

데이터과학을 위한 **R**프로그래밍

1주차. 데이터과학과 오픈소스



이혜선 교수

포항공과대학교 산업경영공학과



목차

1주차. 데이터과학과 오픈소스

1차시

오픈소스와 R프로그램

2차시

R프로그램 실행(RStudio)

3차시

기본스크립트와 R 추가패키지



1주차

1차시

오픈소스와 R프로그램

데이터사이언스

데이터 (Data)



데이터 큐레이션(Data curation)
데이터추출, 변환
- SQL, R, Python

데이터분석
(Data Analytics)



데이터 시각화
(Data visualization)
그래픽 (R의 ggplot)

인사이트 창출
(Insight)



통계모형,
인공지능
(t-검정, 회귀분석,
머신러닝)

데이터애널리틱스

수집
(Data
gathering)

정제
(Data
processing)

시각화
(Data
visualization)

예측모형/분석
(Data
analysis)

● R 프로그램 소개

✓ 오픈소스 통계분석 프로그램

→ 어디서나, 누구나 사용할 수 있는 프로그램

✓ 빅데이터 분석 도구 - 다양한 통계기법과 시각화 도구 제공

통계기법, 데이터마이닝, 기계학습,
텍스트마이닝, 인공지능 분석툴

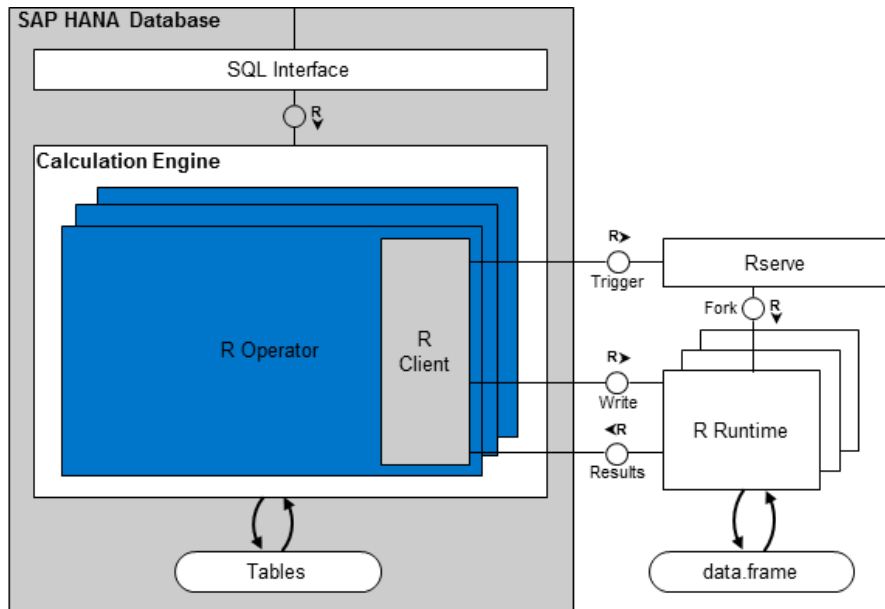
→ ggplot을 이용한 시각적분석, 공간분석맵

✓ 지금까지 17,000여개 이상의 패키지 업로드 제공

✓ R은 C++, 자바, 파이썬 등 다른 프로그램과 쉽게 연동

● R 프로그램 소개

- ✓ R의 개발 : R프로그램은 1995년에 S언어를 개선하여 **Ross Ihaka and Robert Gentleman**가 개발
- ✓ 전세계에서 (학계, 연구기관) 범용적으로 사용
- ✓ 기업체에서도 R을 탑재엔진으로 한 기업형분석 플랫폼개발

