Day17; 20220927(R, 파이썬 시작)

날짜	
유형	
태그	

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/d80ffa7d-a64f-4e03-9ca8-bab86d6e9a7f/01._R_%EC%84%A4%EC%B9%98_%EB%B0%8F_%EA%B0%9C%EC%9A%94.pdf

빅데이터 분석에 관하여 이슈가 되고 있는데 R, 파이썬은 대용량의 빅데이터를 처리하지는 못한다.

대용량 데이터 처리 핵심기술 - Hadoop

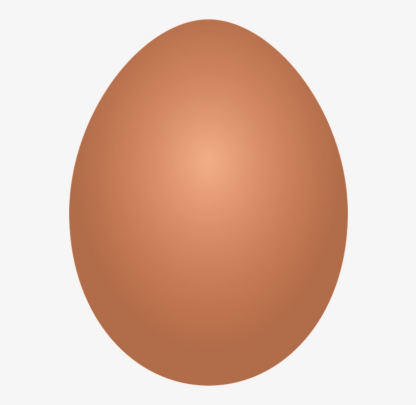
하둡은 길게 배워야 하고 품질도 떨어진다.

그래서 머신러닝, 딥러닝까지만 살펴봄.

Data

수집 → 저장 → 처리(가공) - 분석

R



통계학에서부터 시작.

분석이라는 것은 아주 오래된 학문이다.

통계학 기반으로 데이터를 분석하고 싶을 때 사용.

이전에는 SPSS 등 돈을 주고 사용했음.

다양한 통계학에서부터 예측분석: 머신러닝, 딥러닝도 나온 것이다.

Scikit - learn

- 딥러닝

  tensorFlow

  CNN, RNN

  openCV(Computer Vision)를 통해 CNN을 이해할 수 있다.

분석용 Library(전처리proprocess)

---

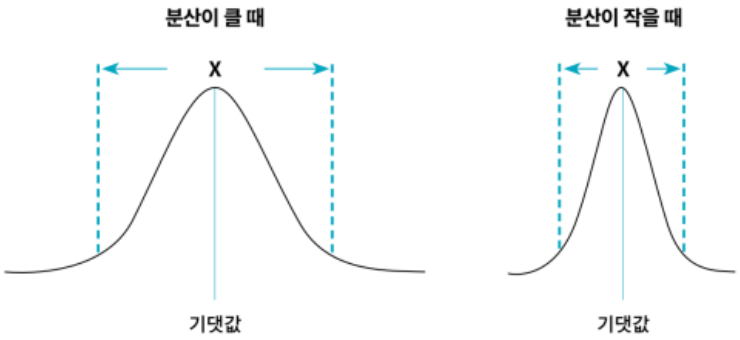
② 분산

- 분산: 데이터들이 중심에서 얼마나 떨어져 있는지를 알아보기 위한 측도다. 관측값에서 평균을 뺀 값을 제곱하고, 그것을 모두 더한 후 전체 개수로 나눠서 구한다. 즉, 차이값의 제곱의 평균이다.
- 확률변수의 분산: 확률변수가 취할 수 있는 값들이 그 중심(모평균)에서 얼마나 떨어져 있는지를 측정하는 측도다. 값이 크면 클수록 확률 X값이 기댓값에서 멀리 떨어져 있을 수 있다.
- 확률변수의 분산은 확률변수의 평균과 마찬가지로 이미 측정되어 있는 값에 대한 것이 아니고 앞으로 측정 또는 관측될 가능성이 있는 값들에 대한 측도다.

이산확률변수의 분산: $Var(X)=E[(X-E(X))^2]=E(X^2)-E(X)^2$

연속확률변수의 분산: $Var(X)=\int(x-E(X))^2 f(x)\,dx=\int x^2 f(x)\,dx-\int xf(x)\,dx^2$

【 분산 】



---

download

[Home]

Download

CRAN

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.

https://cran.r-project.org/mirrors.html

https://ftp.yz.yamagata-u.ac.jp/pub/cran/

Korea

https://ftp.harukasan.org/CRAN/

https://cran.yu.ac.kr/

https://cran.seoul.go.kr/

https://cran.biodisk.org/

Malaysia

The Comprehensive R Archive Network

https://cran.biodisk.org/

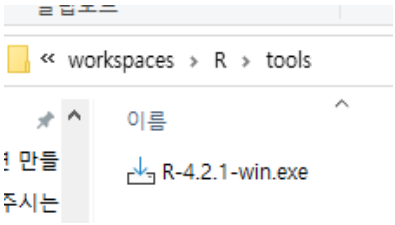
Subdirectories:

base — Binaries for base distribution. This is what you want to **install R for the first time**.

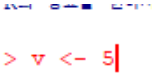
Download R-4.2.1 for Windows (79 megabytes, 64 bit)

README on the Windows binary distribution

New features in this version



변수 선언은 이름만 씀



R은 컴파일이 필요없음

Interaction programming language
Interpreter 언어 - HTML, R

50~60년대에는 딥러닝이 두 번의 시련을 겪음.

컴퓨터가 너무 느림.

