R 프로그래밍 소개

(R programming for Beginners)

이상민

정보융합학부, 소프트웨어융합대학

광운대학교

Agenda

- R 프로그래밍 소개
- R 설치, R studio 설치
- R studio 패널 소개
- R을 이용한 기초통계의 이해
- R을 이용한 데이터 전처리 (Data munging)
- 난수 생성 함수 (Random data)
- 탐색적 데이터 분석 (Explanatory data analysis)

R 소개

- R 프로그래밍

  - R is a free software environment for statistical computing and graphics

  - R is an object-oriented programming language where we create objects and manipulate them as intended. Objects can be Data frames, vectors, matrices, lists, raw data, spatial objects , maps etc.

  - R is a representative tool for statisticians and data scientists for developing statistical software and data analysis

R 역사

- R의 시작?

  - A statistical programming language developed by Bell Labs in 1970s

    . S-plus 제품, John chamber, 1976

  - R was invented by Robert Gentleman and Ross Ihaka at the University Auckland in the mid-1990s

<table>
<tr><td>1993. 뉴질랜드 교수님</td><td>1997. GNU</td><td>2001. Version 1</td></tr>
</table>

로버트 젠틀맨   로스 이하카

R 용도 확장

기존　　　　　　　　　　　　　현재

통계분석용 엔진

Computer science에 기반한 data analytics package로 발달

- Programming language로도 활용
- 개발환경(IDE)으로 인식
- 통계 주도 하의 frame에서 벗어나, CS 주도 하의 frame 진입
- 기존 SPSS, SAS, Matlab 지배적인 시장에서 google 등의 지원으로 R 활용성 증대, 이에 메모리 단가 인하에 따라 메모리 기반 분석 방식의 R 실용성 증대

R vs. Excel?

- R vs. Excel
  - R can handle a huge size of data sets
  - R can import data sets which are used in excel
  - R provide more advanced statistical analysis and data visualization
  - R libraries are keep updating

R vs. Python?

- R vs. Python

    - Python is a general-purpose interpreted, interactive, object-oriented and high-level programming language

    - Python provides a bunch of libraries for software engineering and web programming. Further, Python presents the state-of-the-art algorithm in machine learning!

    - R has a more comprehensive data modeling and visualization base

    - The development using both languages is quite similar!
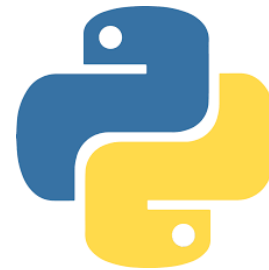
R 유용한 것들

- R-cran
    - The Comprehensive R Archive Network
    - R 라이브리리를 확인, 관리, 확장하는 패키지 관리도구

- https://www.r-bloggers.com/

- https://blog.revolutionanalytics.com/
    - 라이브러리 기본 사용 설명서, 예제 소개
    - 최근 재미있는 라이브러리 소개
    - Revolution R 은 algorithm 기본 처리가 multi-thread 지원

- Github for R 사이트

R 설치

- R 배포 사이트 방문

  - https://www.r-project.org/

  - Download → CRAN 선택

R 설치

- 아래 Korea를 찾고, 해당 mirroring server로부터 R 다운 받을 것

Korea

https://ftp.harukasan.org/CRAN/	Information and Database Systems Laboratory, Pukyong National University
https://cran.yu.ac.kr/	Yeungnam University
http://healthstat.snu.ac.kr/CRAN/	Graduate School of Public Health, Seoul National University, Seoul
https://cran.biodisk.org/	The Genome Institute of UNIST (Ulsan National Institute of Science and Technology)

- 자신의 PC 운영체제에 맞춰서 R 다운로드

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

R 설치

- base 선택

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

- 다운로드 시작

R-4.0.2 for Windows (32/64 bit)

Download R 4.0.2 for Windows (84 megabytes, 32/64 bit)

