

M11 – Introduction

LearningSpoonsR

2019-01-25

강의 소개

설치 및 준비

논 의

수업 과정

강의 소개

강의 소개

강사 소개

- ▶ 심민규 (공학박사)
 - ▶ 현재
 - ▶ 경희대학교 연구교수
 - ▶ FinTech 스타트업 자문
 - ▶ 러닝스푼즈등 강의
 - ▶ 경력
 - ▶ 항공사, Operation 수요 분석
 - ▶ 자산운용사, 시장 분석 및 펀드 운용
 - ▶ FinTech 스타트업, 모형 개발

강의 일정 및 소통

- ▶ 1/28(월), 2/11(월), 2/18(월), 2/25(월), 3/4(월)
- ▶ 2/4(월) 휴강
- ▶ facebook에서 “Minkyu Sim” 검색 후 친구 신청 → 페이스북 그룹에 넣어드림
- ▶ 강의 자료 페이지 <https://github.com/LearningSpoonsR/LS-DS>
- ▶ learingSpoonsR@gmail.com (강사 이메일)
- ▶ 김형중 매니저

프로젝트 소개

수강생 프로젝트 소개

- ▶ <http://bitly.kr/1G0Q>

M41으로 보는 강의 목표

- ▶ 상관관계 (correlation) vs 인과관계 (causation)

1. income \rightarrow 기대수명
2. 부모수명 \rightarrow 자녀수명
3. 쌍둥이 형 \rightarrow 쌍둥이 동생

- ▶ $Y = \beta_0 + \beta_1 X + \epsilon$

1. $y = ax + b$
2. β_1
3. β_0
4. ϵ

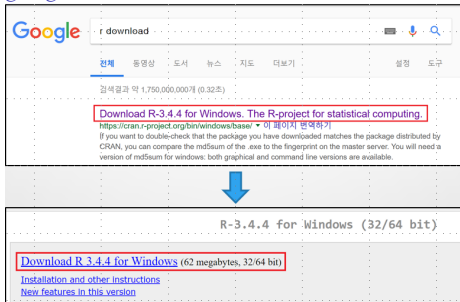
설치 및 준비

R 설치

설치 전 할일

- ▶ 윈도우 사용자 계정을 영어로 바꿔주세요.
 - ▶ 제어판 → 사용자 계정 → 계정 이름 변경 → 영어로 입력
 - ▶ 앞으로도 R파일과 데이터 파일 등도 가급적이면 영어로 다루는게 편리합니다.

google에서 “R download”를 검색하고 다운받아 설치합니다.



R Studio 설치

google에서 “R Studio download”를 검색하고 다운받아 설치합니다.

The image shows a Google search for "R STUDIO download". The search results include a link to the RStudio website: <https://www.rstudio.com/products/rstudio/download/>. Below the search results, the RStudio website is displayed. The website has a navigation bar with links: Products, Resources, Pricing, About Us. The main content area is titled "Choose Your Version of RStudio" and features a large "DOWNLOAD" button. Below the button, there is a table titled "Installers for Supported Platforms" listing various operating systems and their corresponding installer sizes and dates.

R STUDIO download

전체 동영상 이미지 도서 뉴스 더보기 설정 도구

검색결과 약 8,960,000개 (0.36초)

Download RStudio – RStudio
<https://www.rstudio.com/products/rstudio/download/> ▾ 이 페이지 번역하기

2018. 3. 15. - RStudio includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. Learn More about RStudio features.

이 페이지를 여러 번 방문

Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. Learn More about RStudio features.

RStudio Desktop Open Source License FREE \$995

RStudio

Products Resources Pricing About Us

RStudio Desktop 1.1.442 – Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it here.

