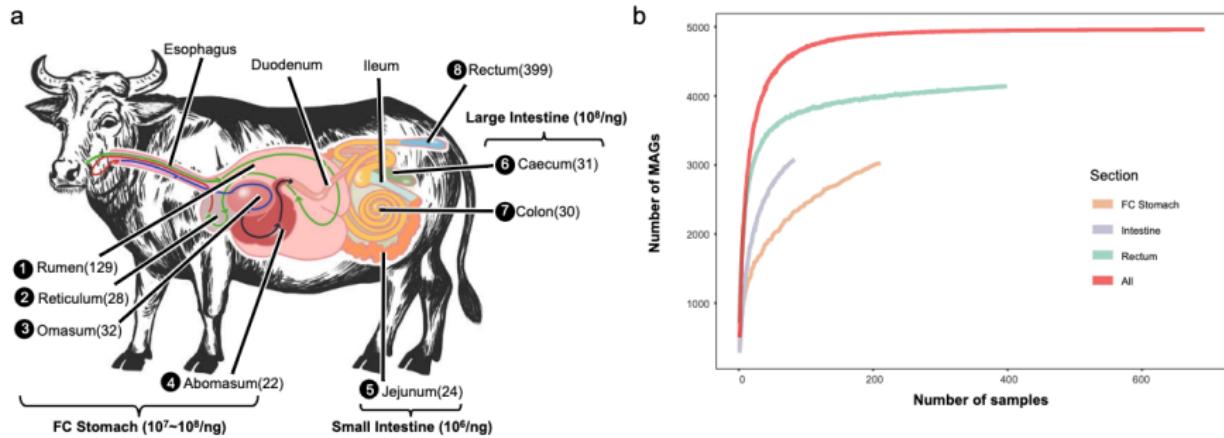


Figure 8: Nature communications (2023 IF: 16.6)

噬菌体研究

RESEARCH ARTICLE

**ADVANCED
SCIENCE**
Open Access

www.advancedscience.com

Long-Read Sequencing Reveals Extensive DNA Methylation in Human Gut Phagenome Contributed by Prevalently Phage-Encoded Methyltransferases

Chuqing Sun, Jingchao Chen, Menglu Jin, Xueyang Zhao, Yun Li, Yanqi Dong, Na Gao, Zhi Liu, Peer Bork,* Xing-Ming Zhao,* and Wei-Hua Chen**

DNA methylation plays a crucial role in the survival of bacteriophages

1. Introduction

Figure 9: Advanced science (2023 IF: 15.1)

基于三代测序的肠道噬菌体研究

Class rules

- ① 每次随机点名
- ② 课堂随机提问
- ③ 每人有 1 次无理由缺课机会
- ④ 考试、平时成绩各占总成绩的 50%

注：严格按以上进行，如多次表现不好，可能会不及格

作业和扣分规则

- ① (几乎) 每次课后都有作业;
- ② 直接从平时成绩 50 分中扣除;
- ③ 作业迟交三天以内每次扣 3 分; 七天以内每次 5 分; 七天及以上每次 10 分;
- ④ 每人有一次迟交机会 (3 天以内), 不扣分。
- ⑤ 补交仅限前三次作业;
- ⑥ 老师有最终解释权

考试

- ① 开卷；允许带资料、书籍，不允许带任何有大屏幕或智能屏幕的电子设备
- ② 内容都在 ppt 上或作业里，不会超纲
- ③ 考试时允许零星英文，但不允许大段英文（抄 ppt）
- ④ 老师有最终解释权

section 2: why choose R?

R 语言简史

1993 到 2000 这段时间 R 只在小范围内流传。2000 年之后开始大爆发，用户数量直线上升。除去 R 本身的优秀之外，这种爆发与多个因素有关，比如自由软件的兴起，Linux 的成熟等等；经济危机也促进大家采用免费的自由软件替代统计领域的传统强者如 SPSS、SAS 和 Matlab 等（注：均为收费软件）。

首先，越来越多的学术文章使用 R 作为分析工具。根据来自著名学术搜索引擎 Google Scholar（谷歌学术）的数据，R 的流行趋势有以下两个特点：1) 在学术领域的市场份额逐年增加，且增势迅猛，2) R 是为数不多市场份额增加的统计软件之一。

接下来我们就用 R 把这个趋势画出来！如下面代码所示，所需代码包括 4 个部分：装入所需要的包，读取数据，处理数据和作图。运行这段代码，既专业又美观的图片就生成了！

R 的流行性调查

代码

```

library("ggplot2"); library("reshape2");

dat <- read.csv(file = "data/talk01/chaper01_preface_scholarly_impact_2012.4.9.csv");

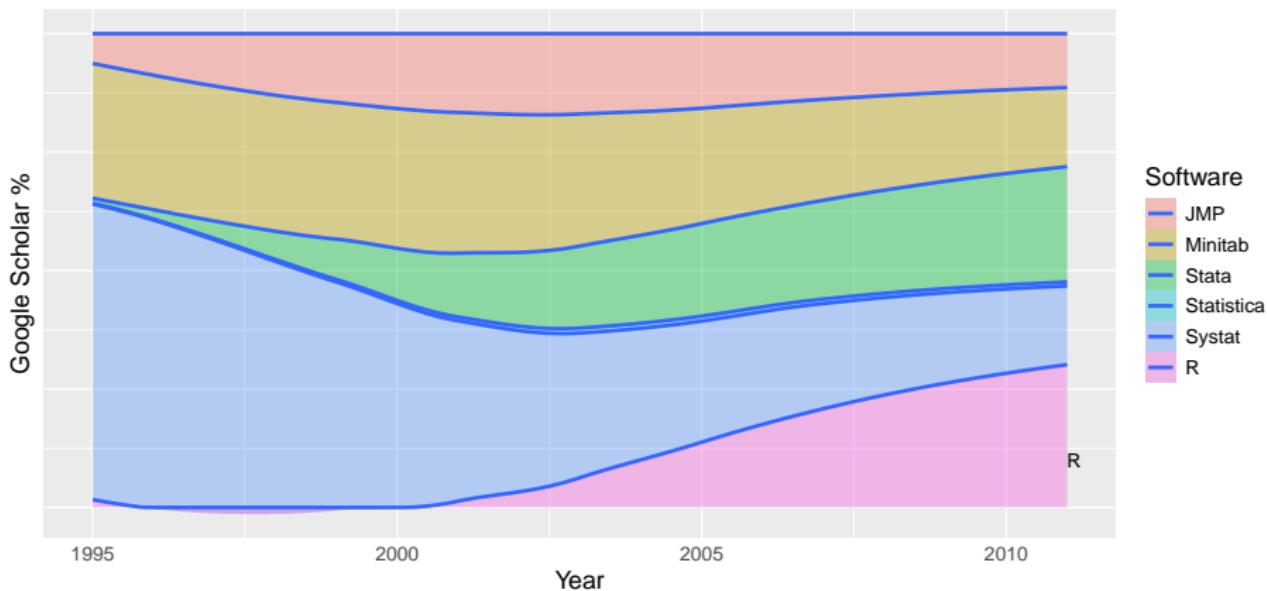
cols.subset <- c("Year", "JMP", "Minitab", "Stata", "Statistica", "Systat", "R");
Subset <- dat[ , cols.subset];
ScholarLong <- melt(Subset, id.vars = "Year");
names(ScholarLong) <- c("Year", "Software", "Hits");

plot1 <-
  ggplot(ScholarLong, aes(Year, Hits, group=Software)) + # 准备
  geom_smooth(aes(fill=Software), position="fill", method="loess") + # 画图
  ggtitle("Market share") + # 设置图标题
  scale_x_continuous("Year") + # 改变 X 轴标题
  scale_y_continuous("Google Scholar %", labels = NULL ) +
  theme(axis.ticks = element_blank(), text = element_text(size=14)) +
  guides(fill=guide_legend( title = "Software", reverse = F )) +
  geom_text(data = data.frame( Year = 2011, Software = "R", Hits = 0.10 ),
            aes(label = Software), hjust = 0, vjust = 0.5);