Installation and other instructions
New features in this version

R 설치

- 설치 진행 (이전 이후에는 가능한 옵션 변경 없이 '다음')
  - 경로상 한글이 없도록 수정할 것
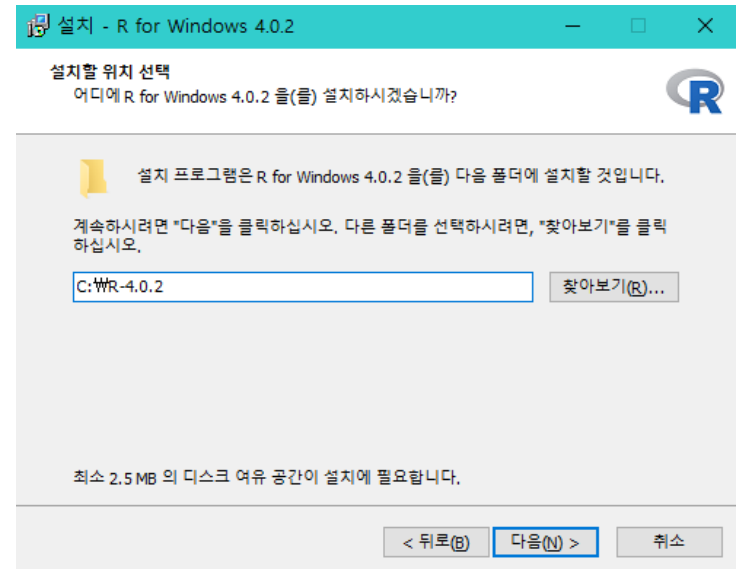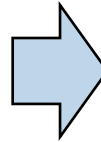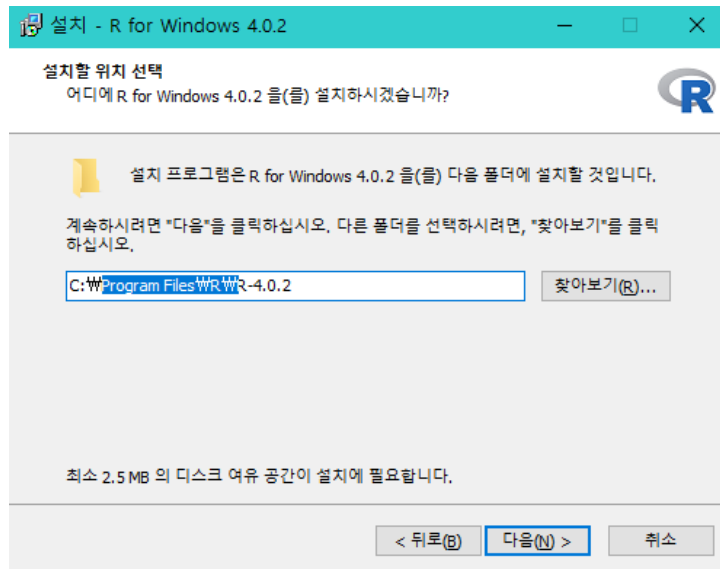
R tools 설치

- R tools: a collection of tools necessary for building R packages in Window PC

  C, C++, Fortran 코드로 제공되는 라이브러리를 설치하기 위한 도구 (MinGW)

  - https://cran.r-project.org/bin/windows/Rtools/

※ R 설치 버전과 버전을 맞출 것

Using Rtools40 on Windows

Starting with R 4.0.0 (released April 2020), R for Windows uses a brand new toolchain bundle called **rtools40**.

This version of Rtools upgrades the mingw-w64 gcc toolchains to version 8.3.0, and introduces a new build system based on msys2, which makes easier to build and maintain R itself as well as the system libraries needed by R packages on Windows. For more information about the latter, follow the links at the bottom of this document.

This documentation is about rtools40, the current version used for R 4.0.0 and newer. For information about previous versions of Rtools that can be used with R 3.6.3 or older, please visit this page.

Installing Rtools40

Note that rtools40 is only needed build R packages with C/C++/Fortran code from source. By default, R for Windows installs the precompiled "binary packages" from CRAN, for which you do not need rtools!

To use rtools40, download the installer from CRAN:

- On Windows 64-bit: rtools40-x86_64.exe (recommended: includes both i386 and x64 compilers)
- On Windows 32-bit: rtools40-i686.exe (i386 compilers only)

Note for RStudio users: please check you are using the latest version of RStudio (at least 1.2.5042) to work with rtools40.
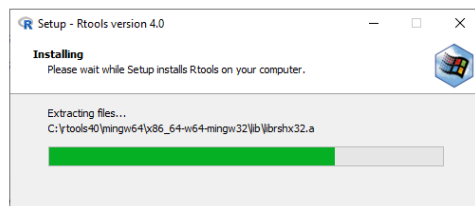
R tools 설치

- Rtools 설치 완료 되었는지 폴더 직접 가서 확인

- Rtools 설치 경로를 환경변수(PATH)에 추가

  제어판 → 시스템 및 보안 → 시스템 → 고급 시스템 설정 → 고급 → 환경 변수 → 사용자 변수

 **→ 새로 만들기 →**

R 설치 (R Engine 대안)

