

통계학 입문

2021-03-14

데이터과학융합스쿨

손 대 순 교수

이번 강의에서는 통계 분야를 포함한, 분석이 필요한 거의 모든 분야에서 활발하게 사용하고 있는 R이라는 소프트웨어를 소개하고자 합니다.

- **R**

- 통계 계산과 그래프 작성을 위한 프로그래밍 언어이자 소프트웨어 환경
- 1993년 뉴질랜드 오클랜드 대학의 로스 이하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)에 의해 공개
- R 코어팀에 의해 지속적으로 개발
- GPL하에 배포되어 비용 부담 없이 자유롭게 사용
- 통계소프트웨어 개발과 자료 분석에 널리 사용되고 있으며, 통계학자들뿐만 아니라 계량 연구를 하는 분야에서 폭넓게 사용

- **홈페이지**

- <http://www.r-project.org>

여러가지 통계분석용 프로그램들이 많이 있습니다. SAS라는 것도 있고, MiniTab, SPSS, MatLab등 다양한 소프트웨어를 통해 분석을 수행할 수 있어요. 하지만, R은 다른 프로그램에 비해 몇가지 장점을 가지고 있습니다. 물론 단점도 있겠지만요.

R이 가진 가장 중요한 장점은 1)무료로 사용할 수 있는 2)플랫폼이라는 것이지요. R은 무료 소프트웨어로 공개되어 있습니다. 학생이고, 기업이고 무료버전을 사용할 수 있지요. 조금 특별한 기능을 사용하고자 한다면 돈을 내고 쓰면 됩니다.

그러면 플랫폼이라는 말은 무엇일까요? 요즘은 기업하는 사람들도 플랫폼 사업을 많이 하고 싶어 하죠. 플랫폼이란 그 위에서 일이 벌어지도록 하고 싶다는 뜻이에요. 구글의 플레이스토어 같은 것이겠지요. 구글이 판을 만들어 놓으면, 앱을 사고 싶은 사람도 그 판에 들어와서 살펴보고 구입할 것이며, 앱을 개발하여 파는 사람도 열심히 개발하여 그 판에 올려 놓고자 하겠지요. 그

판을 제공하는 사업을 플랫폼 사업이라고 합니다.

R에서도 이런 비슷한 기능이 있어요. 새로운 분석 방법을 개발했다거나, 새로운 기능을 구현했다거나, 또 새로운 그래프를 그리는 방법에 대해 고안했다거나 하면 R에서 구동될 수 있도록 '라이브러리'라는 것을 만들어 공개합니다. 데이터를 가지고 어떤 사람이 그린 그래프의 형태로 그리고 싶다면, 그 그래프를 그리는데 필요한 라이브러리만 R에 추가하면 금방 그릴 수 있지요. R에서는 이러한 라이브러리들이 모두 무료입니다.

자, 여러분.

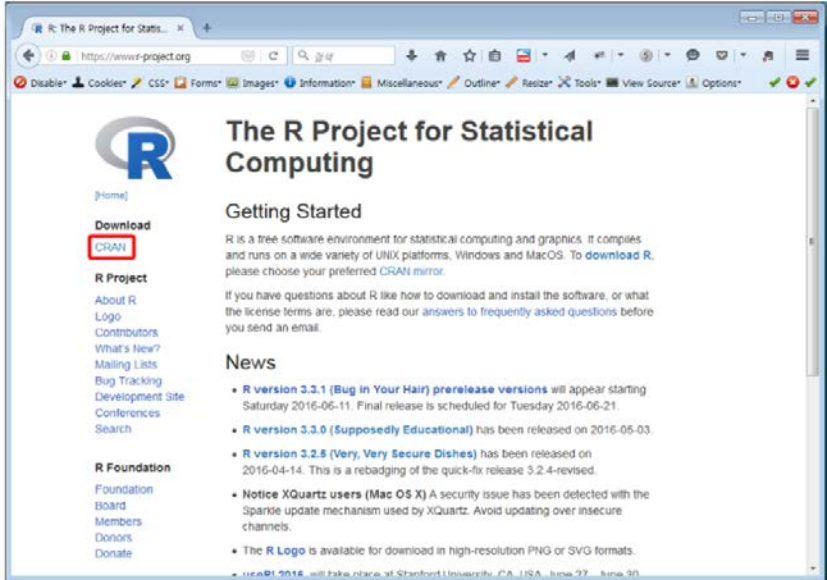
오늘은 간략하게 여러분들이 가진 컴퓨터에서 R을 설치해 보고자 합니다. 어떤 컴퓨터여도 좋습니다. 집에 있는 데스크탑형이든, 노트북이든 상관없습니다. 이 글을 보고 있는 바로 그 컴퓨터에서 해 보면 좋겠어요. 아... 핸드폰으로 보고 있나요? ㅜ.ㅜ 핸드폰은 안되겠네요.

다음의 순서대로 따라 해 봅시다.

<http://www.r-project.org>에 접속합니다.