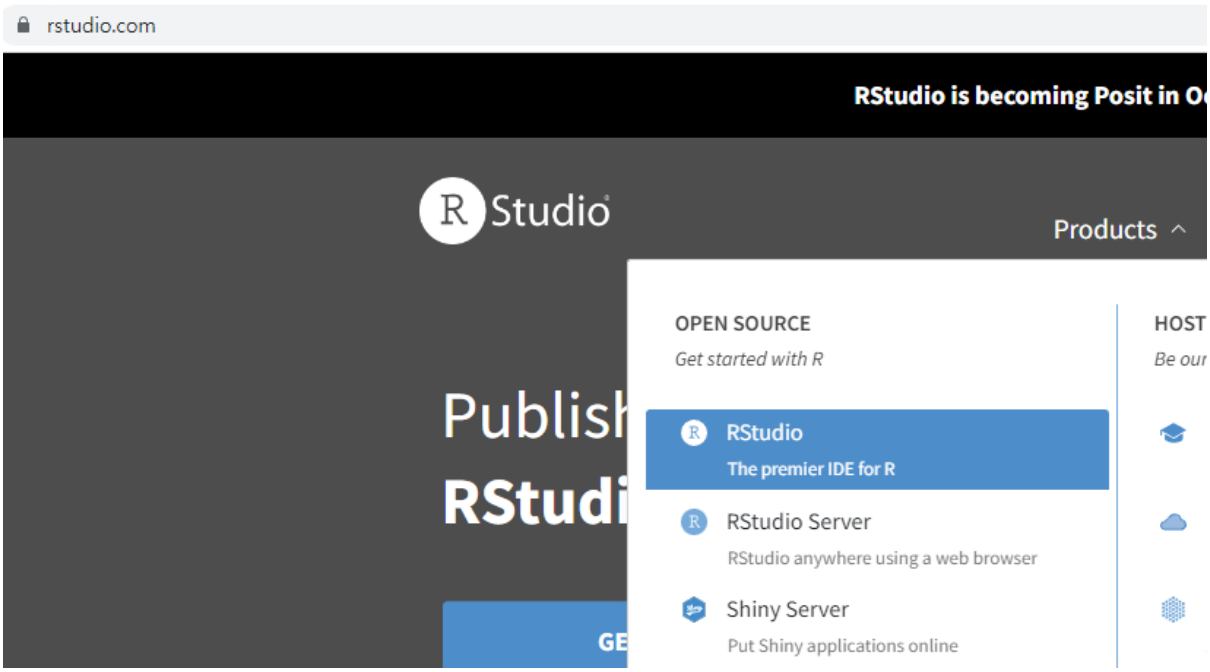
그래서 결과까지 일주일이 걸림.

근데 오타였으면... 또 일주일

RStudio | Open source & professional software for data science teams

Take control of your R and Python code An integrated development environment for R and Python, with a console, syntax-highlighting editor that supports direct code execution, and tools for plotting, history, debugging and

https://www.rstudio.com/



RStudio Desktop

Open Source Edition	
Overview	• Access RStudio locally • Syntax highlighting, code completion, and smart indentation • Execute R code directly from the source editor • Quickly jump to function definitions • View content changes in real-time with the Visual Markdown Editor • Easily manage multiple working directories using projects • Integrated R help and documentation • Interactive debugger to diagnose and fix errors • Extensive package development tools
Support	Community forums only
License	AGPL v3
Pricing	Free

DOWNLOAD RSTUDIO DESKTOP

RStudio Desktop

Open Source License

Free

DOWNLOAD

Learn more

RStudio Desktop 2022.07.2+576 - Release Notes

1. Install R. RStudio requires R 3.3.0+.

2. Download RStudio Desktop. Recommended for your system:

DOWNLOAD RSTUDIO FOR WINDOWS
2022.07.2+576 | 190.49MB

Requires Windows 10/11 (64-bit)

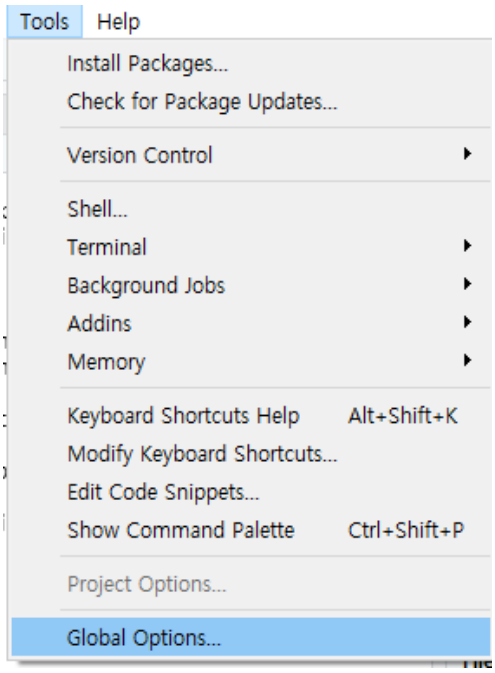
All Installers

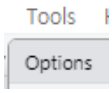
Linux users may need to import RStudio's public code-signing key prior to installation, depending on the operating system's security poli

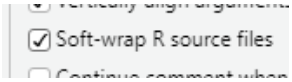
RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an older version of RStudio.