- MRO 배포 사이트 방문

  - https://mran.microsoft.com/download

R Studio 설치

- R Studio 배포 사이트 방문
  - https://rstudio.com/
  - Product → Rstudio 선택
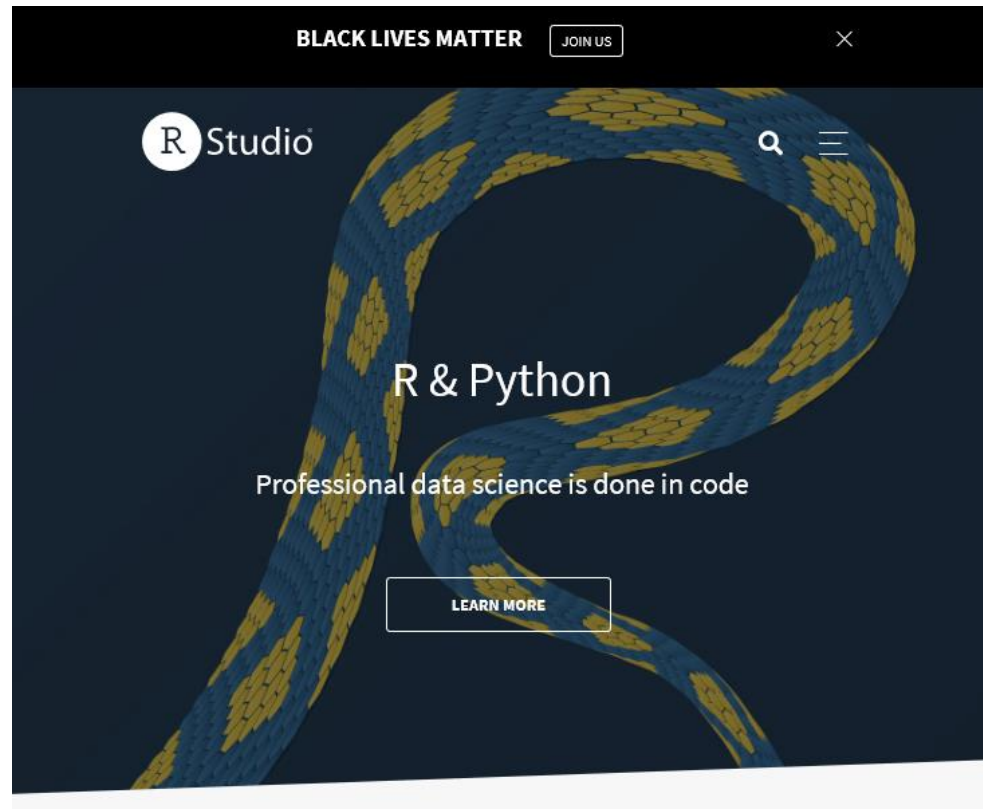
R Studio 설치

- R Studio desktop 선택

There are two versions of RStudio:

RStudio Desktop
Run RStudio on your desktop

RStudio Server
Centralize access and computation
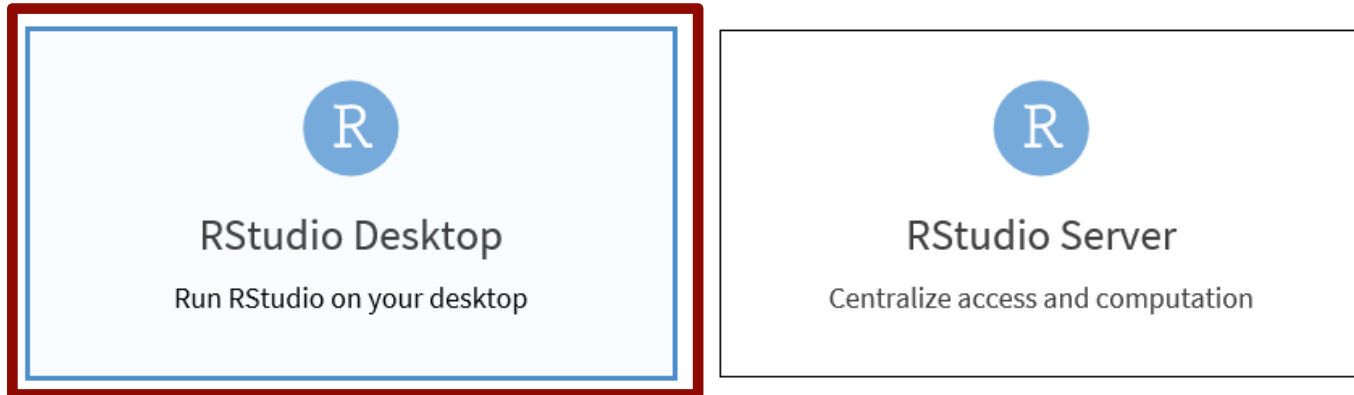
R Studio 설치

- 오픈소스 에디션 선택

R Studio 설치

- 아래와 같이 다운 받을 것

R Studio 실행

- Win 시작 → Rstudio 검색해서 실행

R Studio 실행해보자.