R 설치하기

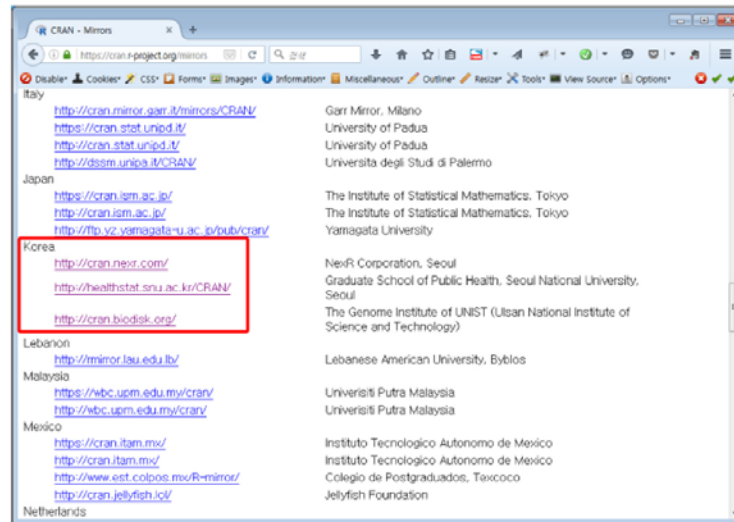
- 홈페이지에서 좌측 항목 중 'Download'의 'CRAN'을 클릭합니다



The screenshot shows the R Project for Statistical Computing website. The left sidebar has a 'Download' link highlighted. The main content area has a 'Getting Started' section with a 'CRAN' link highlighted. The 'CRAN' link is part of a list of links under the 'Getting Started' section.

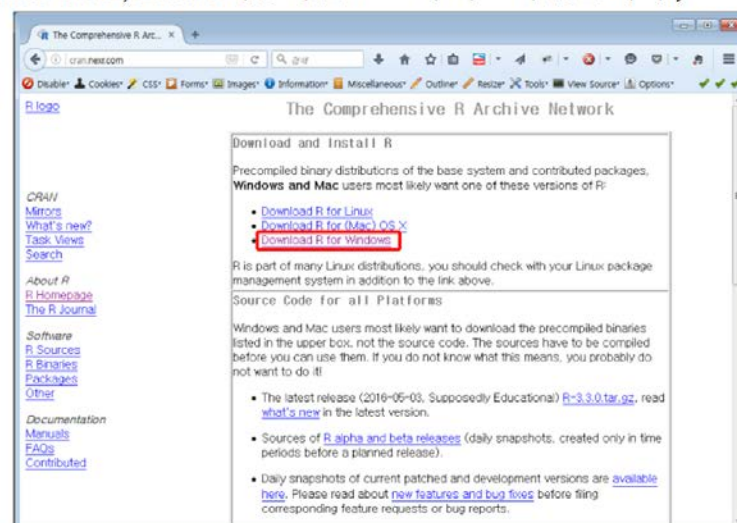
R 설치하기

- 국가별로 다운로드 받을 수 있는 곳의 링크가 나옵니다. 대한민국(Korea)의 링크 사이트 들을 찾아 세 군데 중 한 곳을 선택합니다.



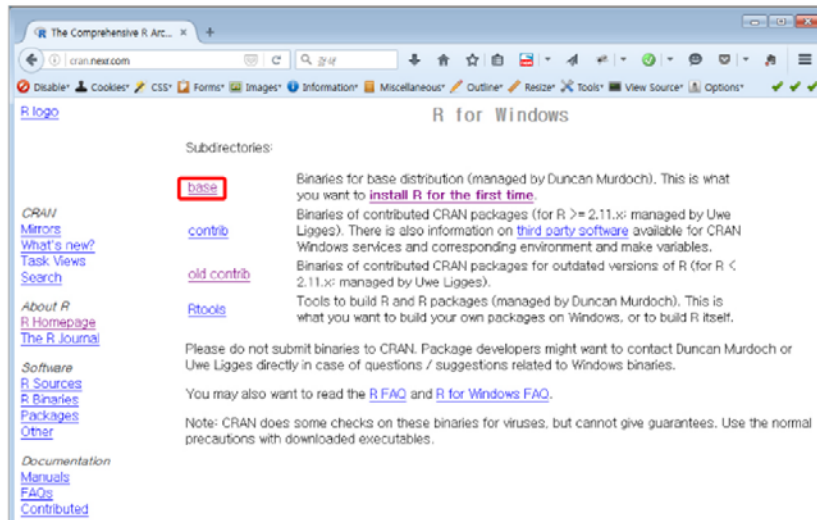
R 설치하기

- 윈도우를 기준으로 설명하므로, 'Download R for Windows'를 클릭합니다. (Max OS X, Linux 사용자들도 크게 다르지 않습니다.)



