

[UST 2021] 데이터 사이언스를 위한 R 프로그래밍

한국생명공학연구원 김하성

2021-03-04

Contents

1	Introduction 강의 개요	5
1.1	Goal 강의 목표	5
1.2	This course	5
1.3	Tips	6
1.4	References books 주요/참고 교재	6
1.5	References 참고 자료	6
1.6	Evaluation 평가 세부 항목	7
1.7	Schedule 강의 계획	7
1.8	R Lecture Youtube Link	8
2	R basics	9
2.1	What is R / Rstudio	9
2.2	R / Rstudio installation	12
2.3	Rstudio interface	17
2.4	Keyboard shortcuts	17
2.5	Set working directory	18
2.6	Hello world	18
2.7	Help	19
2.8	RStudio workspace	21
2.9	R packages	22
2.10	Data sets	24
2.11	Cheatsheet	27

Chapter 1

Introduction 강의 개요

- 매주 목요일 강의노트, 동영상 업데이트
- 강사: 한국생명공학연구원 바이오합성연구센터 김하성
- 연락처: 042-860-4372, haseong [at] kribb.re.kr (생명연 연구동 1143)
- 강의조교: 박성군, tjdrns27 [at] kribb.re.kr
- 강의site: <https://greendaygh.github.io/Rprog2021/>

1.1 Goal 강의 목표

- 이공계열 대학원생이 그들의 원활한 실험 설계와 데이터 분석을 위해 범용 프로그램 언어인 R의 사용법과 프로그래밍 기술을 습득할수 있도록 하는데 목표가 있음. 특히 Data scientist를 위해 개발된 tidyverse 패키지 위주의 강의를 진행함.

1.2 This course

- 이 강좌는 온라인 (강의자료, 동영상) 강의를 기본으로 함
- R에 대한 기본 개념과 사용법, 데이터분석 중심의 설명, 필요시 기초 통계 지식 강의
- 매회 강의에는 해당 강의 내용과 관련된 과제가 주어지며 이메일을 통해 문제 풀이를 제출받음
- 상황에 따라 강의자료 및 동영상 업데이트 일정이 조정될 수 있음
- 참고 온라인 강의 특성상 스스로 학습하려는 의지가 강하지 않으면 효과가 없습니다. 매시간 작은 것이라도 배워간다는 마음으로 강의에 임해주시기 바랍니다

1.3 Tips

- 눈으로 이해하지 않고 스스로 실습 필수
- 각 명령줄이 어떻게/왜 작동하는지 이해하기
- 인터넷 검색을 통한 다른 사람의 코드 이해/적용 필요

1.4 References books 주요/참고 교재

- R for Data Science (<https://r4ds.had.co.nz>, <https://github.com/hadley>)
- Hands-On Programming with R (<https://rstudio-education.github.io/hopr/>)
- Using R for Introductory Statistics by John Verzani
 - Free version of 1st Edition
 - * <https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>
 - Second edition
 - * <https://www.crcpress.com/Using-R-for-Introductory-Statistics-Second-Edition/Verzani/p/book/9781466590731>
- Bioinformatics Data Skills by Vince Buffalo (<http://2.droppdf.com/files/5aTvl/bioinformatics-data-skills.pdf>)
- First Course in Statistical Programming with R by Braun and Murdoch (<https://www.cambridge.org/core/books/first-course-in-statistical-programming-with-r/C9F088122AB40517B07FA77F2F0FDE2F>)
- Introductory Statistics with R by Dalgaard (http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/Introductory_Statistics_with_R__2nd_ed.pdf)
- Modern Applied Statistics with S by Venables and Ripley (http://www.bagualu.net/wordpress/wp-content/uploads/2015/10/Modern_Applied_Statistics_With_S.pdf)
- 일반통계학 (영지문화사, 김우철 외)

1.5 References 참고 자료

- <https://resources.rstudio.com/>
- <http://shiny.rstudio.com/tutorial/>
- R 홈페이지 <https://www.r-project.org/>
- Rstudio 홈페이지 <https://www.rstudio.com/>
- Packages for biologists <https://www.bioconductor.org/>
- R 기본 문서들 (소개, 사용, 설치, 운영) <https://cran.r-project.org/manuals.html>
- R ebooks <https://bookdown.org/>
- Cheat Sheets <https://www.rstudio.com/resources/cheatsheets/>

1.6 Evaluation 평가 세부 항목

- 과제 100%
- 성적부여기준: 최종 평균 70점 이상 S, 70점 미만 U 부여
- 과제 채점: 각 과제당 총점 100점 만점 환산 점수 (답이 틀려도 코드가 있으면 가산점)
- 과제 제출일: 수업 자료 배포 후 1주일 (목요일 강의 자료 배포 → 그 다음주 목요일 까지)
- 과제 솔루션 배포: 과제 제출일 마감 이 후 조교 배포
- 감점 기준
 - 과제 제출일 (1주) 이내 제출: 감점 없음
 - 과제 제출일 (1주) 이후 제출: 20점 감점
 - 솔루션 배포 이후 과제 제출: 40점 감점
 - 과제 미제출: 100점 감점
 - 참고로 S/U 판단은 최종 평가시 평균 70정으로 진행함

1.7 Schedule 강의 계획

- 1주차 – R basics / introduction of data
- 2주차 – Univariate data – Summary statistics
- 3주차 – Bivariate data – Correlation / Independence
- 4주차 – Multivariate data – R data structure
- 5주차 – Populations – Families of distributions
- 6주차 – Sampling – Distribution and CLT
- 7주차 – Statistical inference
- 8주차 – Confidence intervals
- 9주차 – Significance test – parametric
- 10주차 – Significance test – non parametric
- 11주차 – Goodness of fit – parametric
- 12주차 – Goodness of fit – non parametric
- 13주차 – Linear regression – basics & simple LR
- 14주차 – Multiple linear regression
- 15주차 – Analysis of variance
- 16주차 – Logistic / Non-linear regression