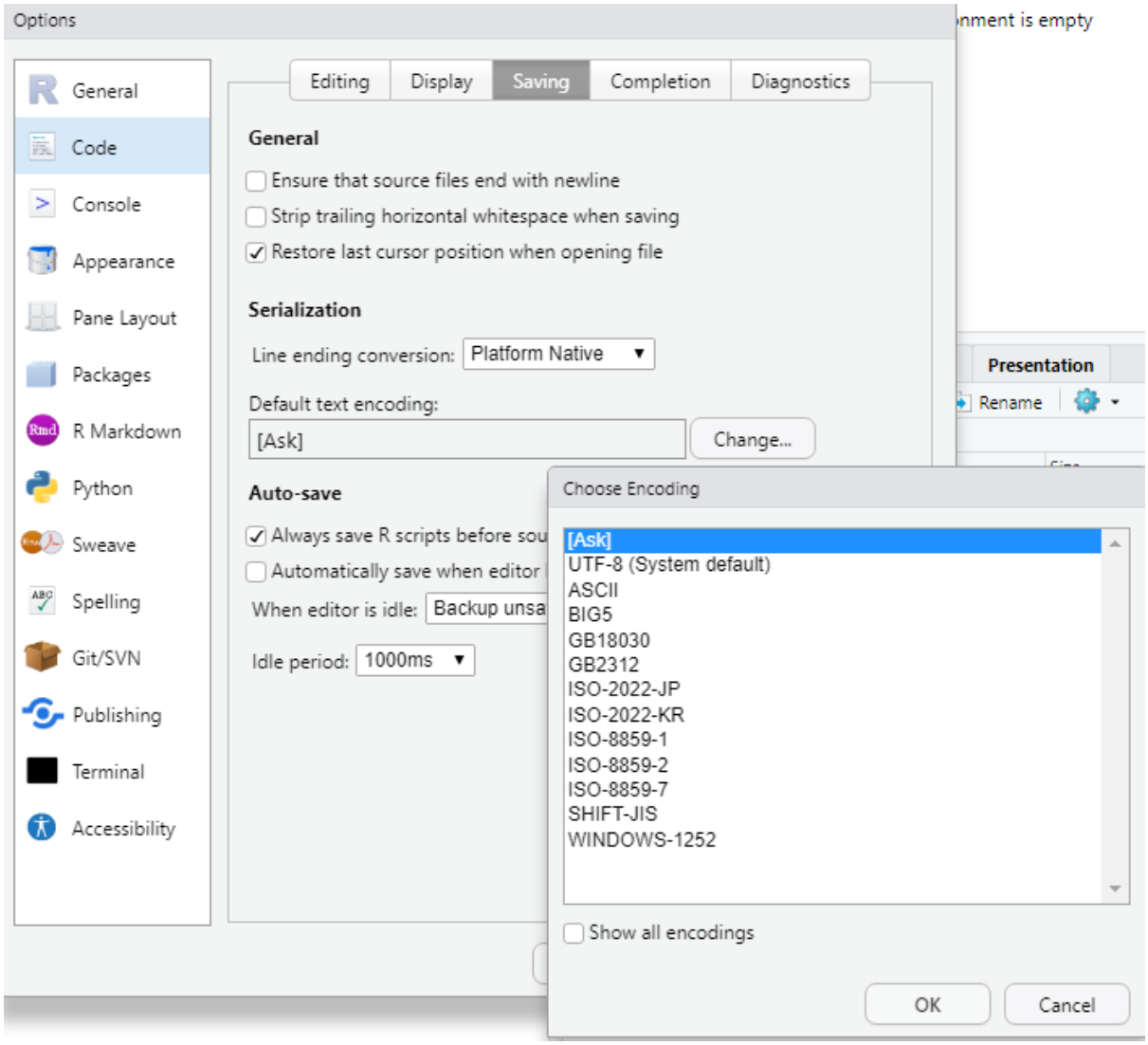
OS	Download	Size	SHA-256
Windows 10/11	RStudio-2022.07.2-576.exe	190.49 MB	b38bf925
macOS 10.15+	RStudio-2022.07.2-576.dmg	224.49 MB	35028d02
Ubuntu 18+/Debian 10+	rstudio-2022.07.2-576-amd64.deb	133.19 MB	b7d0c386
Ubuntu 22	rstudio-2022.07.2-576-amd64.deb	134.06 MB	e1c51003