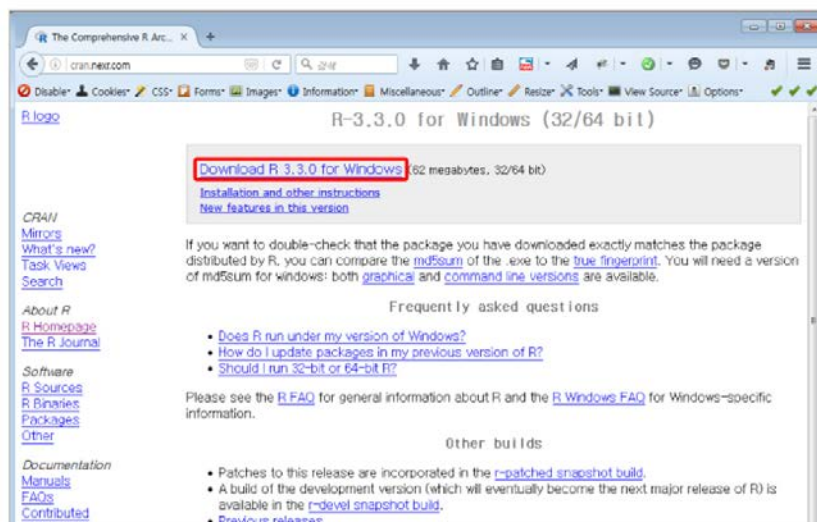
R 설치하기

- R을 설치하기 위해 여기서 'base'를 선택합니다. 나머지 링크들은 패키지 제작 및 R 환경을 위한 링크입니다. ('install R for the first time'을 클릭해도 됩니다.)



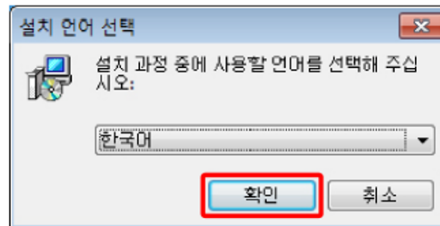
R 설치하기

- 'Download R.x.x.x for Windows'를 클릭하여 설치 파일을 다운로드합니다.



R 설치하기

- 설치과정
 - 다운로드 받은 파일을 실행하여 윈도우 설치과정을 진행합니다.