1.8 R Lecture Youtube Link

1.8.1 Lecture Chapter 1

1.8.2 Lecture 2

1.8.3 Lecture 3

1.8.4 Lecture 4

1.8.5 Lecture 5

1.8.6 Lecture 6

1.8.7 Lecture 7

Chapter 2

R basics

2.1 What is R / Rstudio

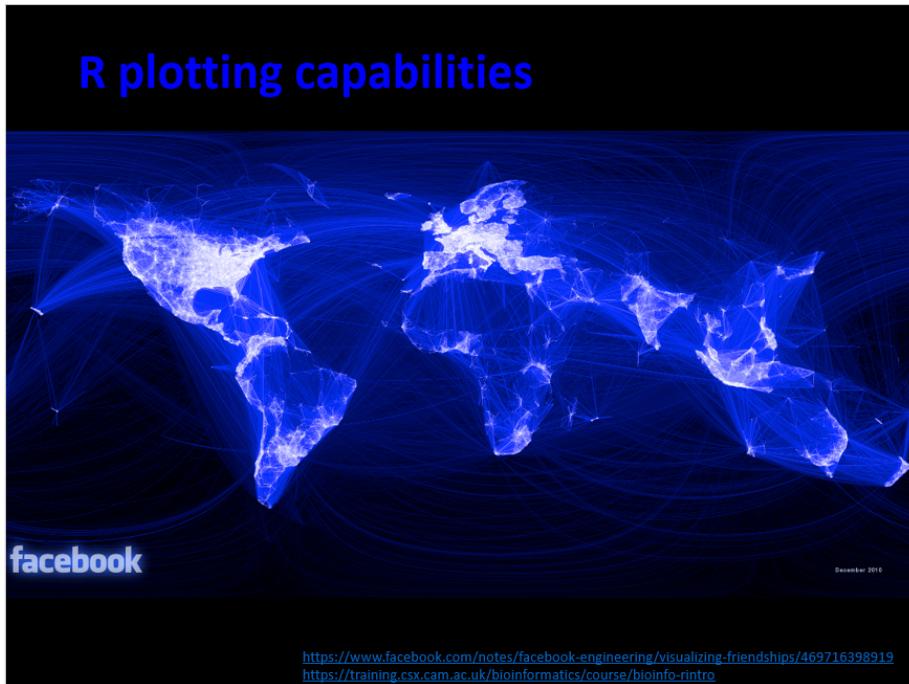


R은 통계나 생물통계, 유전학을 연구하는 사람들 사이에서 널리 사용되는 오픈소스 프로그래밍 언어입니다. Bell Lab에서 개발한 S 언어에서 유래했으며 엄청나게 많은 라이브러리 (다른 사람들이 만들어 놓은 코드)가 있어서 쉽게 가져다 사용할 수 있습니다. R은 복잡한 수식이나 통계 알고리즘을 간단히 구현하고 사용할 수 있으며 C, C++, Python 등 다른 언어들과의 병행 사용도 가능합니다. 2019년 top five language에 랭크 되었으며 이는 빅데이터 증가에 따라 인기가 높아진 것으로 볼 수 있습니다 (참고로 2018년에는 7위).

Rank	Language	Type	Score
1	Python	🌐💻⚙️	100.0
2	Java	🌐📱💻	96.3
3	C	📱💻⚙️	94.4
4	C++	📱💻⚙️	87.5
5	R	💻	81.5
6	JavaScript	🌐	79.4
7	C#	🌐📱💻⚙️	74.5
8	Matlab	💻	70.6
9	Swift	📱💻	69.1
10	Go	🌐💻	68.0

Despite being a much more specialized language than the others, it's maintained its popularity in recent years due to the world being awash in an ever-growing pile of big data. <https://spectrum.ieee.org/computing/software/the-top-programming-languages-2019>

R은 데이터를 통계분석에 널리 사용되는데 이는 데이터를 눈으로 확인하기 위한 visualization이나 벡터 연산 등의 강력한 기능 때문에 점점 더 많은 사람들이 사용하고 있습니다. 기존에는 속도나 확장성이 다른 언어들에 비해 단점으로 지적되었으나 R 언어의 계속적인 개발과 업데이트로 이러한 단점들이 빠르게 보완되고 있습니다. R 사용을 위해서는 R 언어의 코어 프로그램을 먼저 설치하고 그 다음 R 언어용 IDE인 RStudio 설치가 필요합니다.



Interactive web applications

Shiny

from RStudio

Get Started Gallery Articles Reference Deploy Help Contribute

Interact. Analyze. Communicate.

Take a fresh, interactive approach to telling your data story with Shiny. Let users interact with your data and your analysis. And do it all with R.

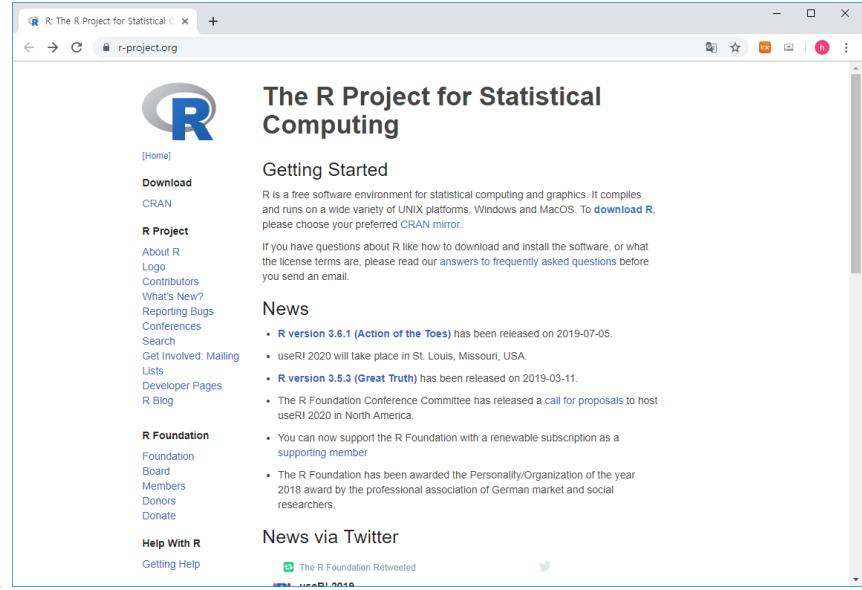
Shiny is an R package that makes it easy to build interactive web apps straight from R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. You can also extend your Shiny apps with CSS themes, htmlwidgets, and JavaScript actions.

ICGC PANCRETIC CANCER (DUCTAL ADENOCARCINOMA) - GENOME VIEWER

Cohort Top ClinVar Gene Summary:					
HMGCR	Chr 19	Start 105801770	End 105801770	From T	To Consequence
TSHZ3	Chr 17	Start 215171410	End 215171410	A	missense_variant
KIAA0512	Chr 12	Start 231092120	End 231092120	C	missense_variant
SMACDC	Chr 19	Start 121441100	End 121441100	G	missense_variant
TSHZ3	Chr 17	Start 215171110	End 215171110	A	missense_variant
KIAA0512	Chr 12	Start 231092120	End 231092120	C	missense_variant
SMACDC	Chr 19	Start 121441100	End 121441100	T	downstream_gene_variant
TSHZ3	Chr 17	Start 215171410	End 215171410	G	downstream_gene_variant

2.2 R / Rstudio installation

- R 사이트에 접속 후 (<https://www.r-project.org/>) 좌측 메뉴 상단에 위치한



CRAN 클릭.