Installers for Supported Platforms

Installer	Size	Date	MD5
RStudio 1.1.442 - Windows 64-bit (64-bit)	85.8 MB	2018-03-12	25a6eb8ecae4f971901c977d9cfb104b
RStudio 1.1.442 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-03-12	89613427803a1e516372075ec2e2d4b2
RStudio 1.1.442 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-03-12	090fcb1fec90e3d621bc89e113c8dc28
RStudio 1.1.442 - Ubuntu 12.04/15.10/Debian 8 (64-bit)	97.4 MB	2018-03-12	2c0005a6a8f12b06c7e6b343692288fd
RStudio 1.1.442 - Ubuntu 16.04+/Debian 9+ (64-bit)	65.1 MB	2018-03-12	c9eb172938b10626f8e4d65fa81c7175
RStudio 1.1.442 - Fedora 19+/Redhat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-03-12	77ced16b9ca8d9c636d388b842a60b1c
RStudio 1.1.442 - Fedora 19+/Redhat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-03-12	8e6435aa53fa9e9878ef9c09b641f4

R Studio를 실행합니다.

현재 변수 및 환경

최근 명령 (Command)

HELP 창

R과 CPU의 통신 기록

탐색기 기능

그래프

```

1 ggplot.Rmd
2 title: "Rich Country = Live Longer??"
3 runtime: shiny
4 output:
5   flexdashboard::flex_dashboard:
6     source: embed
7
8
9 {r setup, include=FALSE}
10 library(flexdashboard)
11 library(ggplot2)
12
13

```

Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,
or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R:

>

Name	Size	Modified
.Rhistory	23.5 KB	Mar 12, 20
data	870.5 KB	Feb 18, 20
ggplot.html	3.9 KB	Feb 18, 20
ggplot.R	2.6 KB	Apr 14, 20
ggplot.Rmd	14.6 KB	Feb 20, 20
rconnect	204 B	Feb 19, 20
script.docx		
새 텍스트 문서.txt		

1. 좌상 패널 (Editor)

- ▶ 메모장과 같은 기능을 제공합니다. (편집 및 저장)
- ▶ **Ctrl + Enter**
 - ▶ 커서가 있는 라인
 - ▶ 선택된 블록이 있을 경우 그것을 실행합니다.
- ▶ **Ctrl + 1**
 - ▶ 다른 패널에서 Editor 패널로 이동합니다.
- ▶ **Ctrl + Shift + 1**
 - ▶ Editor 패널을 전체화면으로 표시합니다.

2. 좌하 패널 (Console)

- ▶ DOS창 처럼 R의 명령어를 실행할 수 있습니다.
- ▶ **Ctrl + 2**
 - ▶ 다른 패널에서 Console 패널로 이동합니다.
- ▶ **Ctrl + Shift + 2**
 - ▶ Console 패널을 전체화면으로 표시합니다.
- ▶ **Terminal** 탭은 R와 CPU의 통신 기록을 보여줍니다.

3. 우상 패널

- ▶ 현재 메모리 상태등을 보여줍니다.

4. 우하 패널

- ▶ 파일 탐색기, 출력된 그래프, HELP등의 기능을 제공합니다.

아래 코드를 콘솔에 입력해보세요.

```
"Hello"  
a <- 10  
b <- a + 5  
install.packages("dplyr")  
library(dplyr)  
b
```

- ▶ 에러가 없이 수행되고 아래의 결과가 나오시나요?

[1] 15

- ▶ 우상단 패널에서 “History” 탭을 누르면 지금까지 실행한 명령을 볼 수 있습니다.
- ▶ 이들을 카피해서 좌상단 Editor 패널에 입력하세요.
- ▶ **File** → **Save As**를 이용해서 **first.R**로 저장하세요.
- ▶ Rstudio를 종료하고 윈도우 탐색기에서 **first.R**을 찾아 더블클릭하세요.

여러가지 설정을 할 수 있습니다.

1. Tool → Global Option →

- ▶ **Appearance**: 폰트와 색상 등을 조정
- ▶ **PANE Layout**: 모니터 크기와 작업 목적에 따라 조정
- ▶ **CODE -> Saving -> Text Encoding in -> UTF-8**: 한글이 깨지는 현상을 방지합니다.