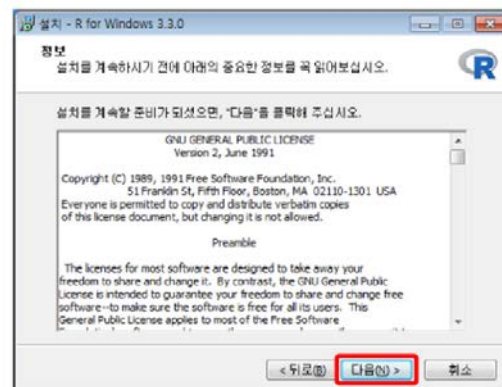


[그림 A-6] 설치 시 사용할 언어 선택

R 설치하기

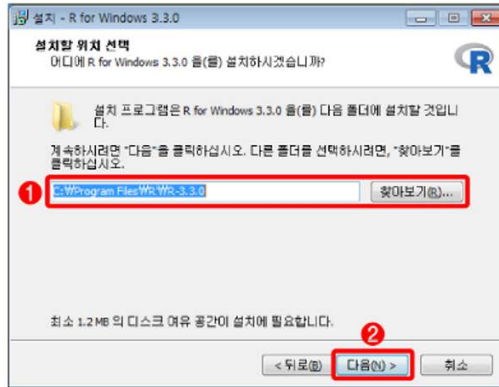


[그림 A-7] 설치 시작

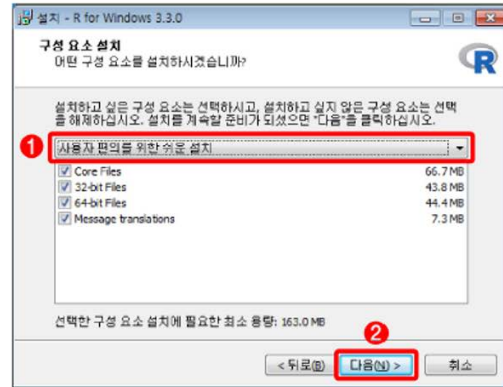


[그림 A-8] 라이선스 확인

R 설치하기



[그림 A-9] 설치 위치 확인

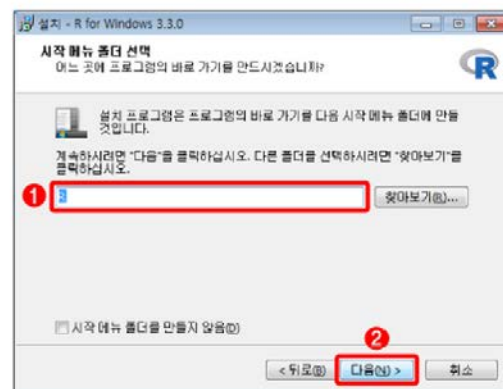


[그림 A-10] 설치 내용 확인

R 설치하기

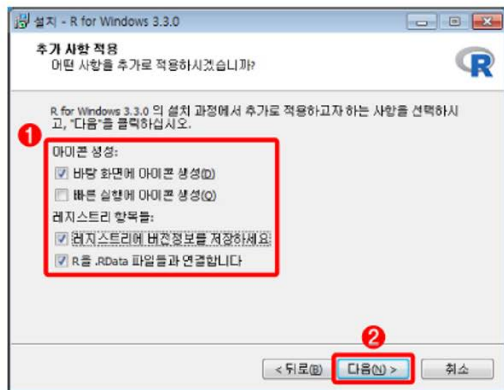


[그림 A-11] 기본 시작 옵션으로 설치

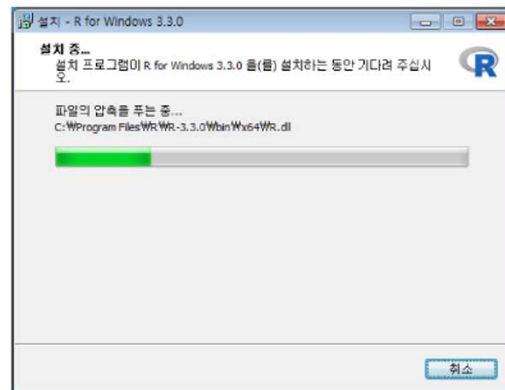


[그림 A-12] Windows의 프로그램 그룹 등록

R 설치하기

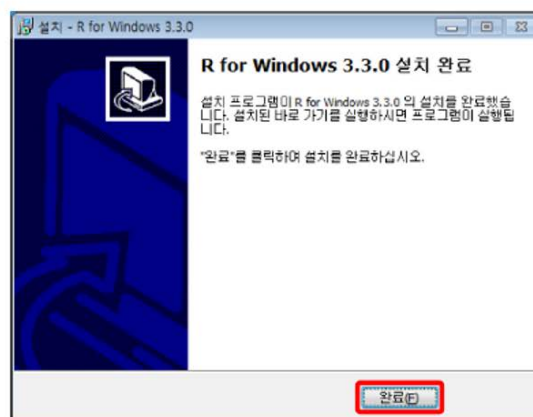


[그림 A-13] 추가 적용사항 확인



[그림 A-14] 설치 진행

R 설치하기



[그림 A-15] 설치 완료

자, 설치가 완료 되었나요?

예전 버전이라 화면이 조금 달라졌을 수는 있는데 설치에 크게 어려운 부분은 없을 것이라 생각합니다. 직접 설치 해 보는 것이 중요하니 꼭 해 보세요.

자, 그럼 이제 R studio라는 것을 설치 해 보겠습니다.

R Studio 설치하기

- **R Studio**

- R Studio는 R을 위한 통합개발환경(IDE)입니다.
 - 통합개발환경(IDE, Integrated Development Environment)
 - 코딩, 디버그, 컴파일, 배포 등 프로그램 개발에 관련된 모든 작업을 하나의 프로그램 안에서 처리하는 환경을 제공하는 소프트웨어이다(위키피디아, 통합개발환경)
 - R 통합개발환경을 위해 R Studio는 콘솔, 직접 코드를 실행시킬 수 있는 구문강조 (Syntax Highlighting) 기능이 있는 편집기와 그림을 그리고 코드 이력을 기록하고 코드 내 버그 찾기 기능과 작업공간 관리 기능을 수행합니다.
 - 오픈소스로 개발되어 있어 무료로 사용할 수 있으며, 일정 비용을 지불하고 각종 지원을 받을 수 있는 유료버전이 있습니다.
 - 홈페이지 : <https://www.RStudio.com>

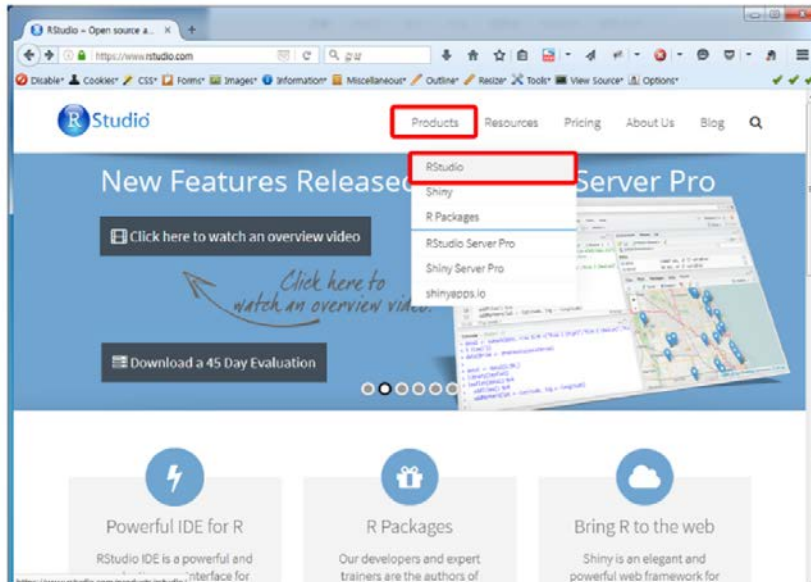
R을 그대로 사용해도 좋은데, R studio를 쓰면 사용자에게 좀 편한 여러가지 기능들을 제공합니다. R studio는 R을 컨트롤 하는 프로그램이에요. 사용해도 되고 안해도 되는 프로그램이긴 하지만, 사용하는 편이 초보자가 배우기에는 훨씬 수월합니다.

자, <https://www.RStudio.com>으로 들어가 볼까요?

R Studio 설치하기

- 설치 파일을 찾아봅시다

- 먼저 홈페이지를 방문하여 상단 메뉴의 Products의 R Studio를 선택합니다.



R Studio 설치하기

- 개인 사용자용의, 'Desktop'과 서버 설치용 'Server' 두 가지가 있습니다.
- 'R Studio Desktop(Free License)'을 선택합니다.