CRAN Mirrors	
The Comprehensive R Archive Network is available at the following URLs, please choose a location or 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 .	
O-Cloud	Automatic redirection to servers we support
https://cloud.r-project.org/	RStudio
Algeria	University of Science and Technology of Oran
https://cran.usthb.dz/	Universidad Nacional de La Plata
Argentina	CSIRO
http://mirror.fcaglo.unlp.edu.ar/CRAN/	AARNET
Australia	School of Mathematics and Statistics
https://cran.csiro.au/	Curtin University of Technology
https://mirror.aarnet.edu.au/pub/CRAN/	Wirtschaftsuniversität Wien
https://cran.ms.unimelb.edu.au/	Patrick Wessa
https://cran.curtin.edu.au/	Ghent University Library
Austria	Universidade Federal do Paraná
https://cran.wu.ac.at/	Oswaldo Cruz Foundation, Rio de Janeiro
Belgium	University of São Paulo, São Paulo
https://www.freestatistics.org/cran/	University of São Paulo, Piracicaba
https://lib.ugent.be/CRAN/	
Brazil	
https://cran.r-c3sl.ufpr.br/	
https://cran.fiocruz.br/	
https://vps.fmvz.usp.br/CRAN/	
https://brieger.esalq.usp.br/CRAN/	
Bulgaria	

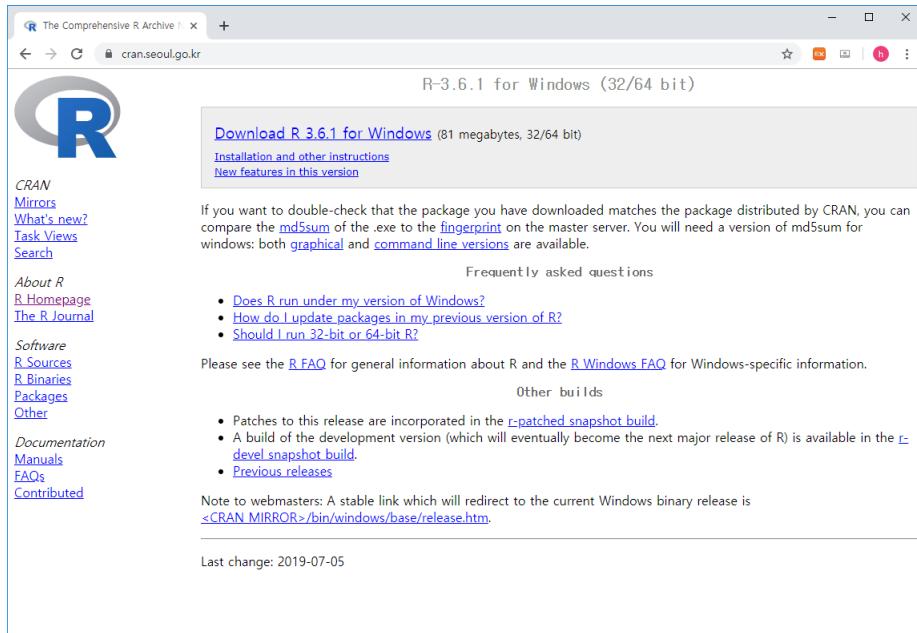
- 미러 사이트 목록에서 Korea의 아무 사이트나 들어감

The screenshot shows the CRAN website for the 'Download and Install R' section. The left sidebar contains links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, and Contributed. The main content area has a large 'R' logo at the top. Below it, there's a heading 'Subdirectories:' with four entries: 'base', 'contrib', 'old_contrib', and 'Rtools'. Each entry has a brief description. At the bottom of the page, there are sections for 'Questions About R' and a link to 'If you have questions about R like how to download and install the software'.

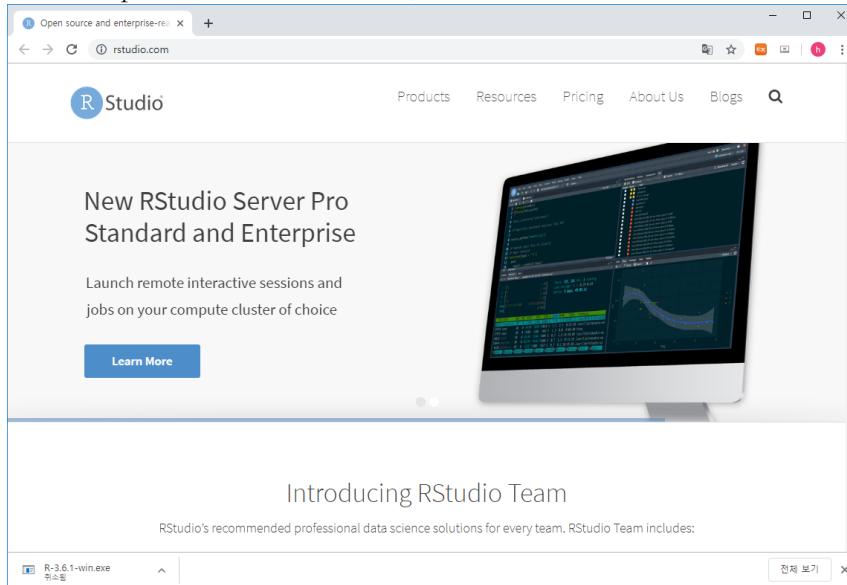
- Download R for Windows를 클릭 후 base 링크 들어가서

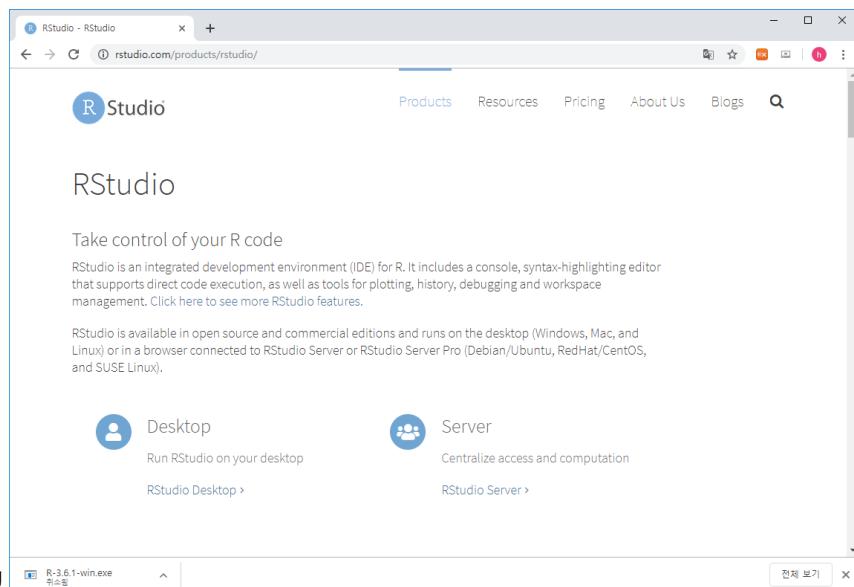
The screenshot shows the 'R for Windows' subdirectory page. It lists four subdirectories: 'base', 'contrib', 'old_contrib', and 'Rtools'. Each subdirectory has a brief description. A note at the bottom says 'Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.' It also mentions the 'RFAQ' and 'R for Windows FAQ'.