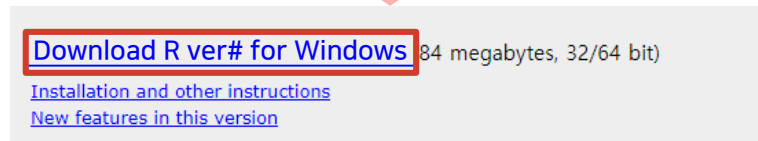
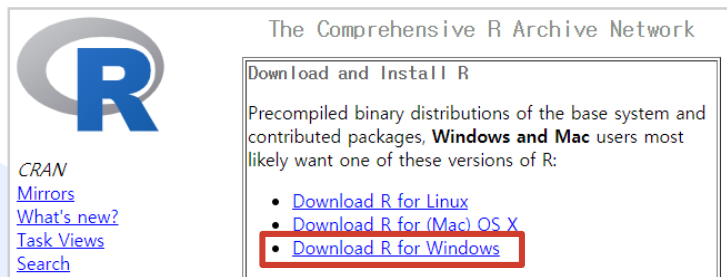


R 프로그램 설치

Step1

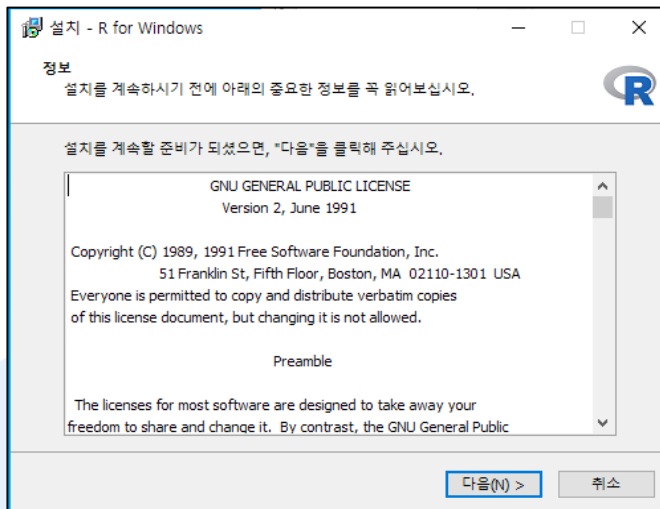
✓ R 프로그램 : Base 프로그램의 설치

🔍 <https://cran.r-project.org/>

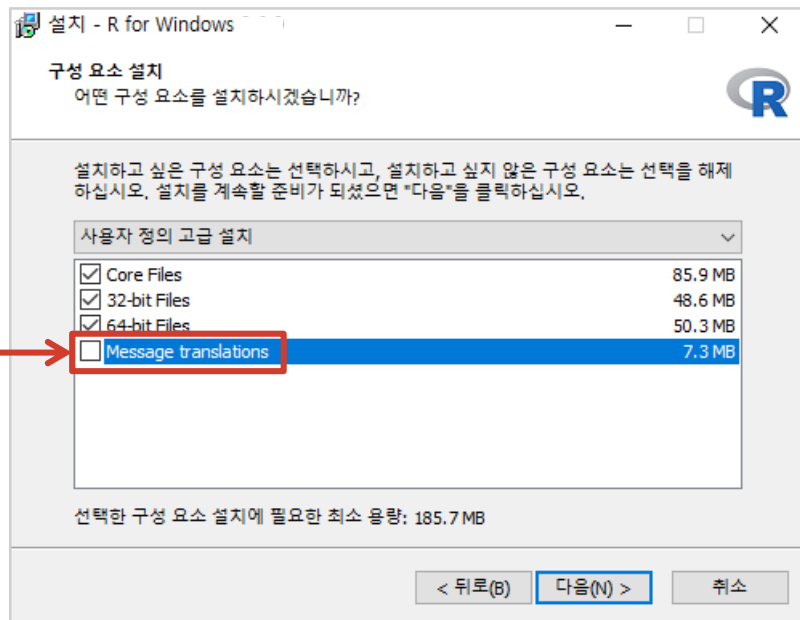


✧ R program의 버전은 계속 update됨

R 프로그램 설치



여기서 ☒를 제거하면 영문메뉴로
프로그램이 설치됩니다.
(파일메뉴 및 로그 메시지를 영문으로
사용할 것을 권장!!)