Choose your version of RStudio				
	RStudio Desktop (Free License)	RStudio Desktop (Commercial License)	RStudio Server (Free License)	RStudio Server Pro (Commercial License)
Integrated Development Environment for R	✓	✓	✓	✓
Priority support		✓		✓
Access via Web Browser			✓	✓
Enterprise Security and Access Controls				✓
Project Sharing				✓
Access to Multiple Versions of R				✓
Multiple Concurrent Sessions				✓
Administrative Dashboard				✓
Load Balancing and Resource Management				✓
License	AGPL	Commercial	AGPL	Commercial
	DOWNLOAD	DOWNLOAD	DOWNLOAD	DOWNLOAD

R Studio 설치하기

- 윈도우 하에서 설치를 진행합니다. (다른 운영체제도 크게 다를 바 없습니다.)
- Installers 목록 중 'R Studio-x.yy.zzz-Windows Vista/7/8/10'을 클릭합니다.
 - 다운로드가 진행됩니다.

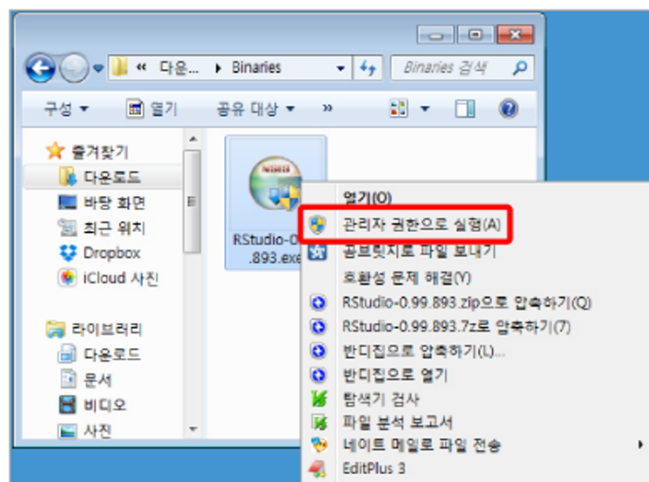
RStudio requires R 2.11.1+. If you don't already have R, download it [here](#).

Installers for Supported Platforms

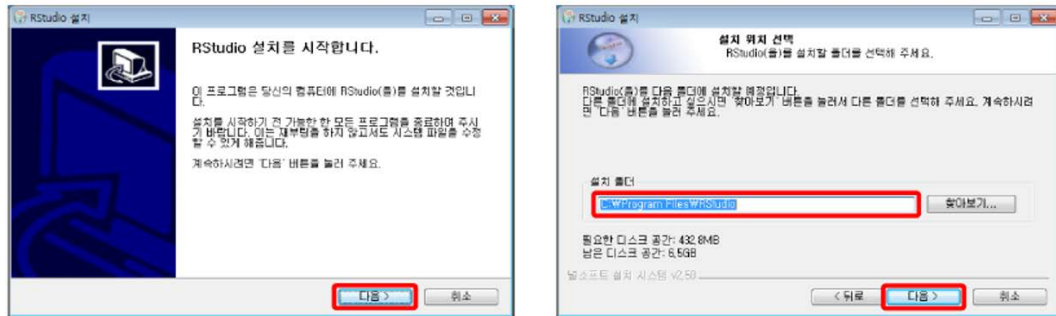
Installers	Size	Date	MD5
RStudio 0.99.903 - Windows Vista/7/8/10	77.1 MB	2016-07-18	716f28f2143c5e21f4acea5752e284f8
RStudio 0.99.903 - Mac OS X 10.6+ (64-bit)	60 MB	2016-07-18	d14a1585b5a5ac0839507b9c04d460d6
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (32-bit)	81.6 MB	2016-07-18	761eae80b0ba4d4cd9051a802a2c44e2
RStudio 0.99.903 - Ubuntu 12.04+/Debian 8+ (64-bit)	88.3 MB	2016-07-18	98ea59d3db00e0083d3e4053514f764d
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	81 MB	2016-07-18	ce2ea1023d99175cb909def0fe66eba7
RStudio 0.99.903 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	81.9 MB	2016-07-18	152f247255e86904cf3354afbc7b3b99

R Studio 설치하기

- 설치과정
 - 다운로드 받은 파일을 실행하여 윈도우 설치과정을 진행합니다.
 - 마우스 오른쪽 버튼을 클릭하여 관리자 권한으로 실행합니다.