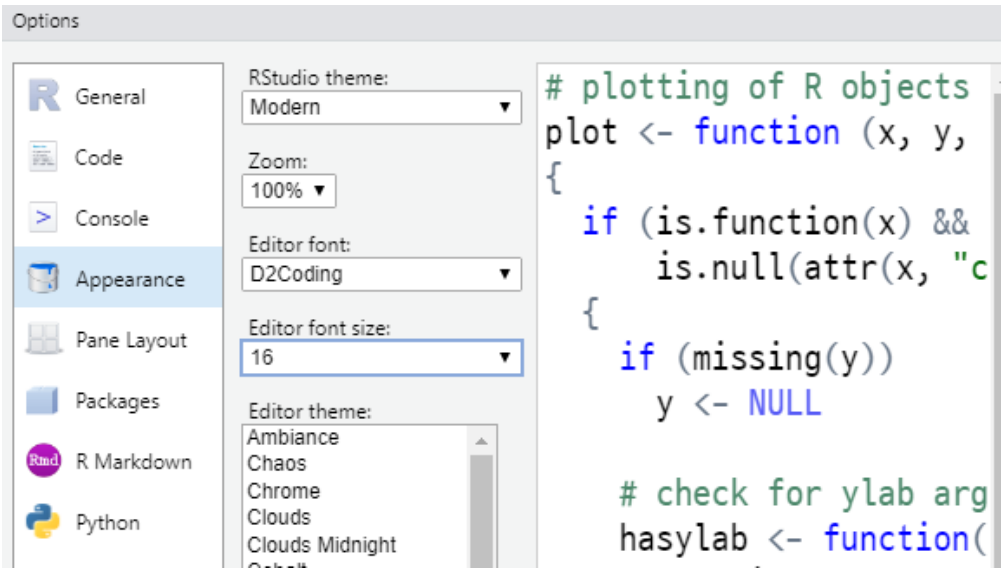
RStudio

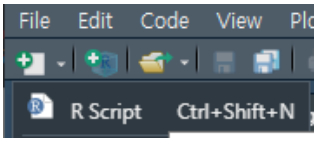

Tools
Help
Install Packages...
Check for Package Updates...
Version Control
Shell...
Terminal
Background Jobs
Addins
Memory
Keyboard Shortcuts Help Alt+Shift+K
Modify Keyboard Shortcuts...
Edit Code Snippets...
Show Command Palette Ctrl+Shift+P
Project Options...
Global Options...


Tools
Options


Soft-wrap R source files

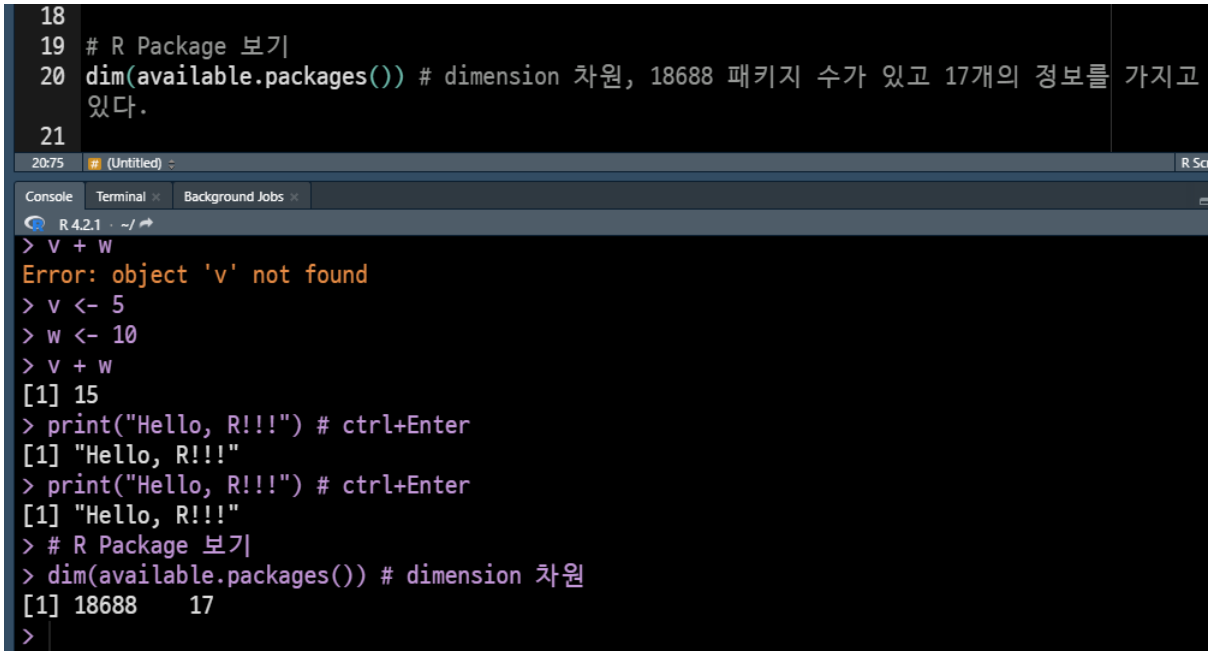

Options
General
Code
Console
Appearance
Pane Layout
Packages
R Markdown
Python
Sweave
Spelling
Git/SVN
Publishing
Terminal
Accessibility
Editing
Display
Saving
Completion
Diagnostics
General
Ensure that source files end with newline
Strip trailing horizontal whitespace when saving
Restore last cursor position when opening file
Serialization
Line ending conversion: Platform Native
Default text encoding:
[Ask]
Change...
Auto-save
Always save R scripts before sou
Automatically save when editor
When editor is idle: Backup unsa
Idle period: 1000ms
Presentation
Rename
Choose Encoding
[Ask]
UTF-8 (System default)
ASCII
BIG5
GB18030
GB2312
ISO-2022-JP
ISO-2022-KR
ISO-8859-1
ISO-8859-2
ISO-8859-7
SHIFT-JIS
WINDOWS-1252
Show all encodings
OK
Cancel