```

Market share, result

Market share



注：这里移除了市场占有率较大的 SAS 和 SPSS

R 的招聘趋势

其次，统计分析相关工作的招聘信息中要求申请者会用 R 的也越来越多了。根据美国招聘搜索引擎 indeed.com 的数据，自 2005 年（此搜索引擎提供的最早数据）起，需要用到 R 的招聘信息占总体招聘的比例逐年上升，目前仅排在 SAS 和 Matlab 之后，处于第 3 位。而且，除了 Stata 之外，R 是唯一一个占比上升。

同样的，我们用 R 把这个趋势画出来！

R job trends

代码

```

library("ggplot2"); ## 主作图包

##2. -- 读取数据 --
dat <- read.table(file ="data/talk01/chaper01_preface_indeed_com_stats_2015.txt",
                  header = T, as.is = T);

##3. 处理数据
dat$date <- as.Date(dat$date); ## 把第一列改为日期

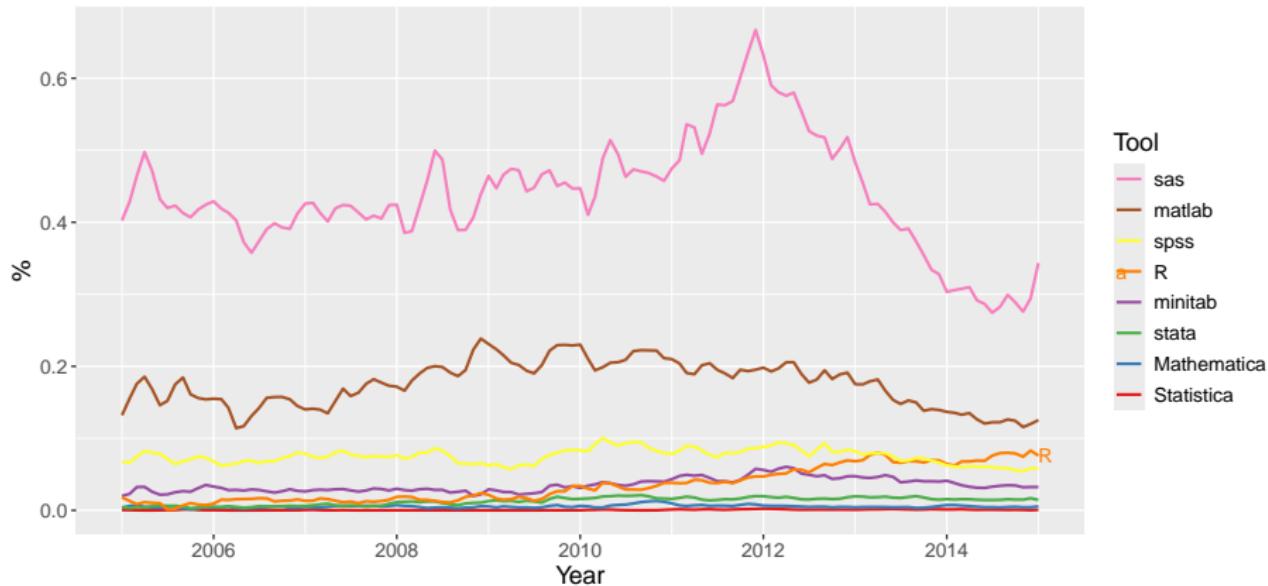
# 根据 job 对 software 进行调整
dat <- transform(dat, software = reorder(software, job));

plot2 <-
  ggplot( dat, aes( date, job, group = software, colour = software) ) +
  geom_line( size = 0.8 ) +
  ggtitle("Job trends (data from indeed.com)") + # 设置图标题
  xlab("Year") + ylab("%") +
  # 改变字体大小; 要放在 theme_grey() 后面
  theme( text = element_text(size=14) ) +
  guides(colour=guide_legend( title = "Tool", reverse = TRUE )) +
  scale_colour_brewer(palette="Set1") + # 改变默认颜色
  geom_text(data = dat[dat$date == "2015-01-01" & dat$software %in% c("R"), ],
            aes(label = software), hjust = 0, vjust = 0.5);

```

R job trends, plot

Job trends (data from indeed.com)



Popularity of Programming language 2020

Worldwide, Sept 2023 :				
Rank	Change	Language	Share	1-year trend
1		Python	27.99 %	+0.1 %
2		Java	15.9 %	-1.1 %
3		JavaScript	9.36 %	-0.1 %
4		C#	6.67 %	-0.4 %
5		C/C++	6.54 %	+0.3 %
6		PHP	4.91 %	-0.4 %
7		R	4.4 %	+0.2 %
8		TypeScript	3.04 %	+0.2 %
9	↑↑	Swift	2.64 %	+0.6 %
10		Objective-C	2.15 %	+0.1 %