- Download R 3.6.3 for Windows 링크로 실행 프로그램 다운로드 (2020.3 현재 R 버전은 3.6.3). 로컬 컴퓨터에 Download 된 R-3.6.3-win.exe 를 실행하고 설치 프로그램의 지시에 따라 R 언어 소프트웨어 설치를 완료합니다.

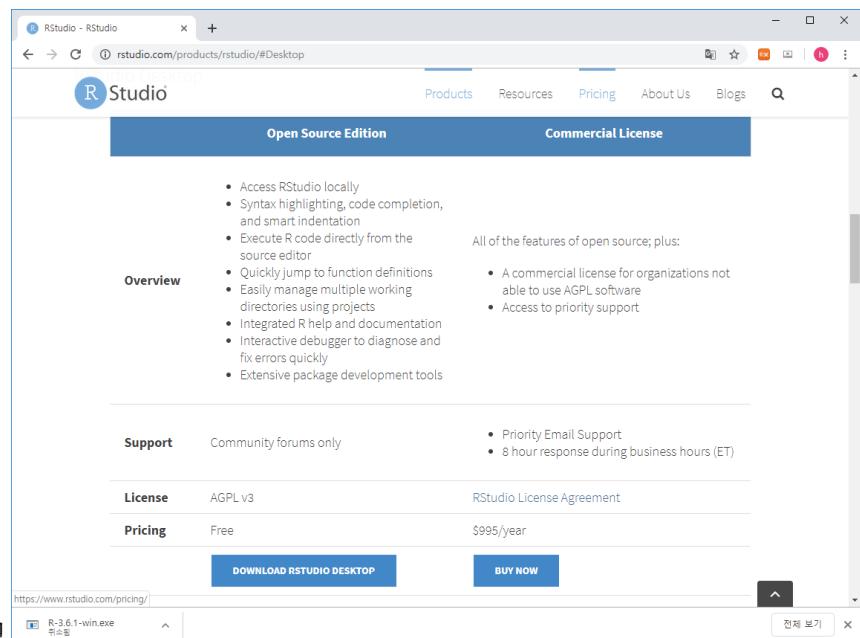


- Rstudio는 R 언어를 위한 오픈소스 기반 통합개발환경(IDE)으로 R 프로그래밍을 위한 편리한 기능들을 제공해 줍니다. 다음 사이트에 접속 (<https://www.rstudio.com/>), 상단의 Products > RStudio 클릭

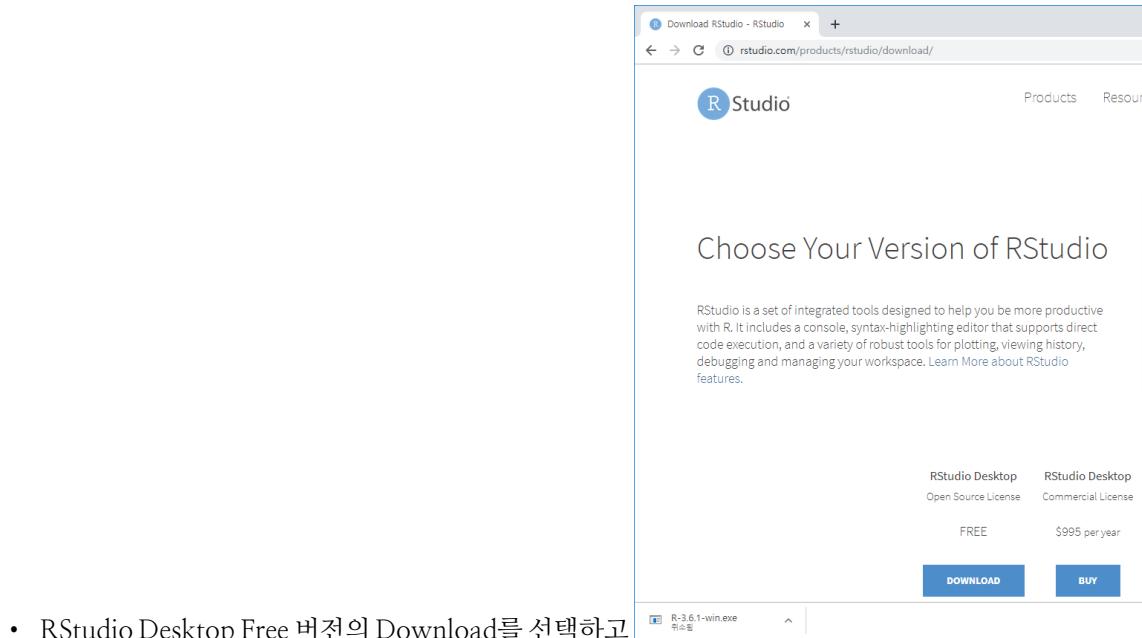




- RStudio Desktop 선택



- Download RStudio Desktop 클릭



- RStudio Desktop Free 버전의 Download를 선택하고

- Download RStudio for Windows (2020.03 현재 version 1.2.5033) 클릭, 다운로드. 로컬 컴퓨터에 다운로드된 RStudio-1.2.5033.exe를 실행하고

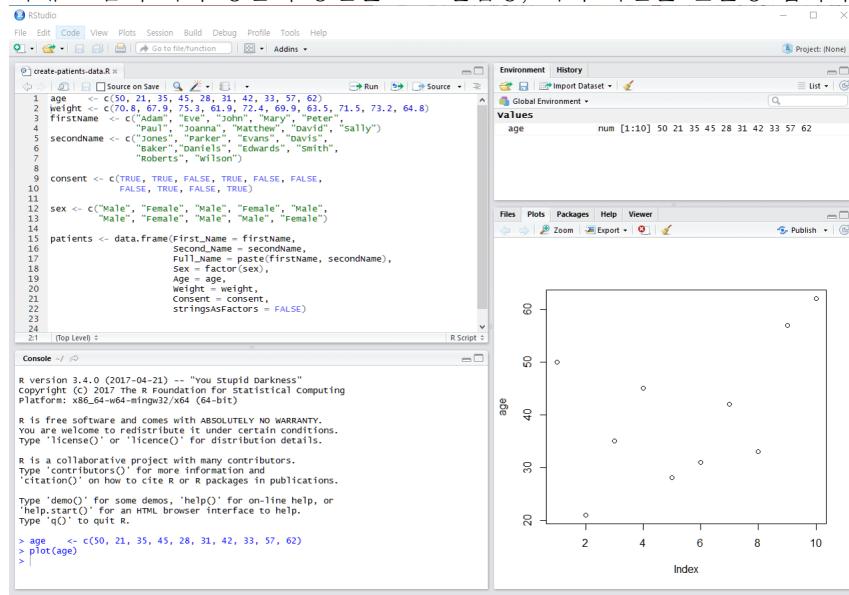
The screenshot shows the 'Release Notes' for RStudio Desktop 1.2.1335. It includes sections for 'Installers' and 'Zip/Tarballs'. The 'Installers' section lists various operating system versions with their file names, sizes, dates, and MD5 checksums. The 'Zip/Tarballs' section lists the same files in a compressed format. Both sections include a note about RStudio requiring R 3.0.1+ and a link to download R if needed.