2. 한영 전환

- ▶ 한글로 전환: `Sys.Setlocale("LC_ALL", "ko_KR.UTF-8")`
- ▶ 영어로 전환: `Sys.Setlocale("LC_ALL", "en_KR.UTF-8")`

RStudio Cheatsheet (1)

RStudio IDE : : CHEAT SHEET

Documents and Apps



Open Shiny, R Markdown, knitr, Sweave, LaTeX, JAR files and more in Source Pane

Check spelling Render output Choose output format Choose output location Insert code chunk

Jump to previous chunk Run selected lines Publish selected to server outline

Access markdown guide at Help > Markdown Quick Reference

Set knitr chunk options Run this and all previous code chunks

RStudio recognizes that this name is **app.R**, **server.R**, **ui.R**, and **global.R** belong to a shiny app

Run Choose Publish Manage app location to shinyapps.io view app or server accounts

Write Code

Navigate Open in new window Save Find and replace Compile as notebook Run selected code

Multiple cursors/column selection with **Alt + mouse drag**

Source with or without Echo Show file outline

Code diagnostics that appear in the margin. Hover over diagnostic symbols for details.

Syntax highlighting based on your file's extension

Tab completion to finish function names, file paths, arguments, and more.

Multi-language code snippets to quickly use common blocks of code.

Change file type

Working Directory Pins > to see command history

Maximize, minimize panes Drag panes

R Support

Import data with wizard History of past commands to run/copy Display RPres slideshows File > New File > R Presentation

Save Delete all workspace saved objects Search inside environment

Choose environment to display from list of parent environments Display objects as list or grid

Displays saved objects by type with short description View in data viewer View function source code

Create Upload Delete Rename File folder Set as Working Directory Click on file or directory name to open.

Turn on at Tools > Project Options > Git/SVN

Show file diff

Added Deleted Modified Renamed Untracked

Open shell to type commands

Pro Features

Share Project with Collaborator Active shared collaboration Start new R Session in current project Close R Session in project Select R Version

PROJECT SYSTEM File > New Project

RStudio saves the call history, workspace, and working directory associated with a project. It reloads each when you re-open a project.

RStudio opens plots in a dedicated Plots pane

Navigate recent plots Export Delete Delete all plots

GUI Package manager lists every installed package

Install Update Packages library for your project

Click to load package with library(). Unlick to detach package with detach()

Package version installed library

Viewer Pane displays HTML content, such as Shiny apps, Jupyter notebooks, and interactive visualizations

Stop Shiny app Publish to shinyapps.io, refresh app

Debug Mode

Open with debug(), browser(), or a breakpoint. RStudio will open the debugger mode when it encounters a breakpoint while executing code.

Click next to line number to add/remove a breakpoint.

Highlighted line shows where execution has paused

Run commands in environment where execution has paused

Examine variables in current environment

Select function in traceback to debug

Launch debugger mode from origin

Open traceback to examine the functions that the error called before the error occurred

Console > IDE Cheatsheet

Next Continue Stop

Step through code one line at a time

Step into and out of functions to run

Resume execution mode

Quit debug

Version Control with Git or SVN

Turn on at Tools > Project Options > Git/SVN

Show file diff

Stage files

Push/Pull staged files to remote View history

Open shell to type commands

Package Writing

File > New Project > New Directory > R Package

Turn project into package, Enable roxygen documentation with Tools > Project Options > R Package Tools

Roxygen guide at Help > Roxygen Quick Reference

View (data*) opens spreadsheet like view of data set

Filter rows by value < or value range

Sort by value

Search for value

Cheatsheet?

- ▶ 모든 문법과 명령어를 외우는 것은 불가능합니다.
 - ▶ Cheatsheet은 사용법이 정리되어 있는 문서입니다.
 - ▶ 경험을 바탕으로 Cheat Sheet과 웹 검색을 하면서 프로그래밍을 합니다.
 - ▶ “Know-how” 보다 “Know-where”가 중요합니다.
- ▶ English?
 - ▶ Cheat Sheet을 가득 채운 영어가 겁나시나요?
 - ▶ 웹 검색 결과와 설명서 등은 대부분 영어로 되어있습니다.
 - ▶ 그러나 컴퓨터와 프로그래밍에 관련된 용어일 뿐입니다.
 - ▶ 약 1000개만 익히면 되므로 경험과 함께 금방 익숙해 집니다.
 - ▶ 프로그래밍에 관련된 설명은 구글/네이버 번역기으로도 쉽게 이해할 수 있습니다.