Figure 10: PYPL ranking (<https://pypl.github.io/>)

Programming languages for bioinformatics

Perl 或 Python

- 强大的文本处理能力（包括序列）
- 不错的运行速度（尤其是 Python）
- 强大的生信和统计学扩展包（尤其是 Python）
- 方便的并行计算

R

- 强大的格式数据处理能力（二维表格, dplyr）
- 无以伦比的统计学专业性
- 专业而好看的数据可视化软件（ggplot2）
- 专业的生信扩展包（Bioconductor）
- 超级好用的整合开发环境 IDE（RStudio）

我用过的 programming languages

- - C
 - - Perl
 - - R
 - - PHP
 - - Java
 - - MySQL
 - - HTML
 - - Javascript

Evolview ver3.0
cited 225 times in 2021 (ver2+3),
166 so far (as of Aug 30, 2022)

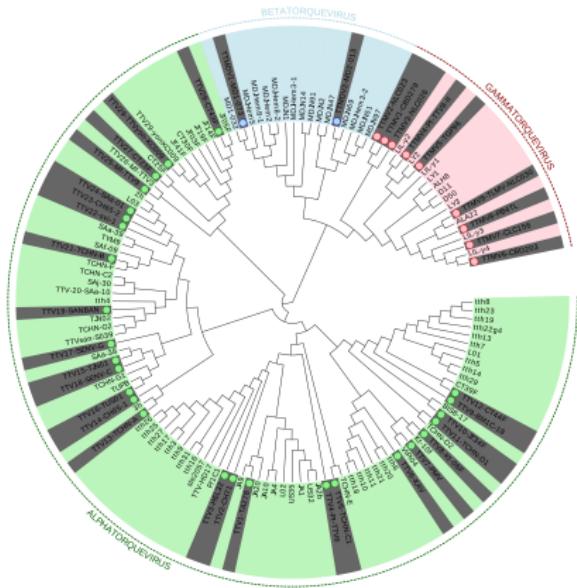


Figure 11: .Evolview showcase 3

网站链接、参考文献和扩展阅读

综上所述，R 已经是最流行的免费统计分析软件，排名仅在几个传统的分析软件之后，而且大有赶超它们的趋势。学好 R，不仅有助于在学术研究领域的发展，对找工作也有不少的帮助。

- R 的官方网站: <http://www.r-project.org>
- R 档案综合网络，即 CRAN(Comprehensive R Archive Network):
<http://cran.r-project.org/>
- ggplot2: <http://ggplot2.org/>
- RStudio: <https://posit.co/products/open-source/rstudio/>
- 如何从 Google Scholar 抓取引用数据:
<http://librestats.com/2012/04/12/statistical-software-popularity-on-google-scholar/>
- indeed 招聘趋势: www.indeed.com/jobtrends
- R for data science: [https://r4ds.had.co.nz \(必读!!\)](https://r4ds.had.co.nz)

Section 3: setting up working environment

Install R

Go to <https://mirrors.tuna.tsinghua.edu.cn/CRAN/> (清华镜像),
 R supports mainstream operating systems including Linux, Windows and MacOS, please download the corresponding installation files according to the operating system. As shown in the figure below:

The Comprehensive R Archive Network

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 used. If you do not know what this means, you probably do not want to do it!

- The latest release (2023-06-16, Beagle Scouts) [R-4.3.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing correction requests or bug reports.
- 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 frequent](#) before you send an email.

Install R on Linux

目前大多 Linux 发行版都带有 R，因此可直接使用。从 CRAN 下载文件进行安装稍嫌复杂，要求用户对 Linux 系统有一定的了解，而且需要有管理员权限。建议初级用户在 Linux 高手指导下安装。点击上图中的"Download R for Linux" 后，发行版为 Redhat (红帽) 或 Suse 的用户要先阅读网站上提供的 readme 或 readme.html 文件，然后其中的指示进行安装。这里就不再累述了。

r-base-core_3.1.3-1lucid_amd64.deb 或 r-base-core_3.1.2-1lucid0_i386.deb
----- ----- ----- -----
1 2 3 4 5 1 2 3 4 5

Figure 13: R 安装包文件名

R studio

RStudio 可以从 <https://posit.co/downloads/> 下载，支持等主流的操作系统。

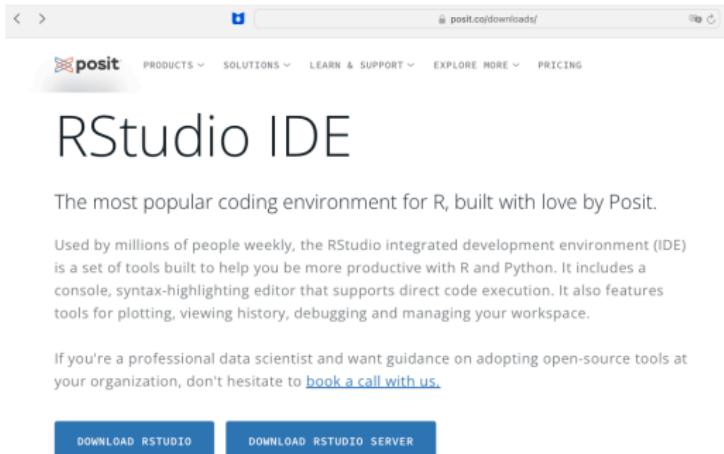
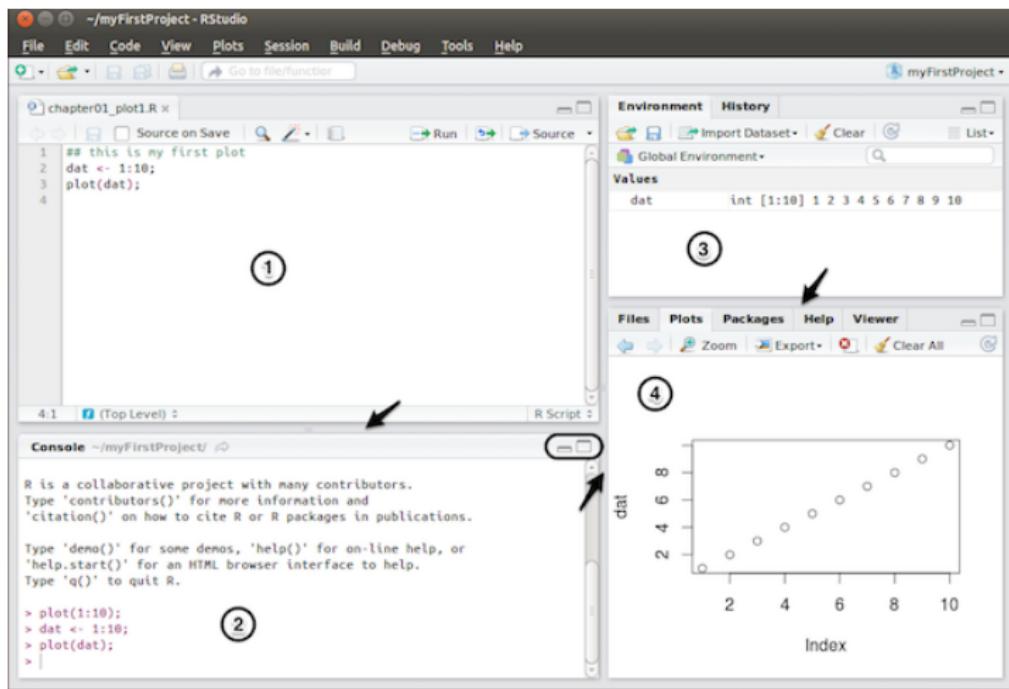