Options
General
Code
Console
Appearance
Pane Layout
Packages
R Markdown
Python
RStudio theme:
Modern
Zoom:
100%
Editor font:
D2Coding
Editor font size:
16
Editor theme:
Ambiance
Chaos
Chrome
Clouds
Clouds Midnight
# plotting of R objects
plot <- function (x, y,
{
if (is.function(x) &&
is.null(attr(x, "c
{
if (missing(y))
y <- NULL
# check for ylab arg
hasylab <- function(

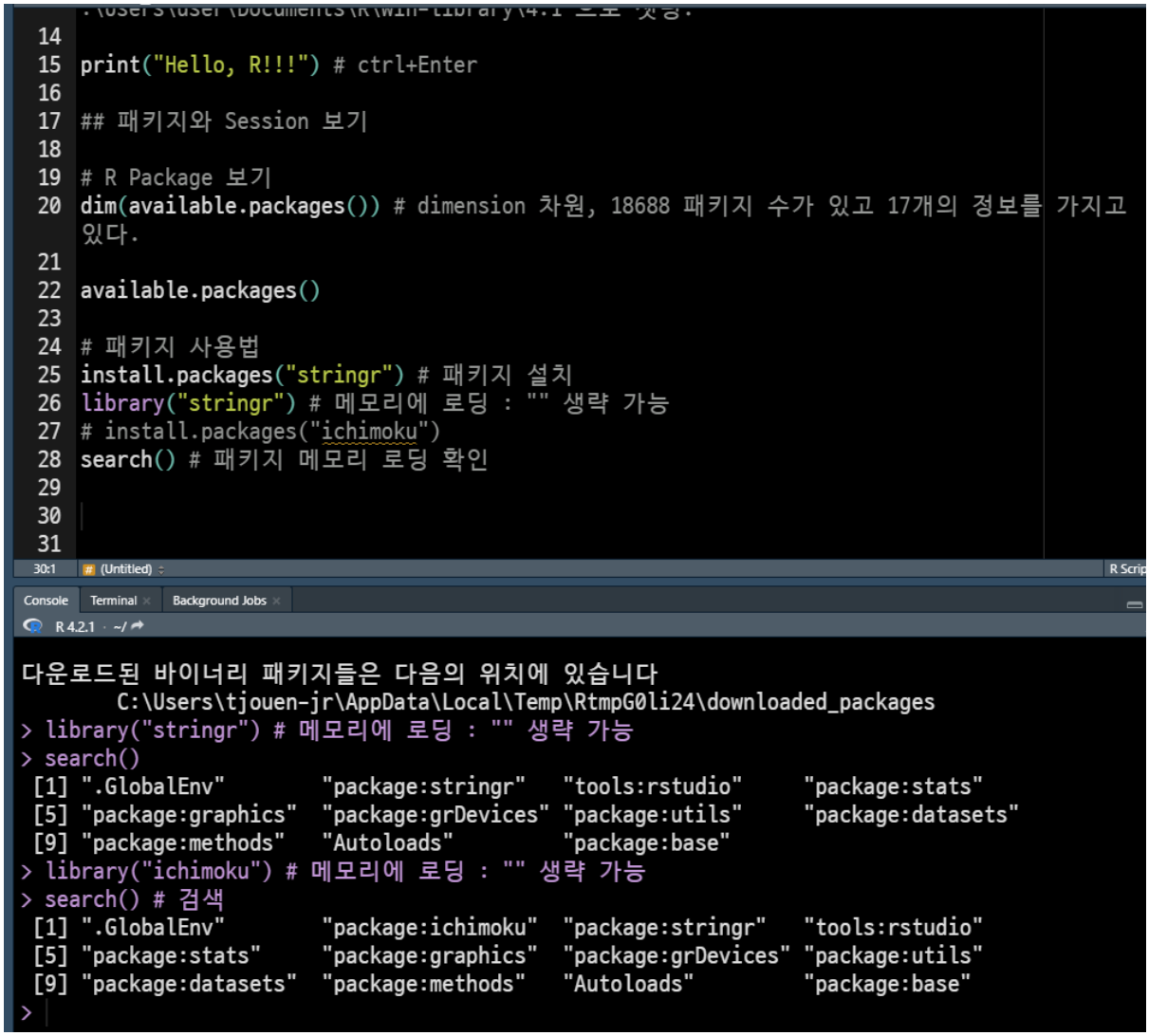


chap01_Basic.R

```
# chap01_Basic : 주석문 기호

##############################
#  Chapter01. R 설치와 개요
##############################

# 주요 단축키
# - script 실행 : ctrl+Enter, ctrl+R
# - script 저장 : ctrl+s

# Library 등록.
# - C:/Program Files/R/R-4.1.2/etc/Rprofile.site 파일에 .libPaths("C:/myRLib/Library") 추가 혹은
# - 사용자 변수 새로 만들기 : 변수 이름 -> R_LIBS, 변수 값 : C:\Users\user\Documents\R\win-library\4.1 으로 셋팅.

print("Hello, R!!!") # ctrl+Enter

## 패키지와 Session 보기

# R Package 보기
dim(available.packages()) # dimension 차원, 18688 패키지 수가 있고 17개의 정보를 가지고 있다.

available.packages()

# 패키지 사용법
install.packages("stringr") # 패키지 설치
library("stringr") # 메모리에 로딩 : "" 생략 가능
# install.packages("ichimoku")
search() # 패키지 메모리 로딩 확인

# 패키지 제거
remove.packages("stringr")
search()

# R session 보기
sessionInfo()

data()

# 기본 데이터 셋으로 히스토그램 그리기
# 단계1 : 빈도수(frequency)를 기준으로 히스토그램 그리기

hist(Nile)
```



Date	Package	Title
2022-09-26	ACEt	Estimating Dynamic Heritability and Twin Model Comparison
2022-09-26	actilifecounts	Generate Activity Counts from Raw Accelerometer Data
2022-09-26	basemaps	Accessing Spatial Basemaps in R
2022-09-26	Bergm	Bayesian Exponential Random Graph Models
2022-09-26	calculus	High Dimensional Numerical and Symbolic Calculus
2022-09-26	climatrends	Climate Variability Indices for Ecological Modelling
2022-09-26	coefa	Meta Analysis of Factor Analysis Based on CO-Occurrence Matrices