논 의

R vs Excel

▶ R에서는

1. 한 번 했던 작업을 쉽게 재현할 수 있습니다.
2. 처음에는 어렵고 오래걸릴 수도 있습니다.
3. 익숙해지면 빠르고 정확하게 분석을 할 수 있습니다.
4. 데이터 구조와 특성에 대해서 논리적으로 사고하게 됩니다.

-	Excel	R
인터페이스	GUI	Console
결과 재현	반복해서 작업	정확한 재현 가능 (Reproducible)
난이도	쉬움	초반의 학습 곡선
속도, 용량	느리고 한정적	빠르고 대용량
실수	관대함	결과물의 실수 없음이 보장
장점	눈에 보여서 직관적	추상적, 논리적 사고가 요구됨

R vs 다른 언어들

▶ R Studio의 Support

1. 안정적인 개발 환경을 제공합니다.
2. 확장 프로그램들을 통합하여 관리합니다.
3. 쉽고 직관적인 메뉴얼을 제공합니다.

▶ 다른 언어들과 비교

- ▶ MATLAB: R과 가장 유사하며 수리 계산에 강점. 유료.
- ▶ SAS: 초고용량 데이터 처리. 매우 유료.
- ▶ C, C++: 배우기 어렵지만 빠르고 다른 언어의 토대.
- ▶ JAVA: C와 유사하지만 주로 웹개발 용으로 쓰임.
- ▶ Python
 - ▶ R에 비해서 난이도, 시각화, 문서화가 약점
 - ▶ 분석과 개발의 양쪽을 지원하여 인기가 높음

R은 왜 배워야 하나요?

▶ 시대의 흐름

- ▶ 데이터에 기반한 분석과 의사결정이 필수인 시대
- ▶ 반복작업은 최대한 컴퓨터에게 맡겨야 하는 시대
- ▶ 분석과 의사 결정을 하는 도구는 시대에 따라 진화
- ▶ 암산 → 주판 → 엑셀 → 프로그래밍
- ▶ 불과 20년 전에만 해도 엑셀 1급이면 전문가였습니다. 지금은 모두가 사용합니다.
- ▶ 스마트폰에서 사진을 찍고 카톡에 첨부해서 보내는 기능도 예전에는 생소했습니다.

▶ 지적인 발전

- ▶ 유소년 교육도 구구단 → 19단 → 프로그래밍으로 트렌드 변화
- ▶ 커리어 발전을 위해 다루는 데이터 구조를 파악하고 자동으로 처리해야 합니다.
- ▶ 비전공자가 처음으로 배우기에 가장 쉽습니다.

수업 과정

목표

1. R 프로그래밍을 배웁니다.
2. 데이터 분석 프로젝트의 흐름을 익힙니다.
3. 기수강생 결과물 수준의 결과물을 만들어 냅니다.
4. Take it to your workplace!

프로젝트 과정

1. 분석
 - ▶ Design (가설 설정 및 디자인)
 - ▶ Analysis (데이터 분석)
2. 공유
 - ▶ Visualization (시각화)
 - ▶ Documentation (문서화)

강의 순서

주제	Module	Note
0. 강의 소개	M11, M41	.
1. 기본 문법	M12, M13	.
2. 첫 번째 프로젝트	M21, M22	분석, 시각화, 문서화
3. 문서화 templates	M25, M26, M27	pdf, docx
4. advanced 시각화	M23, M24	.
5. advanced 문서화	M32	html 대시보드
6. 실무 프로젝트	M43, M46	.
7. 데이터 다루기	M51, M52	전처리, 시계열 데이터
8. advanced 문서화	M33	동적인 html

Survey

1. 이름, 이메일
2. 전공
3. 프로그래밍 경험
4. 직무 소개와 수강후에 직무에 어떻게 도움이 되리라 기대하는지 말씀 부탁드립니다.