Figure 14: RStudio website main page

RStudio 有商业和免费版本；也有 server 版

R studio, cont.

RStudio 运行时的界面如下图所示，除了顶部的菜单栏工具栏之外，主界面还包括 4 个子窗口：



R studio, cont.

1. 代码编辑器

- 具有代码编辑、语法高亮、代码和变量提示、代码错误检查等功能
- 选中并向 R 控制台（窗口 2）发送并运行代码。用快捷键 Ctrl+Enter (MacOS 下是 Cmd+Enter) 进行代码发送。没有代码选中时，发送光标所在行的代码
- 可同时打开编辑多个文件
- 除 R 代码外，还支持 C++、R MarkDown、HTML 等其它文件的编辑
- 也可用于显示数据

2. R console

- 可在此直接输入各种命令并查看运行结果。支持代码提示

3. 变量列表及代码运行的历史记录

R studio, cont.

4. 其它窗口

- 当前工作目录下的文件列表
- 作图结果
- 可用和已安装的扩展包；在这里可以直接安装新的和升级已有的扩展包
- 帮助

注意，子窗口之间可以通过快捷键 $\text{Ctrl} +$ 子窗口编号进行切换。如 $\text{Ctrl} + 1$ 可以切换到代码编辑子窗口， $\text{Ctrl} + 2$ 则切换到 R 控制台。

其它特点

- 创建、管理 projects

R studio 特点详解

代码提示/自动完成

子窗口 1 和 2 都提供有代码提示功能，即：用户输入 3 个字母时，RStudio 会列出所有前 3 个字母相同的变量或函数名供用户选择；用户可通过键盘的上下键选择，然后用 Enter（回车）选定，非常方便。变量或函数名前面的小图标表示了它们的类型；如果当前高亮的是函数，RStudio 还会显示其部分帮助内容。

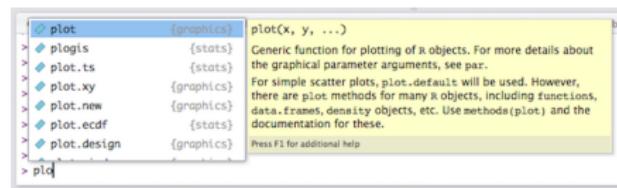


Figure 16: R studio code autocomplete

R studio 特点详解, cont.

查看变量内容

子窗口 3 内会列出所有当前使用的变量、变量的类型以及大小，如下图：

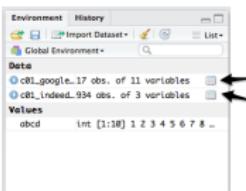


Figure 17: loaded variables

有些简单变量，如数组，RStudio 会直接显示其部分值；对于复杂一些的变量，比如 data.frame（类似于二维表格），则可以点击变量名前边的小三角标识展开其内容。当变量的最右侧出现小网格状图标时（如上图箭头所指位置），点击它们后可以在子窗口 2 内察看。

R studio 特点详解, cont.

导出作图并选择导出格式

RStudio 的第 4 子窗口里集中了许多有用的功能，组织在不同的'Tab'（标签）内。比如作图（plots），不仅可以察看画图的结果，还可以导出当前图像至硬盘，或拷贝至剪贴板；如下图所示。支持导出格式有 png 和 pdf。

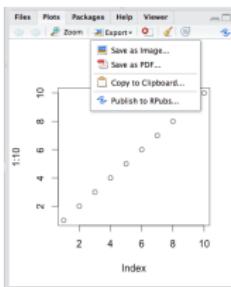


Figure 18: export active plot to various graphical formats

R studio 特点详解, cont.

查看已安装的包

通过第 4 子窗口的“包”(Packages)标签内的工具，用户可以很方便的查看已安装的包：

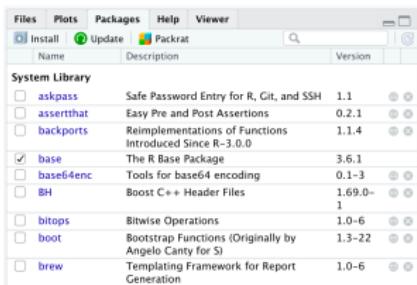


Figure 19: Check installed packages

install new package(s)

同样通过第 4 子窗口的“包”（Packages）标签内的工具，安装新的包：

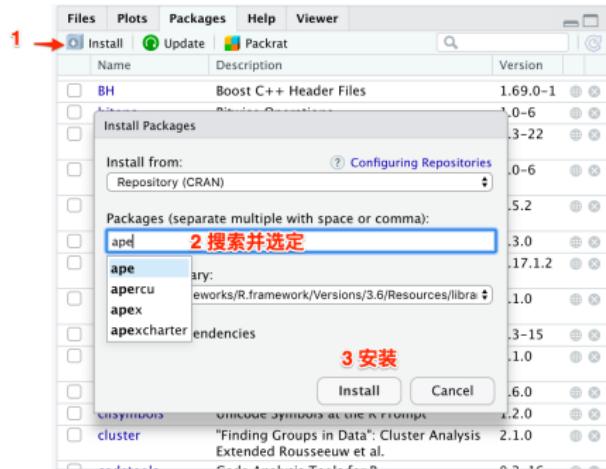


Figure 20: install new package

Packages needed for this study

大部分包都由 Posit（前身为 RStudio）公司提供；包括：ggplot2, tidyverse, readr, stringr 等。可以用以下命令一次性安装。不过，为方便读者直接从后面章节阅读，在每一次使用新包时，我们会再次进行提示安装方法。

```
install.packages( c("ggplot2", "tidyverse", "readr", "stringr") );
```

也可以单独安装：

```
install.packages( "ggplot2" ); # 安装作图用的 ggplot2  
install.packages( "tidyverse" ); # 数据处理用等
```

第一次运行命令 `install.packages()` 时，系统会提示选择镜像网站；请选择地理位置上距你最近的镜像（比如中国）。

install packages, cont.