Installers	Size	Date	MD5
RStudio 1.2.1335 - Windows 7+ (64-bit)	126.9 MB	2019-04-08	d0e2470f1ff8ef4cd35a669aa323a213d
RStudio 1.2.1335 - macOS 10.12+ (64-bit)	121.1 MB	2019-04-08	6c570b0e21445b3f7c48c284ce299ee
RStudio 1.2.1335 - Ubuntu 14/Debian 8 (64-bit)	92.2 MB	2019-04-08	c1b07d0511469abf582919b183eee83
RStudio 1.2.1335 - Ubuntu 16 (64-bit)	99.3 MB	2019-04-08	c142d99c210257fb10d18c045ff13c7
RStudio 1.2.1335 - Ubuntu 18/Debian 10 (64-bit)	100.4 MB	2019-04-08	71a8d1990c0d97939804b46cfb0aea75
RStudio 1.2.1335 - Fedora 19/Red Hat 7 (64-bit)	114.1 MB	2019-04-08	29666feff99699a912971ab6545f256a7a
RStudio 1.2.1335 - Debian 9 (64-bit)	100.6 MB	2019-04-08	1e32d4d6f6e216f086a81ca82ef65a91
RStudio 1.2.1335 - OpenSUSE 15 (64-bit)	101.6 MB	2019-04-08	2795a63c7efdb0e2aa2dae6ba09a81e5
RStudio 1.2.1335 - SLES/OpenSUSE 12 (64-bit)	94.4 MB	2019-04-08	c65424b06ef6737279d982db9eefcae

설치 가이드에 따라 설치 완료합니다.

2.3 Rstudio interface

- 아래 그림의 좌측 상단의 공간은 코드편집창, 좌측 하단은 콘솔창입니다.



2.4 Keyboard shortcuts

- 참고사이트
 - <https://support.rstudio.com/hc/en-us/articles/200711853-Keyboard-Shortcuts>
 - Tools → Keyboard shortcut Quick Reference (**Alt + Shift + K**)
- 코드편집창 이동 (**Ctrl+1**) 콘솔창 이동(**Ctrl+2**)
 - 한 줄 실행 (**Ctrl+Enter**)
 - 주석처리 (**Ctrl + Shift + C**)
 - 또는 #으로 시작하는 라인
- 실습
 - 코드편집창에서 다음 입력

```

Source on Save | 🔎 | 🖊️ | 📋
  1 x <- 10
  2 y <- 20
  3

```

- 단축키 Ctrl + enter로 코드 실행
- 단축키 Ctrl + 2로 커서 콘솔창으로 이동
- x값 x+y값 확인
- 단축키 Ctrl + 1로 코드편집창 이동
- 단축키 Ctrl + Shift + C 사용

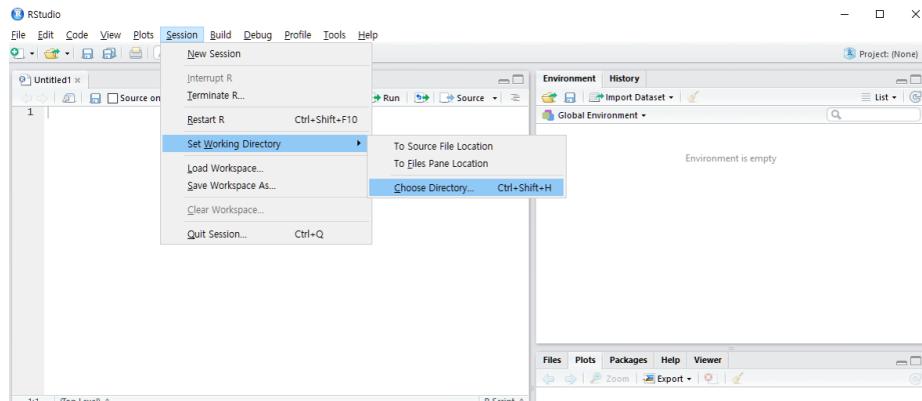
```
# x <- 10
# y <- 20
```

2.5 Set working directory

시작 전 항상 작업 디렉토리 설정. 예를 들어 c: 아래 새로운 디렉토리 rprog2020 을 만들고 작업공간으로 설정

```
getwd()
dir()
setwd("C:\\rprog2020")
getwd()
dir()
```