R Studio 설치하기



R Studio 설치하기



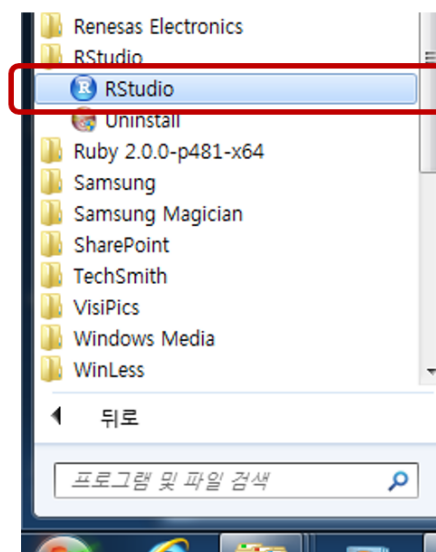
R Studio 설치하기



R Studio 실행과 기본설정 : 작업환경 구축

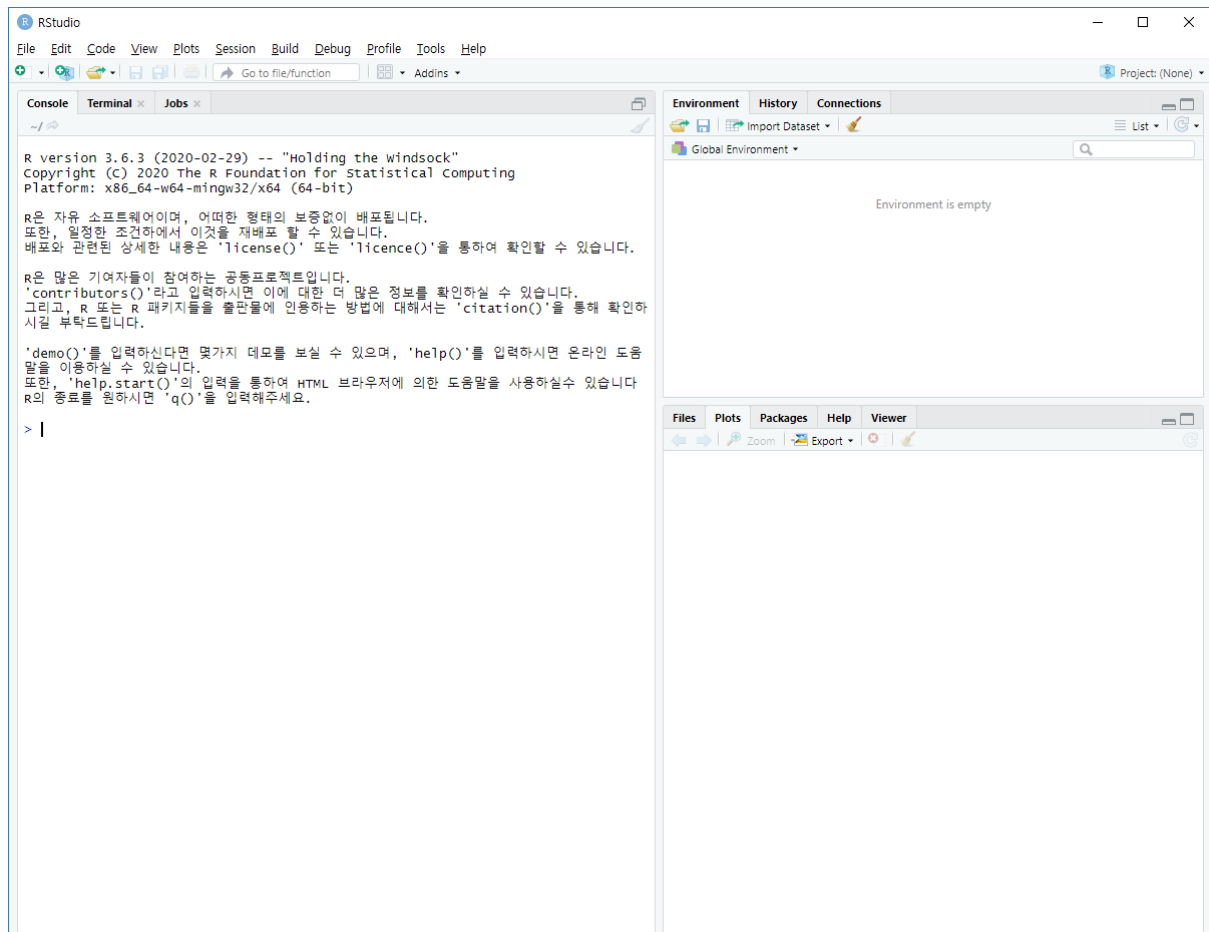
• R Studio 실행

- 프로그램 그룹에서 “R Studio” 항목을 찾습니다.
- 항목안에 “R Studio”를 클릭하여 실행합니다.



이제 R과 R Studio까지 모두 설치가 끝났습니다.

R studio를 실행해 보세요. 아래와 같은 화면이 나오면 성공입니다. !!



R은 우선 계산기 역할을 할 수 있어요.

일반적인 계산기보다 더 똑똑한 계산도 가능합니다.