You can also choose your CRAN mirror manually (recommended when installing takes a long time):

```
chooseCRANmirror();
```

```
> chooseCRANmirror()
Secure CRAN mirrors

 1: 0-Cloud [https]          2: Algeria [https]           3: Australia (Canberra) [https]
 4: Australia (Melbourne 1) [https] 5: Australia (Melbourne 2) [https] 6: Australia (Perth) [https]
 7: Austria [https]          8: Belgium (Ghent) [https]        9: Brazil (PR) [https]
10: Brazil (RJ) [https]       11: Brazil (SP 1) [https]       12: Brazil (SP 2) [https]
13: Bulgaria [https]         14: Chile 1 [https]            15: Chile 2 [https]
16: China (Hong Kong) [https] 17: China (Guangzhou) [https] 18: China (Lanzhou) [https]
19: China (Shanghai 1) [https] 20: China (Shanghai 2) [https] 21: Colombia (Cali) [https]
```

Figure 21: Choose CRAN mirror

You can also use `chooseBioCmirror()`; to choose mirror for BioConductor packages.

Packages needed for this study, cont.

实际上，以上包属于一个 meta-package，我们只需要安装它就可以了：

```
install.packages("tidyverse")
```

它是以下包的集合，都由 <https://www.tidyverse.org> 开发：

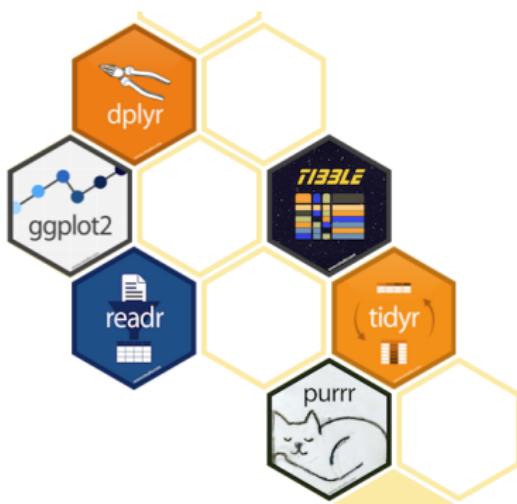


Figure 22: tidyverse: a mega package

R studio server

特点：

- 在服务器上安装，使用服务器的强大计算资源
- 通过网页登录，使用服务器帐号密码（方便，安全）
- 一直运行

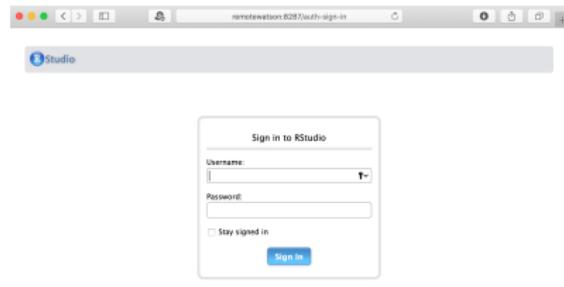


Figure 23: RStudio server web login form (w/ linux account)

R studio server, cont.