2022-09-26	colorDF	Colorful Data Frames in R Terminal
2022-09-26	comorbidity	Computing Comorbidity Scores
2022-09-26	conos	Clustering on Network of Samples
2022-09-26	dsfa	Distributional Stochastic Frontier Analysis
2022-09-26	dttr2	Manipulate Date, POSIXct and hms Vectors
2022-09-26	eatATA	Create Constraints for Small Test Assembly Problems
2022-09-26	EEMDlstm	EEMD Based LSTM Model for Time Series Forecasting
2022-09-26	extrafrail	Estimation and Additional Tools for Alternative Shared Frailty Models
2022-09-26	fso	Fuzzy Set Ordination
2022-09-26	ggOceanMaps	Plot Data on Oceanographic Maps using 'ggplot2'
2022-09-26	ggridges	Ridgeline Plots in 'ggplot2'

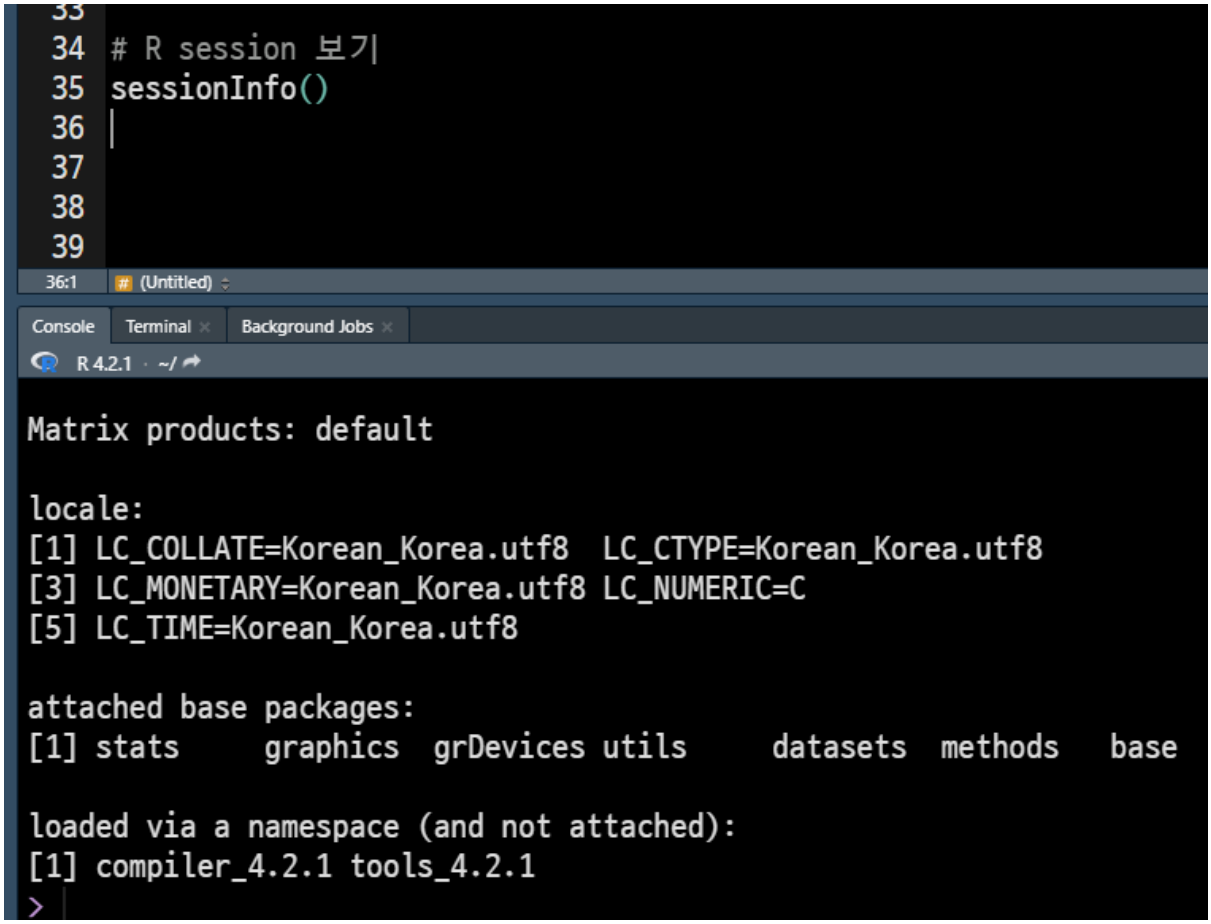


```
14
15 print("Hello, R!!!") # ctrl+Enter
16
17 ## 패키지와 Session 보기
18
19 # R Package 보기
20 dim(available.packages()) # dimension 차원, 18688 패키지 수가 있고 17개의 정보를 가지고
   있다.
21
22 available.packages()
23
24 # 패키지 사용법
25 install.packages("stringr") # 패키지 설치
26 library("stringr") # 메모리에 로딩 : "" 생략 가능
27 # install.packages("ichimoku")
28 search() # 패키지 메모리 로딩 확인
29
30
31
```

```
Console  Terminal  Background Jobs
R 4.2.1 · ~/

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
        C:\Users\tjouen-jr\AppData\Local\Temp\RtmpG0li24\downloaded_packages
> library("stringr") # 메모리에 로딩 : "" 생략 가능
> search()
 [1] ".GlobalEnv"        "package:stringr"   "tools:rstudio"     "package:stats"
 [5] "package:graphics"  "package:grDevices" "package:utils"     "package:datasets"
 [9] "package:methods"   "Autoloads"         "package:base"
> library("ichimoku") # 메모리에 로딩 : "" 생략 가능
> search() # 검색
 [1] ".GlobalEnv"        "package:ichimoku"  "package:stringr"   "tools:rstudio"
 [5] "package:stats"     "package:graphics"  "package:grDevices" "package:utils"
 [9] "package:datasets"  "package:methods"   "Autoloads"         "package:base"
>
```