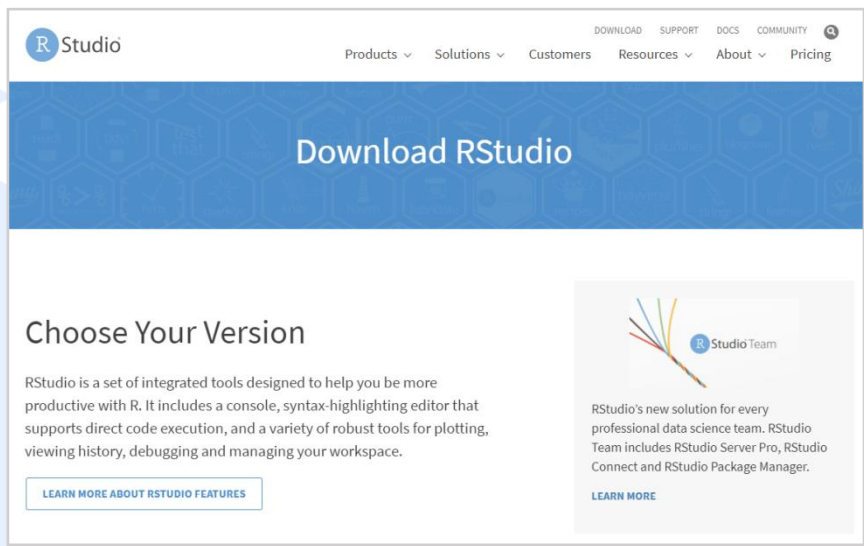


● RStudio 설치

Step2

- ☑ RStudio : R 프로그램 실행을 지원하는 통합툴 프로그램
(프로그램 입력, 작업정보관리, 그래픽, 패키지 윈도우)

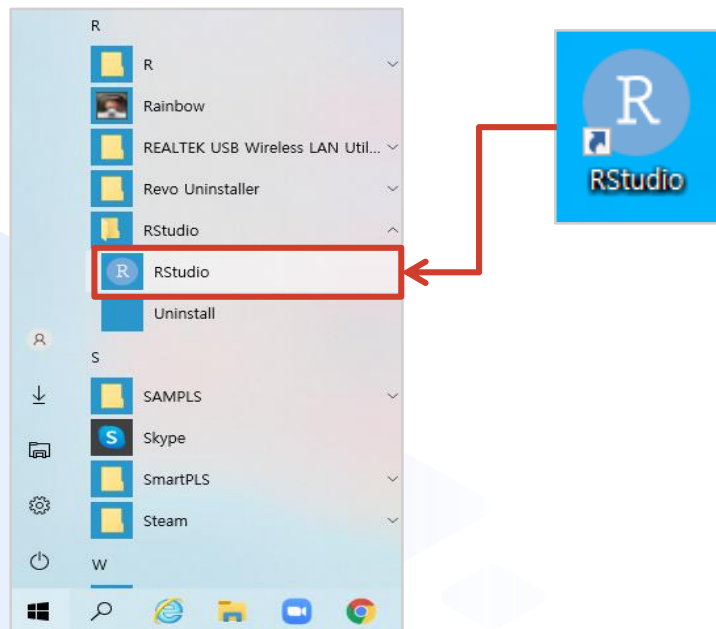
🔍 <https://www.RStudio.com/products/RStudio/download/>



● RStudio 설치

☑ RStudio 설치 후 실행아이콘

➤ RStudio 설치 후 윈도우 메뉴에서 RStudio의 실행아이콘을 바탕화면에 생성하면 편리함!



RStudio 화면구성

The screenshot shows the RStudio interface with the following components highlighted:

- 1. R 프로그램 입력창**: The source editor window showing R code for creating a vector `x1` and checking its class.
- 2. 콘솔 (프로그램 수행)**: The console window showing the R version (3.4.3) and startup messages.
- 3. R데이터/환경**: The Environment pane showing the variable `x1` with values `1 3 5 7 9`.
- 4. 패키지/그래픽/도움말**: The bottom toolbar containing icons for Files, Plots, Packages, Help, and Viewer.