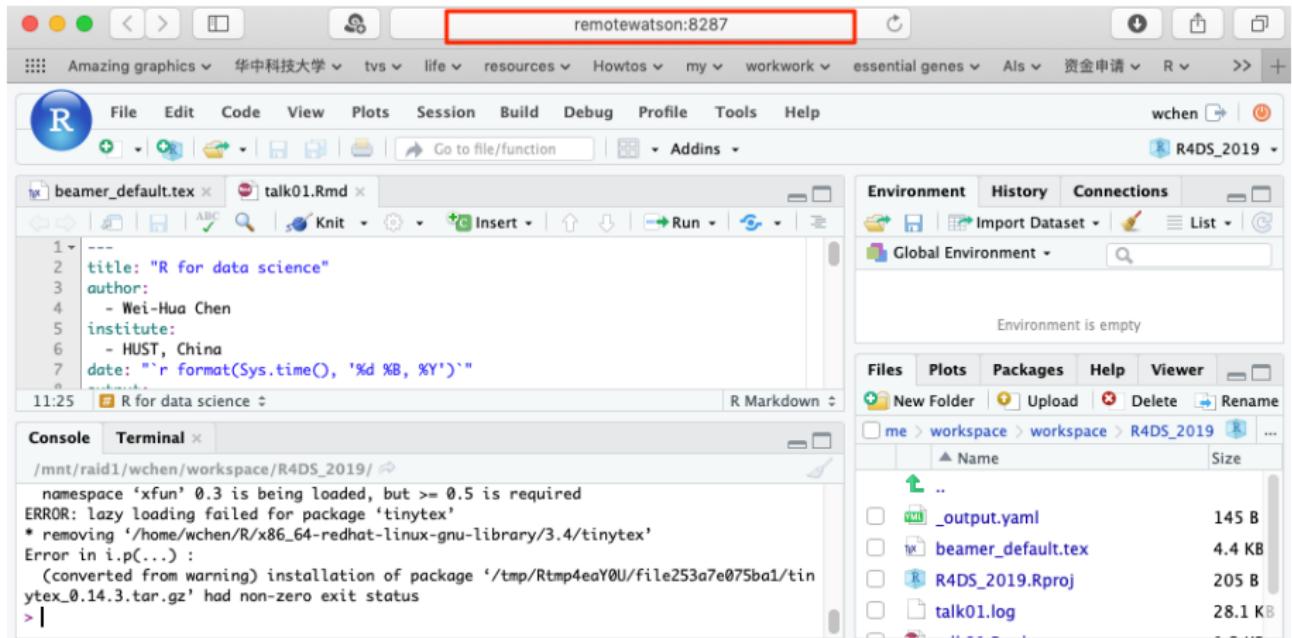


Figure 24: RStudio server 界面（通过浏览器）

R studio packages for data science

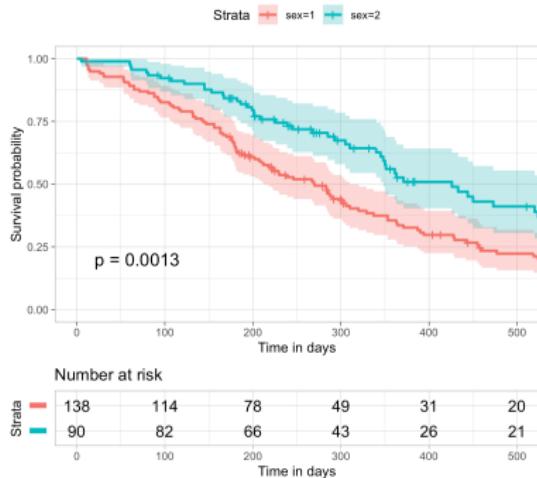
all part of (业界良心) tidyverse

- dplyr: 强大且方便的数据处理
- tydvr: 数据转换工具
- readr: 方便的文件 IO
- stringr: 文本处理
- Tibble: 代替 data.frame 的下一代数据存储格式
- purr: (暂时还未用到的包 ~~)

R studio packages for data visualisation

tidyverse

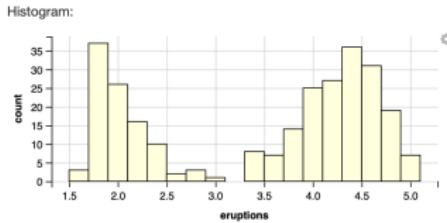
- ggplot2: 专业好用（但学习曲线很陡）的画图工具
 - <http://ggplot2.tidyverse.org>
 - gallery: <http://www.ggplot2-exts.org/gallery/>



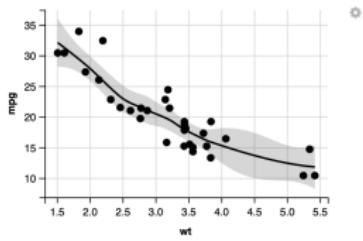
RStudio packages for data visualisation, cont.

ggvis (currently ver0.4): <http://ggvis.rstudio.com>

- from the **ggplot2** team
- create interactive graphics in RStudio and web browser
- top 50 ggplot2 visualisations



Scatterplot with smooth curve and interactive control:



RStudio packages for data visualisation, cont.

Shiny: <https://shiny.posit.co/r/gallery/>

- build professional, interactive visualizations
- equipped with popular web widgets
- can be deployed as independent websites

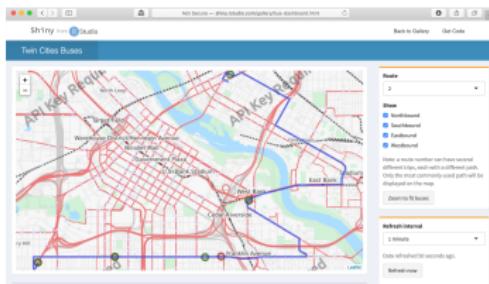


Figure 27: Shiny website example: <https://shiny.posit.co/r/gallery/interactive-visualizations/bus-dashboard/>

RStudio packages for data visualisation, cont.

other packages (默认已安装)

- rmarkdown : create professional documents
- knitr: convert rmarkdown to pdf, html and more ...

Section 4: 如何做作业?

Windows 用户

- 最好新建一个英文帐户，用户名只包括英文字符，比如姓名全拼或英文名：WeihuaChen or JackMa
- 尽量不要更改当前用户，以免数据丢失
- 将以下软件安装到默认目录，包括 R and RStudio

完成作业并提交?

1. 安装必要软件

- R 4.0 or newer
- RStudio 最新版

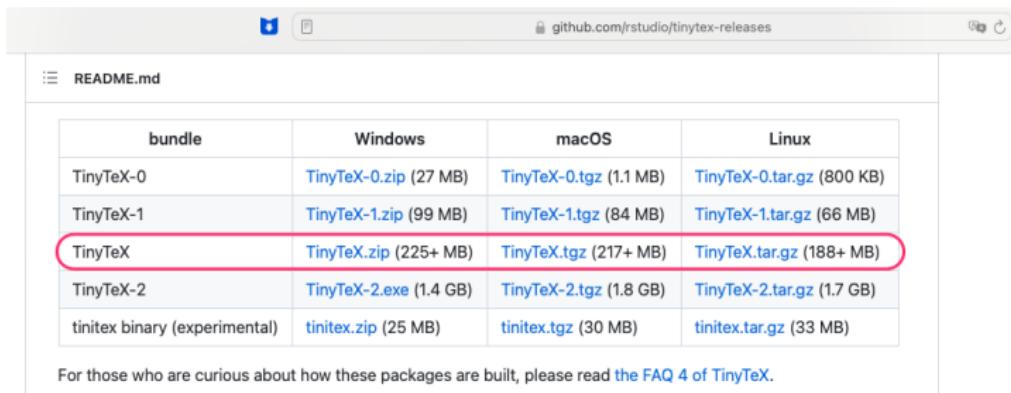
2. 安装 tinytex 包, 辅助将 Rmd 文件转为 PDF

```
install.packages("tinytex");
```

完成作业并提交? cont.

3. 下载 TinyTeX.zip 文件

下载页面: <https://github.com/rstudio/tinytex-releases>
 下载 precompiled TinyTeX.zip (找到对应的平台)。



bundle	Windows	macOS	Linux
TinyTeX-0	TinyTeX-0.zip (27 MB)	TinyTeX-0.tgz (1.1 MB)	TinyTeX-0.tar.gz (800 KB)
TinyTeX-1	TinyTeX-1.zip (99 MB)	TinyTeX-1.tgz (84 MB)	TinyTeX-1.tar.gz (66 MB)
TinyTeX	TinyTeX.zip (225+ MB)	TinyTeX.tgz (217+ MB)	TinyTeX.tar.gz (188+ MB)
TinyTeX-2	TinyTeX-2.exe (1.4 GB)	TinyTeX-2.tgz (1.8 GB)	TinyTeX-2.tar.gz (1.7 GB)
tinitex binary (experimental)	tinitex.zip (25 MB)	tinitex.tgz (30 MB)	tinitex.tar.gz (33 MB)

For those who are curious about how these packages are built, please read the [FAQ 4](#) of TinyTeX.