또는 아래와 같이 RStudio 메뉴에서 설정



2.6 Hello world

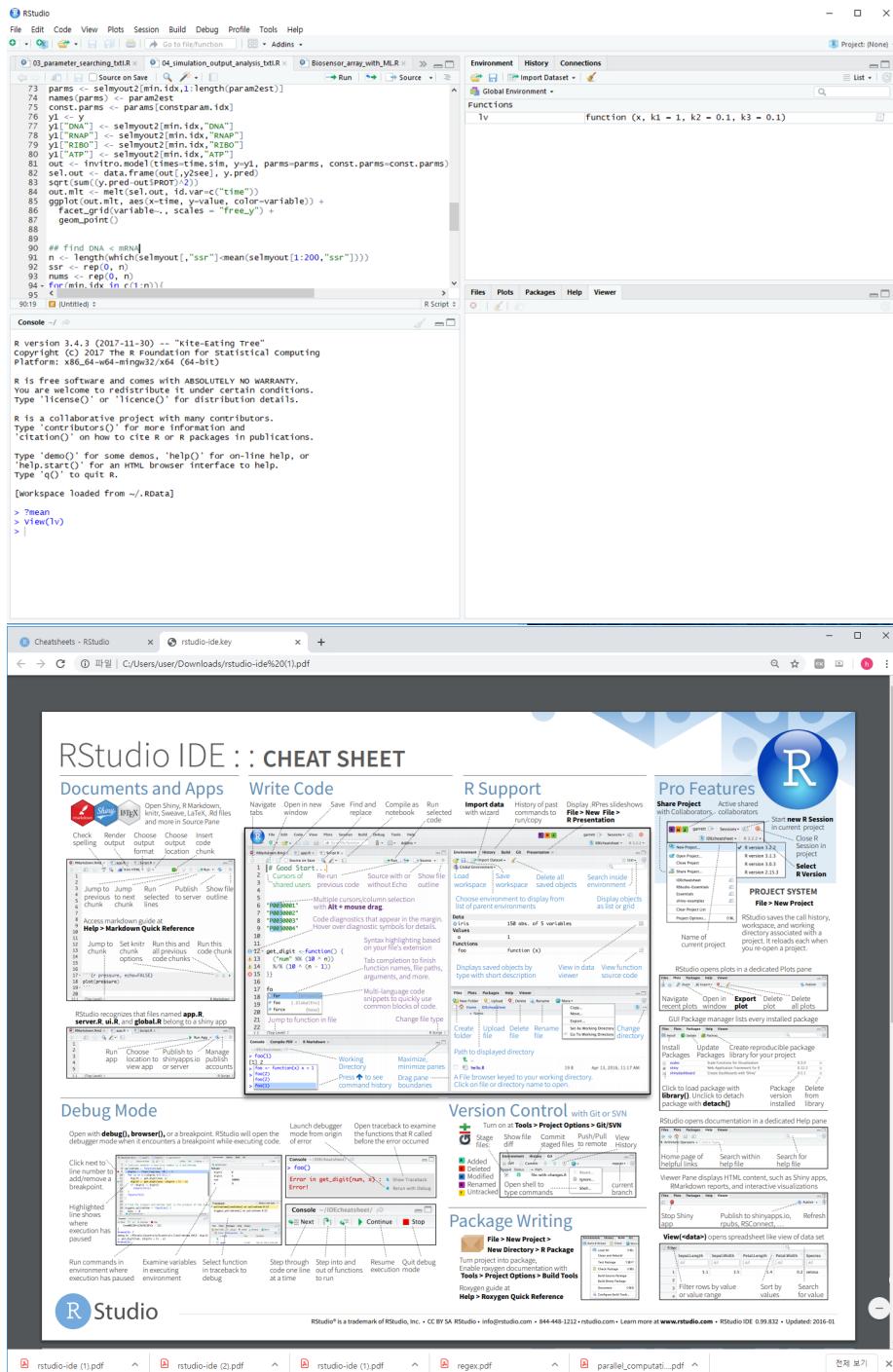
```
mystring <- "Hello \n world!"
cat(mystring)
print(mystring)
```

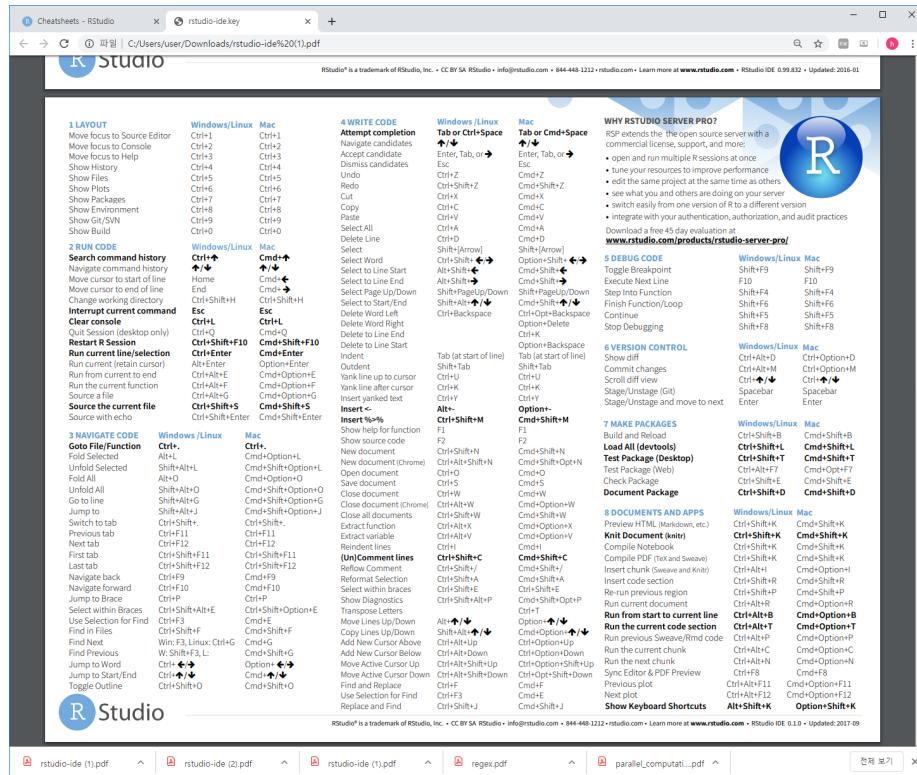
2.7 Help

R의 장점 중 하나로 방대한 양의 도움말 페이지가 제공됩니다. ? 명령을 사용하면 되며 구글이나 웹에서도 도움을 얻을 수 있습니다.

```
?cat  
?print  
?mean  
help("mean")  
example("mean")  
help.search("mean")  
help(package="MASS")
```

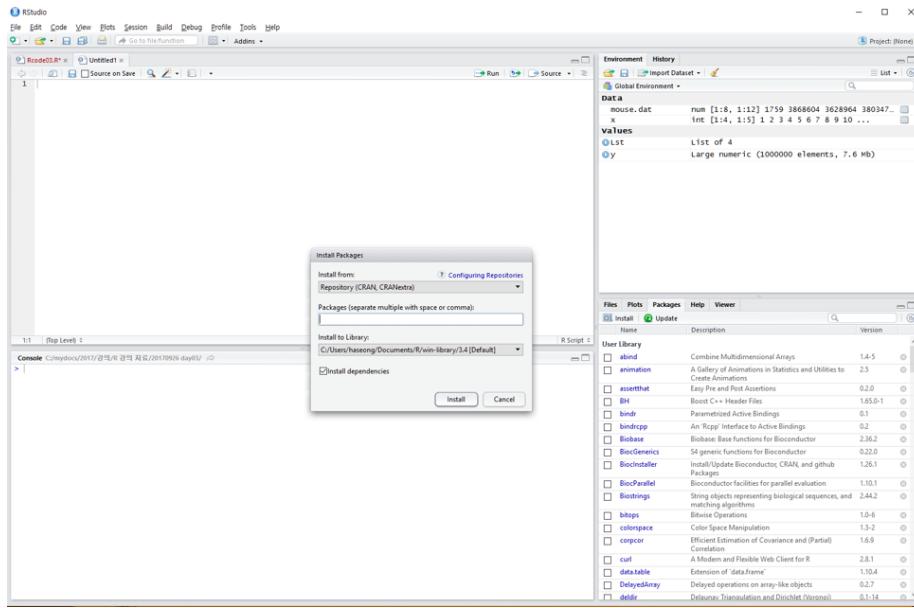

2.8 RStudio workspace





2.9 R packages

R은 “package”라 불리우는 다양한 함수 라이브러리를 사용할 수 있습니다. 예를 들어 `sum()`이나 `sd()`와 같은 함수는 `stats`이라는 패키지에서 구현된 함수입니다. 이러한 패키지는 인터넷의 `repository`에서 구할 수 있으며 대표적으로 The Comprehensive R Archive Network (CRAN) <http://cran.r-project.org/web/views/> 와 생물학자를 위한 Bioconductor specialised in genomics http://www.bioconductor.org/packages/release/BiocViews.html#_Software 가 있습니다. 이러한 패키지의 설치는 아래와 같이 RStudio를 이용하거나 콘솔창에서 `install.packages()` 함수를 이용할 수 있습니다.