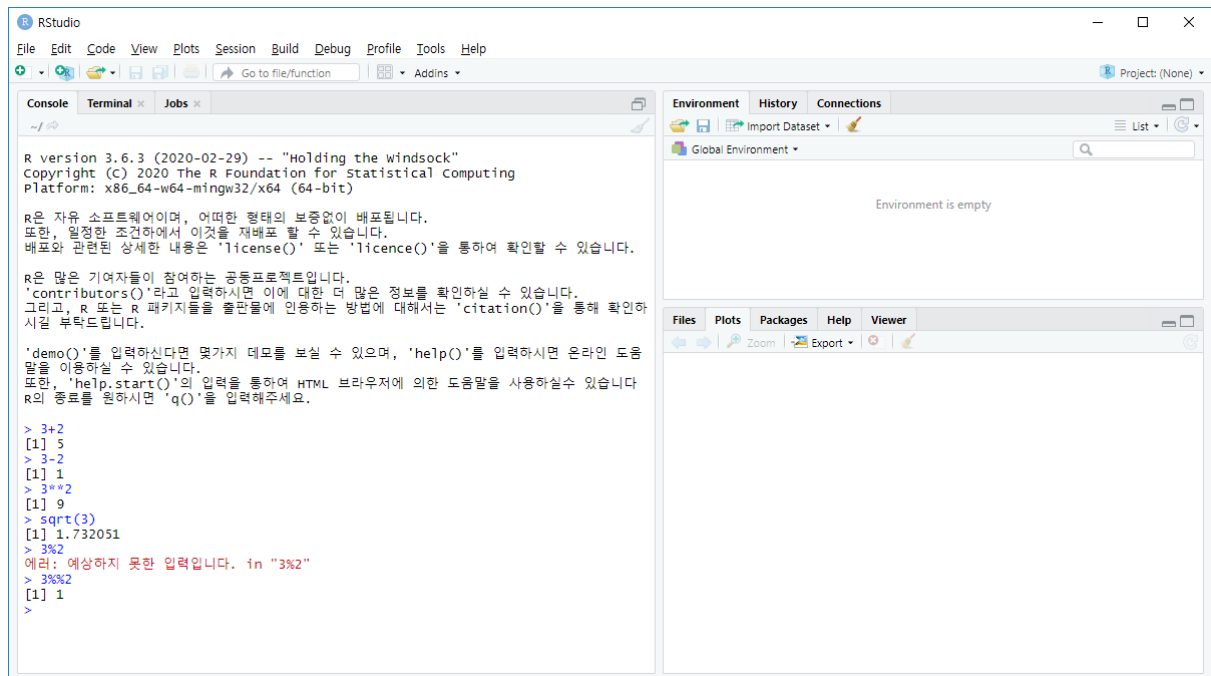
[표 1-3] R의 산술연산자

연산자	설명	예	결과
+	더하기	3 + 2	5
-	빼기	3 - 2	1
*	곱하기	3 * 2	6
/	나누기	3 / 2	1.5
^ 혹은 **	승수	3 ^ 2	9
x %% y	X를 y로 나눈 나머지 값 반환	3 %% 2	1
x %/% y	나누기의 결과를 정수로	3 %/% 2	1

위의 표에서 보는 것처럼, 제곱이나 나머지 계산도 훌륭하게 잘 해 냅니다.

R studio화면 왼쪽에 'Console'창이 보이죠? 여기에 명령어를 쳐 넣고 엔터를 치면 됩니다.

한 번 해 보세요.



몇가지 명령어를 쳐 봤는데, 잘 되지요?

화면에서 사용한 sqrt는 루트입니다. Square root를 sqrt라는 함수로 쓰는 것이지요.

따라서, sqrt(3)하면 3의 제곱근을 구하라는 명령이 됩니다.

한림대학교 학생 중 1학년이 500명, 2학년이 400명, 3학년이 250명, 4학년이 95명 있다고 가정 합시다. 학교 전체 학생들의 학년 비율을 그래프로 한 번 그려볼까요?

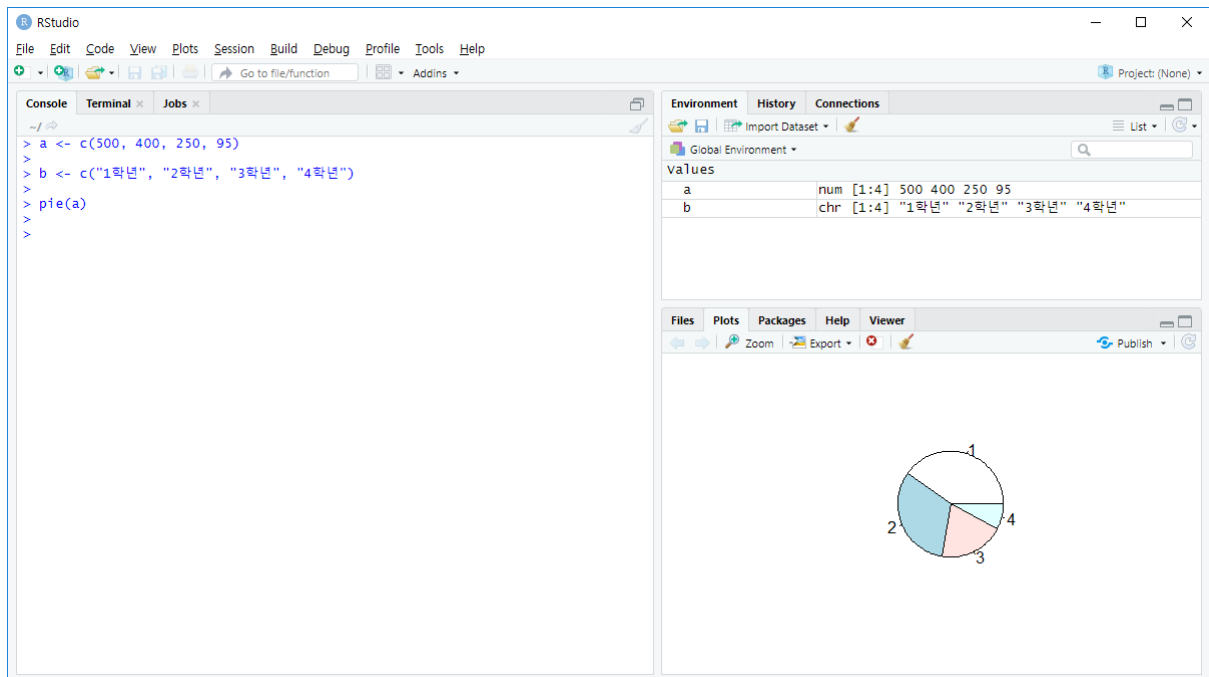
다음과 같이 입력해 봅시다.

```
> a <- c(500, 400, 250, 95)
> b <- c("1학년", "2학년", "3학년", "4학년")
> pie(a, labels=b)
```

한 줄씩 콘솔 창에 입력하면 됩니다.

a라는 변수에는 학생 수를 넣어 주고, b라는 변수에는 각각이 무엇을 뜻하는지에 대한 표식을 넣어 주었습니다. 그리고는 파이차트라고 하는 것을 그려 보죠.