Figure 28: TinyTeX.zip 下载

注意不要解压缩!!

完成作业并提交? cont.

4. 安装下载的 TinyTex.zip 文件

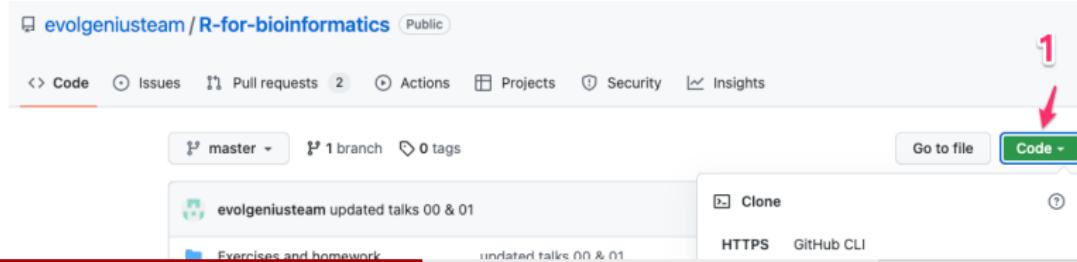
```
tinytex:::install_prebuilt("C:/wchen/Downloads/TinyTex.zip");
```

注: 请将 "C:/wchen/Downloads/TinyTex.zip" 更换为实际下载路径

5. 安装 Github Desktop

- 安装 Github Desktop 并登入, 如无帐户, 则先创建;
- 在浏览器打开网址:

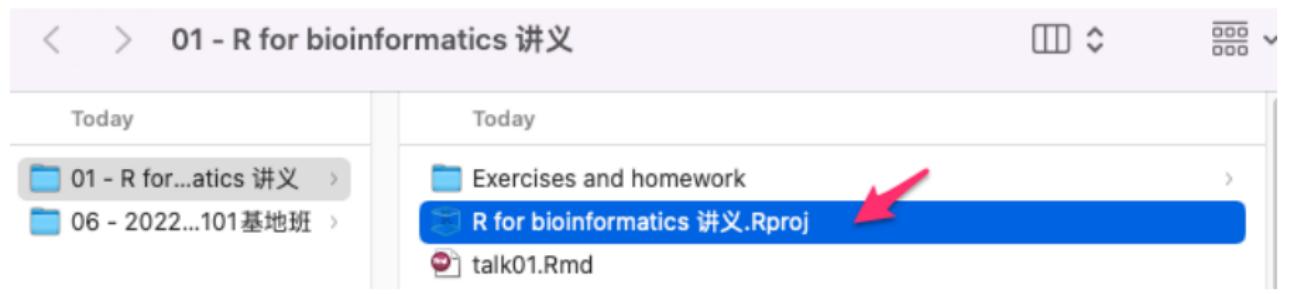
<https://github.com/evolgeniusteam/R-for-bioinformatics>



完成作业并提交? cont.

6. 通过 Github Desktop 下载教程和作业

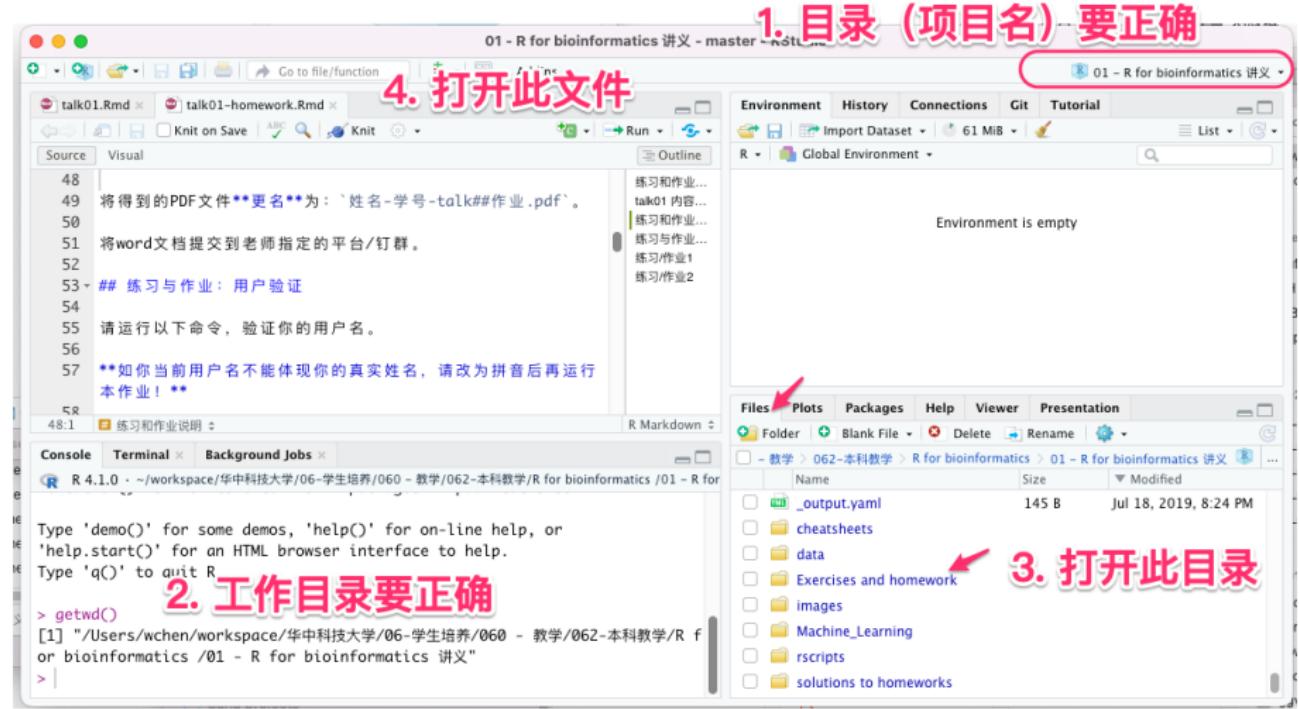
- 下载到默认或指定目录
- 打开下载的目录，找到 R for bioinformatics 讲义.Rproj 文件，双击用 RStudio 打开