- UsingR package installation

Packages → Install

Packages → Install

The 'Install Packages' dialog shows 'UsingR' selected for installation. The 'Console' tab shows the R command and its execution output:

```
> install.packages("usingR")
Installing package into 'C:/Users/user/Documents/R/win-library/3.4'
(as 'lib' is unspecified)
also installing the dependency 'Histdata'

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.4/HistData_0.8-4.zip'
Content type 'application/zip' length 359785 bytes (351 KB)
downloaded 351 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.4/UsingR_0.6.zip'
Content type 'application/zip' length 2081603 bytes (2.0 MB)
downloaded 2.0 MB

package 'HistData' successfully unpacked and MD5 sums checked
package 'usingR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/users\user\AppData\Local\Temp\RtmpwdxEq7\downloaded_package
>
```

The 'Packages' tab shows the installed packages: UsingR, ggplot2, munsell, rbsml, and stats4. The 'Console' tab shows the library loading command and its output:

```
> library("usingR", lib.loc = "~/R/win-library/3.4")
필요한 패키지를 로딩중입니다: MASS
필요한 패키지를 로딩중입니다: Histdata
필요한 패키지를 로딩중입니다: Hmisc
필요한 패키지를 로딩중입니다: lattice
필요한 패키지를 로딩중입니다: survival
필요한 패키지를 로딩중입니다: Formula
필요한 패키지를 로딩중입니다: ggplot2

다음의 패키지를 부착합니다: 'Hmisc'

The following objects are masked from 'package:base':
  format.pval, units

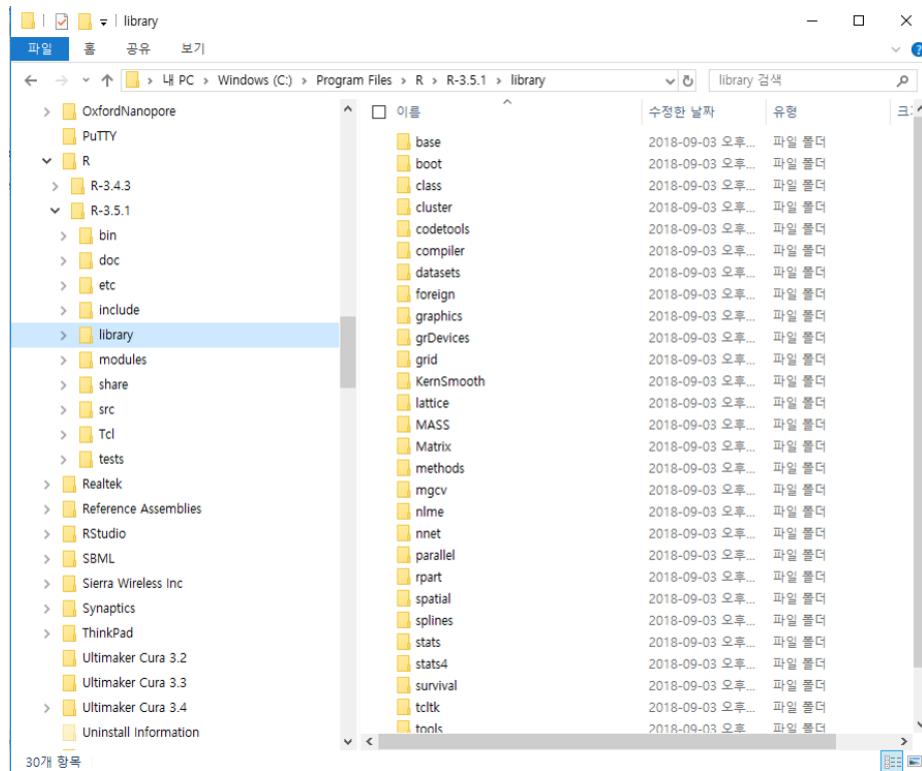
다음의 패키지를 부착합니다: 'usingR'
```

패키지를 설치하고 사용하기 위해서는 `library()` 함수를 사용해서 미리 loading 해 두어야 합니다. 한 번 로딩으로 작업 세션이 끝날때까지 관련된 함수를 사용할 수 있으나 R 세션이나 RStudio를 재시작 할 경우 다시 로딩해야 사용할 수 있습니다.

```
library(UsingR)
```

- R 설치 디렉토리
- R 패키지 설치 디렉토리

```
.libPaths()  
path.package()
```



2.10 Data sets

대부분의 패키지는 함수와 함께 관련된 도움말, 예제, 그리고 데이터셋을 같이 제공해 줍니다. `library()` 함수를 사용할 때 자동으로 같이 로딩이 되며 `data()` 함수를 사용해서 사용자 작업공간에 복사본을 만들어서 사용할 수 있습니다.

```
head(rivers)  
length(rivers)  
class(rivers)
```

```
data(rivers)
data(package="UsingR")
library(HistData)
head(Cavendish)
str(Cavendish)
head(Cavendish$density2)
data(package="HistData")
```


2.11 Cheatsheet

Base R Cheat Sheet

Getting Help

- `?mean`: Get help of a particular function.
- `help.search('weighted mean')`: Search the help files for a word or phrase.
- `help(package = 'dplyr')`: Find help for a package.
- `More about an object`
- `str(x)`: Get a summary of an object's structure.
- `class(x)`: Find the class an object belongs to.

Using Packages

- `install.packages('dplyr')`: Download and install a package from CRAN.
- `library(dplyr)`: Load the package into the session, making all its functions available to use.
- `dplyr::select`: Use a particular function from a package.
- `data(iris)`: Load a built-in dataset into the environment.

Working Directory

- `getwd()`: Find the current working directory (where inputs are found and outputs are sent).
- `setwd("C:/file/path")`: Change the current working directory.
- Use projects in RStudio to set the working directory to the folder you are working in.