위치 한국, 한국통화, 한국시간 기준

기본 패키지 알려줌.



R은 시각화의 기능을 많이 사용한다.

히스토그램

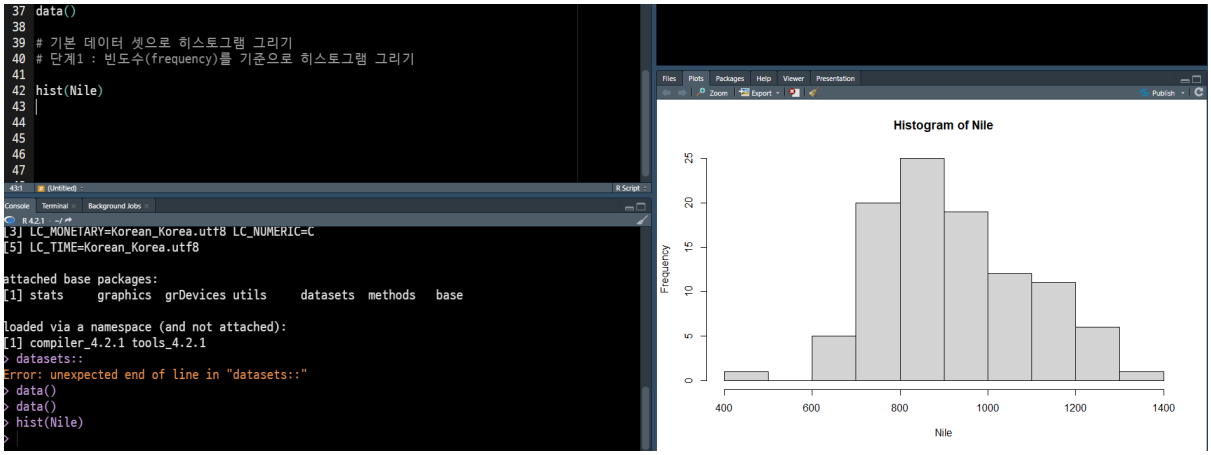
```
JohnsonJohnson          Quarterly Earnings per Johnson & Johnson Share
LakeHuron               Level of Lake Huron 1875-1972
LifeCycleSavings        Intercountry Life-Cycle Savings Data
Loblolly                Growth of Loblolly pine trees
Nile                    Flow of the River Nile
Orange                  Growth of Orange Trees
OrchardSprays           Potency of Orchard Sprays
PlantGrowth             Results from an Experiment on Plant Growth
Puromycin               Reaction Velocity of an Enzymatic Reaction
Seatbelts               Road Casualties in Great Britain 1969-84
Theoph                  Pharmacokinetics of Theophylline
Titanic                 Survival of passengers on the Titanic
ToothGrowth             The Effect of Vitamin C on Tooth Growth in Guinea
                        Pigs
UCBAdmissions           Student Admissions at UC Berkeley
UKDriverDeaths          Road Casualties in Great Britain 1969-84
```