명령어는 간단합니다. 'pie(a)'라고 입력하면 됩니다. a라는 변수를 가지고 파이차트를 그리라는 명령어예요.



오른쪽 아래에 파이차트가 그려진 것이 보이나요?

그냥 1,2,3,4로 구별되어 있으니 보기가 불편하죠? 레이블(표식)을 넣어 봅시다.

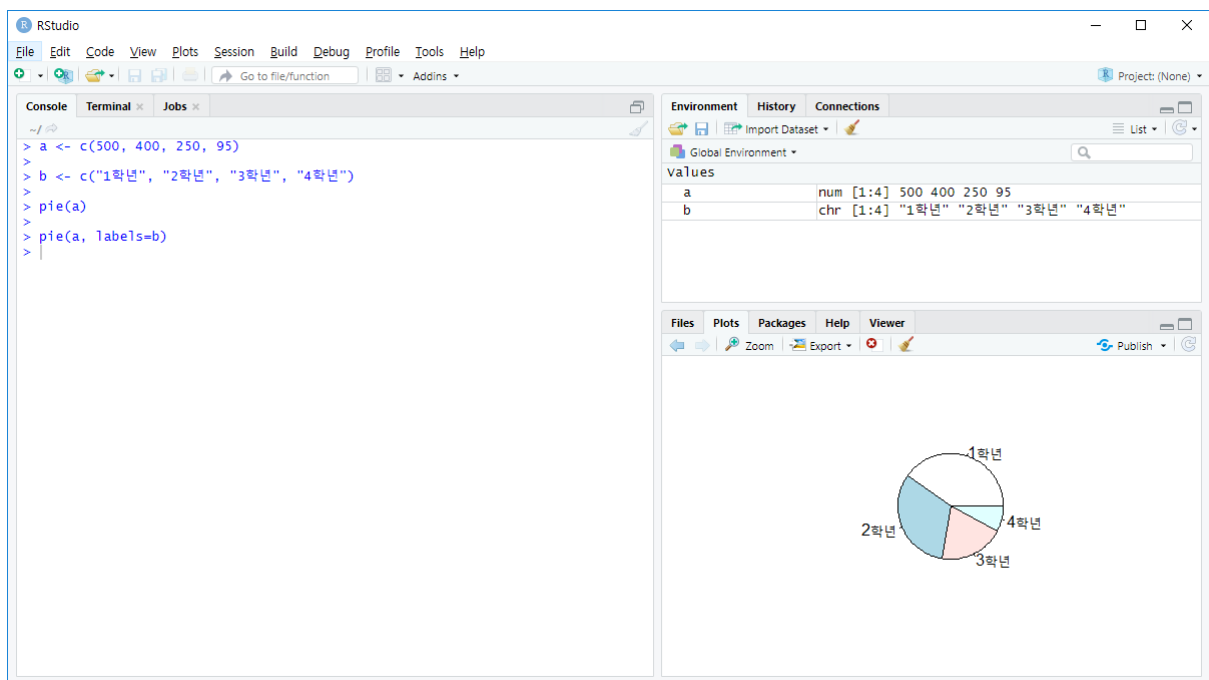
레이블도 변수로 만들어 놓았으니, 그 변수가 무엇인지만 pie함수에 알려주면 됩니다.

pie(a, labels=b)라고 입력해 볼까요?

어때요? 그림에 학년 표시가 제대로 되어 나오나요?

R은 이렇게 그래프도 그려 줍니다.

아직은 보기에 썩 예쁜 그림은 아니지만, 색깔도 바꿀 수 있고, 제목을 넣을 수도 있고, 3차원으로 그릴 수도 있어요. 얼마든지 예쁘게 꾸밀 수 있습니다.



자, 여기까지... 여러분들 해 보고 있는거죠?

과제의 최소한의 요구사항은 파이차트의 인증샷입니다.

본인이 직접 해 보고, 인증샷을 스마트캠퍼스 '과제제출'에 올려 주세요. 검색 해 보고, 조금씩 응용해 보면 좋겠습니다. 색깔도 이쁘게 바꾸어 보고, 제목도 붙일 수 있겠네요. 특별한 다른 기능이 있다면 시도해 보세요. 프로그램은 스스로 찾아보면서 학습하는 것이 가장 빠른 방법입니다.

어떻게 하면 본인이 직접 해 보았다는 인증이 될까요? '확실하게 내가 한거다!!'라는 것이 증명되는 단 하나의 인증샷. 여러분들의 창의적 아이디어를 기대합니다. ^^