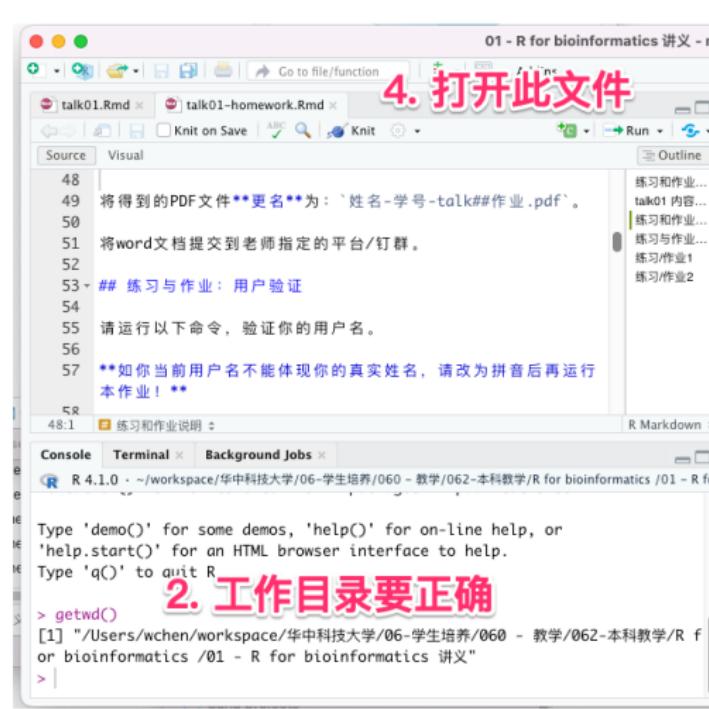
完成作业并提交? cont.

7. 通过 Rstudio 打开相应的 Rproj 文件

1. 目录(项目名)要正确



4. 打开此文件



```

48 将得到的PDF文件**更名**为: `姓名-学号-谈##作业.pdf`。
49 将word文档提交到老师指定的平台/钉群。
50
51 将word文档提交到老师指定的平台/钉群。
52
53 ## 练习与作业：用户验证
54
55 请运行以下命令，验证你的用户名。
56
57 **如你当前用户名不能体现你的真实姓名，请改为拼音后再运行本作业！**
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
279
280
281
282
283
284
285
286
287
288
289
289
290
291
292
293
294
295
296
297
298
299
299
300
301
302
303
304
305
306
307
308
309
309
310
311
312
313
314
315
316
317
318
319
319
320
321
322
323
324
325
326
327
328
329
329
330
331
332
333
334
335
336
337
338
339
339
340
341
342
343
344
345
346
347
348
349
349
350
351
352
353
354
355
356
357
358
359
359
360
361
362
363
364
365
366
367
368
369
369
370
371
372
373
374
375
376
377
378
379
379
380
381
382
383
384
385
386
387
388
389
389
390
391
392
393
394
395
396
397
398
399
399
400
401
402
403
404
405
406
407
408
409
409
410
411
412
413
414
415
416
417
418
419
419
420
421
422
423
424
425
426
427
428
429
429
430
431
432
433
434
435
436
437
438
439
439
440
441
442
443
444
445
446
447
448
449
449
450
451
452
453
454
455
456
457
458
459
459
460
461
462
463
464
465
466
467
468
469
469
470
471
472
473
474
475
476
477
478
479
479
480
481
482
483
484
485
486
487
487
488
489
489
490
491
492
493
494
495
496
497
498
499
499
500
501
502
503
504
505
506
507
508
509
509
510
511
512
513
514
515
516
517
518
519
519
520
521
522
523
524
525
526
527
528
529
529
530
531
532
533
534
535
536
537
538
539
539
540
541
542
543
544
545
546
547
548
549
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
687
688
689
689
690
691
692
693
694
695
696
697
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
787
788
789
789
790
791
792
793
794
795
796
797
797
798
799
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
887
888
889
889
890
891
892
893
894
895
896
897
897
898
899
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
987
988
989
989
990
991
992
993
994
995
996
997
997
998
999
999
1000
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1096
1097
1098
1099
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1188
1189
1189
1190
1191
1192
1193
1194
1195
1196
1196
1197
1198
1199
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1296
1297
1298
1299
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1388
1389
1389
1390
1391
1392
1393
1394
1395
1396
1396
1397
1398
1399
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1488
1489
1489
1490
1491
1492
1493
1494
1495
1496
1496
1497
1498
1499
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1588
1589
1589
1590
1591
1592
1593
1594
1595
1596
1596
1597
1598
1599
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1688
1689
1689
1690
1691
1692
1693
1694
1695
1696
1696
1697
1698
1699
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1748
1749
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1788
1789
1789
1790
1791
1792
1793
1794
1795
1796
1796
1797
1798
1799
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1848
1849
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1888
1889
1889
1890
1891
1892
1893
1894
1895
1896
1896
1897
1898
1899
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1948
1949
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1988
1989
1989
1990
1991
1992
1993
1994
1995
1996
1996
1997
1998
1999
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2088
2089
2089
2090
2091
2092
2093
2094
2095
2096
2097
2097
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2148
2149
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2188
2189
2189
2190
2191
2192
2193
2194
2195
2196
2196
2197
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2248
2249
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2288
2289
2289
2290
2291
2292
2293
2294
2295
2296
2296
2297
2298
2299
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2348
2349
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2388
2389
2389
2390
2391
2392
2393
2394
2395
2396
2396
2397
2398
2399
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408

```

完成作业并提交? cont.

8. 完成作业并提交

- 按 Rmd 文件的要求回答问题或/和提供代码
- Install required packages
 - 比如: tidyverse
- 完成作业
- 用工具栏中的 Knit 按钮生成与 Rmd 同名的 PDF 文件, 将文件名改为: 姓名-学号-talk## 作业.pdf
- 通过钉钉, 在规定的时间内提交
- 现场演示 ...

Section 5: other tools used in our group

Tools used in our group

programming languages

- R
- Python
- Perl

code depository

- Github
- Github desktop
- Atom (general-purpose text editor and more)

web development

- AngularJS
- Vue.js
- Bootstrap
- plot.ly (interactive visualisation)

section 6: Concluding remarks & 作业

Concluding remarks

本期回顾

- 生信分析必备编程语言
- 强大、专业又好用
- RStudio 及其众多扩展包
- To be continued

下期预告

- 数字和字符串
- 整数、小数、逻辑
- 数据类型之间转换；自动规则
- matrix

About the course

- all codes are available at Github:
<https://github.com/evolgeniusteam/R-for-bioinformatics>

作业

- Exercises and homework 目录下 talk01-homework.Rmd 文件；
- 完成时间：见钉群的要求