Vectors		Programming	
Creating Vectors		For Loop	
<code>c(2, 4, 6)</code>	2 4 6 Join elements into a vector	<code>for (variable in sequence){</code>	<code>do something</code>
<code>2:6</code>	2 3 4 5 6 An integer sequence	<code>}</code>	<code>}</code>
<code>seq(2, 3, by=0.5)</code>	2.0 2.5 3.0 A complex sequence	Example	
<code>rep(1:2, times=3)</code>	1 2 1 2 1 2 Repeat a vector	<code>for (i in 1:4){</code>	
<code>rep(1:2, each=3)</code>	1 1 1 2 2 2 Repeat elements of a vector	<code>j <- i + 10</code>	
		<code>print(j)</code>	
		<code>}</code>	
Vector Functions		While Loop	
<code>sort(x)</code>	Return x sorted.	<code>if (condition){</code>	<code>do something</code>
<code>rev(x)</code>	Return x reversed.	<code>} else {</code>	<code>do something different</code>
<code>table(x)</code>	See counts of values.	<code>}</code>	<code>}</code>
		Example	
		<code>while (i < 5){</code>	
		<code>print(i)</code>	
		<code>i <- i + 1</code>	
Selecting Vector Elements		Functions	
By Position		Example	
<code>x[4]</code>	The fourth element.	<code>function_name <- function(var){</code>	
<code>x[-4]</code>	All but the fourth.	<code>do something</code>	
<code>x[2:4]</code>	Elements two to four.	<code>return(new_variable)</code>	
<code>x[-(2:4)]</code>	All elements except two to four.		
<code>x[c(1, 5)]</code>	Elements one and five.		
<code>x[x == 10]</code>	Elements which are equal to 10.		
<code>x[x < 0]</code>	All elements less than zero.		
<code>x[x %in% c(2, 5)]</code>	Elements in the set 1, 2, 5.		
By Value		Example	
<code>x['apple']</code>	Element with name 'apple'.	<code>square <- function(x){</code>	
		<code>squared <- x*x</code>	
		<code>return(squared)</code>	
Reading and Writing Data			
Input		Output	
<code>df <- read.table('file.txt')</code>		<code>write.table(df, 'file.txt')</code>	Read and write a delimited text file.
<code>df <- read.csv('file.csv')</code>		<code>write.csv(df, 'file.csv')</code>	Read and write a comma separated value file. This is a special case of readable/writable.
<code>Load('file.Rdata')</code>		<code>save(df, file = 'file.Rdata')</code>	Read and write an R data file, a file type special for R.
Conditions			
<code>a == b</code>	Are equal	<code>a > b</code>	Greater than
<code>a != b</code>	Not equal	<code>a < b</code>	Less than
<code>a <= b</code>	Greater than or equal to	<code>a >= b</code>	Less than or equal to
<code>is.na(a)</code>	Is missing	<code>is.null(a)</code>	Is null

RStudio® is a trademark of RStudio, Inc. • CC BY Mhairi McNeill • mhairimcneill@gmail.com

Learn more at [web page](#) or [vignette](#) • package version • Updated: 3/15

Types		Matrices		Strings	
Converting between common data types in R. Can always go from a higher value in the table to a lower value.		<code>m <- matrix(x, nrow = 3, ncol = 3)</code> Create a matrix from x.		Also see the <code>string</code> package	
<code>as.logical</code>	TRUE, FALSE, TRUE Boolean values (TRUE or FALSE)	<code>m[2,]</code> - Select a row	<code>t(m)</code>	<code>paste(x, y, sep = " ")</code>	Join multiple vectors together.
<code>as.numeric</code>	1, 0, 1 Integers or floating point numbers	<code>m[, 1]</code> - Select a column	<code>Transpose</code>	<code>paste(x, collapse = "")</code>	Join elements of a vector together.
<code>as.character</code>	'1', '0', '1' Character strings. Generally preferred to factors.	<code>m[2, 3]</code> - Select an element	<code>m %*% n</code>	<code>grep(pattern, x)</code>	Find regular expression matches in x.
<code>as.factor</code>	'1', '0', '1' levels: '1', '0' Character strings with present levels. Needed for some statistical models.		<code>Matrix Multiplication</code>	<code>gsub(pattern, replace, x)</code>	Replace matches in x with a string.
			<code>solve(m, n)</code>	<code>toupper(x)</code>	Convert to uppercase.
			Find x:m * n	<code>tolower(x)</code>	Convert to lowercase.
				<code>nchar(x)</code>	Number of characters in a string.
Lists				Factors	
<code>l <- list(x = 1:5, y = c('a', 'b'))</code> A list is a collection of elements which can be of different types.				<code>factor(x)</code>	<code>cut(x, breaks = 4)</code>
<code>l[[2]]</code>	Second element of l.	<code>l[1]</code>	New list with only the first element.	Turn a vector into a factor. Can set the levels of the factor and the order.	Turn a numeric vector into a factor by 'cutting' into sections.
		<code>l\$x</code>	Element named x.		
		<code>l[('y')]</code>	New list with only element named y.		
Also see the <code>dplyr</code> package				Statistics	
				<code>lm(y ~ x, data=df)</code>	<code>t.test(x, y)</code>
				<code>glm(y ~ x, data=df)</code>	<code>prop.test</code>
				<code>summary</code>	Test for a difference between means.
				<code>pairwise.t.test</code>	Test for a difference between proportions.
				<code>aoov</code>	Analysis of variance.
Data Frames				Distributions	
List subsetting				<code>Random Variates</code>	<code>Normal</code>
<code>df\$x</code>	<code>df[[2]]</code>	<code>nrow(df)</code>	<code>rnorm</code>	<code>Density Function</code>	<code>dnorm</code>
		<code>ncol(df)</code>	<code>rpois</code>	<code>Cumulative Distribution</code>	<code>pnorm</code>
		<code>dim(df)</code>	<code>dpois</code>	<code>Quantile</code>	<code>qnorm</code>
			<code>Binomial</code>		
			<code>rbinom</code>		
			<code>dbinom</code>		
			<code>pbinom</code>		
			<code>Uniform</code>		
			<code>runif</code>		
			<code>dunif</code>		
			<code>punif</code>		
			<code>qunif</code>		
Plotting				Dates	
				<code>plot(x)</code>	<code>Also see the <code>ggplot2</code> package</code>
				<code>plot(x, y)</code>	
				<code>hist(x)</code>	
				<code>See the <code>lubridate</code> package</code>	

RStudio® is a trademark of RStudio, Inc. • CC BY Mhairi McNeill • mhairimcneill@gmail.com

Learn more at [web page](#) or [vignette](#) • package version • Updated: 3/15

이 저작물은 크리에이티브 커먼즈 저작자표시-비영리-변경금지 4.0 국제 라이선스에 따라 이용할 수 있습니다.