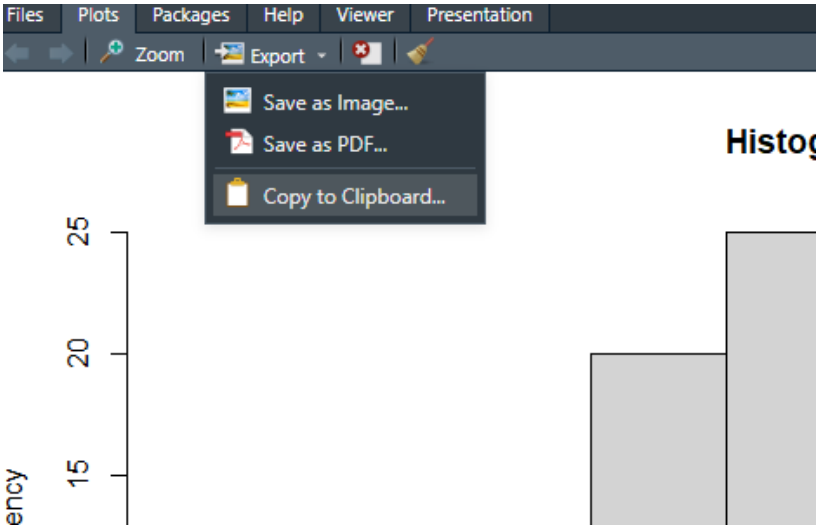
Console Terminal × Background Jobs ×

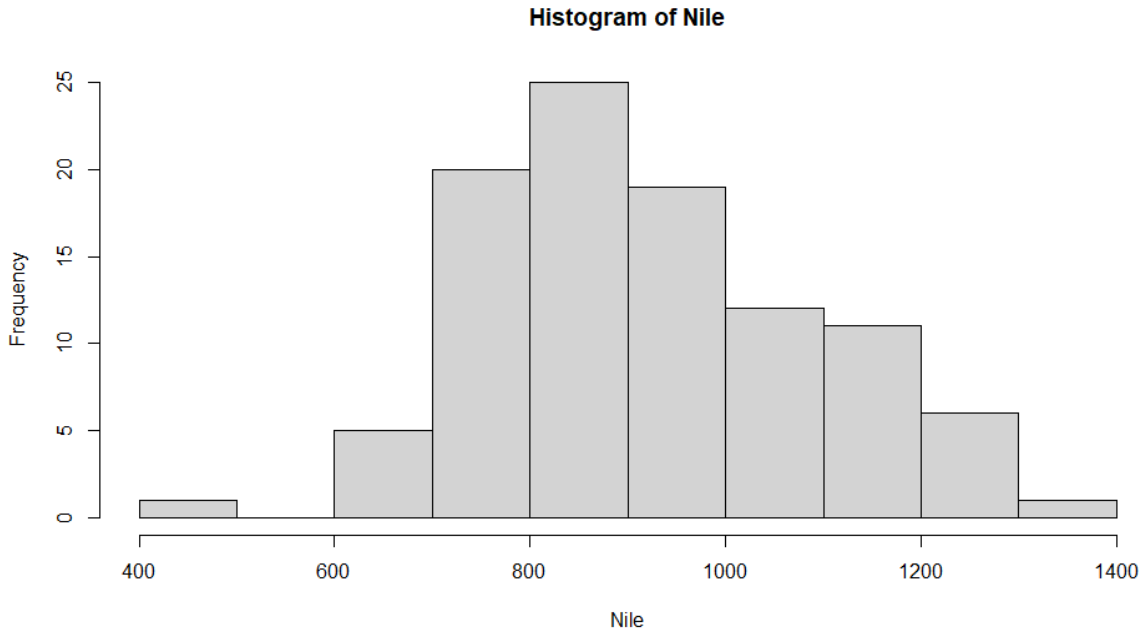
R 4.2.1 · ~/

```
[1] LC_COLLATE=Korean_Korea.utf8  LC_CTYPE=Korean_Korea.utf8
[3] LC_MONETARY=Korean_Korea.utf8 LC_NUMERIC=C
[5] LC_TIME=Korean_Korea.utf8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  method

loaded via a namespace (and not attached):
[1] compiler_4.2.1 tools_4.2.1
> datasets::
Error: unexpected end of line in "datasets::"
> data()
> data()
>
```

칸이 작아서 그림이 안 나와서 에러가 날 수도 있음. 그러면 plot 창을 넓히면 됨.




Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Save as Image...
Save as PDF...
Copy to Clipboard...
Histo
ency
15
20
25


Histogram of Nile
Frequency
0
5
10
15
20
25
400
600
800
1000
1200
1400
Nile