몇 가지 단축키를 알아 둡시다.

R은 라인 by 라인으로 실행시킬 수 있다.

특정 라인에서 'CTRL + ENTER'를 누르고 실행하거나,
아니면 라인들을 선택해서 실행하면 된다.
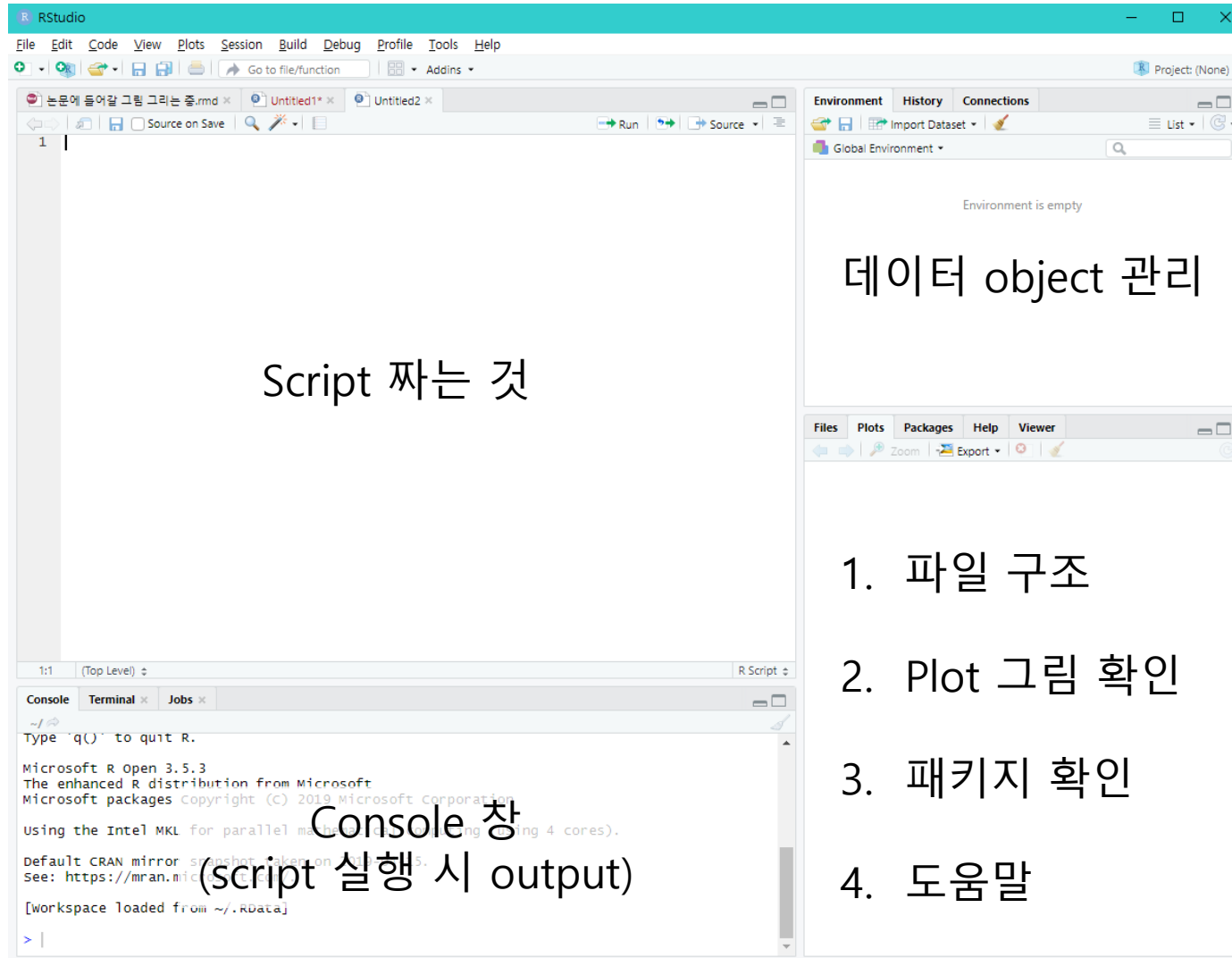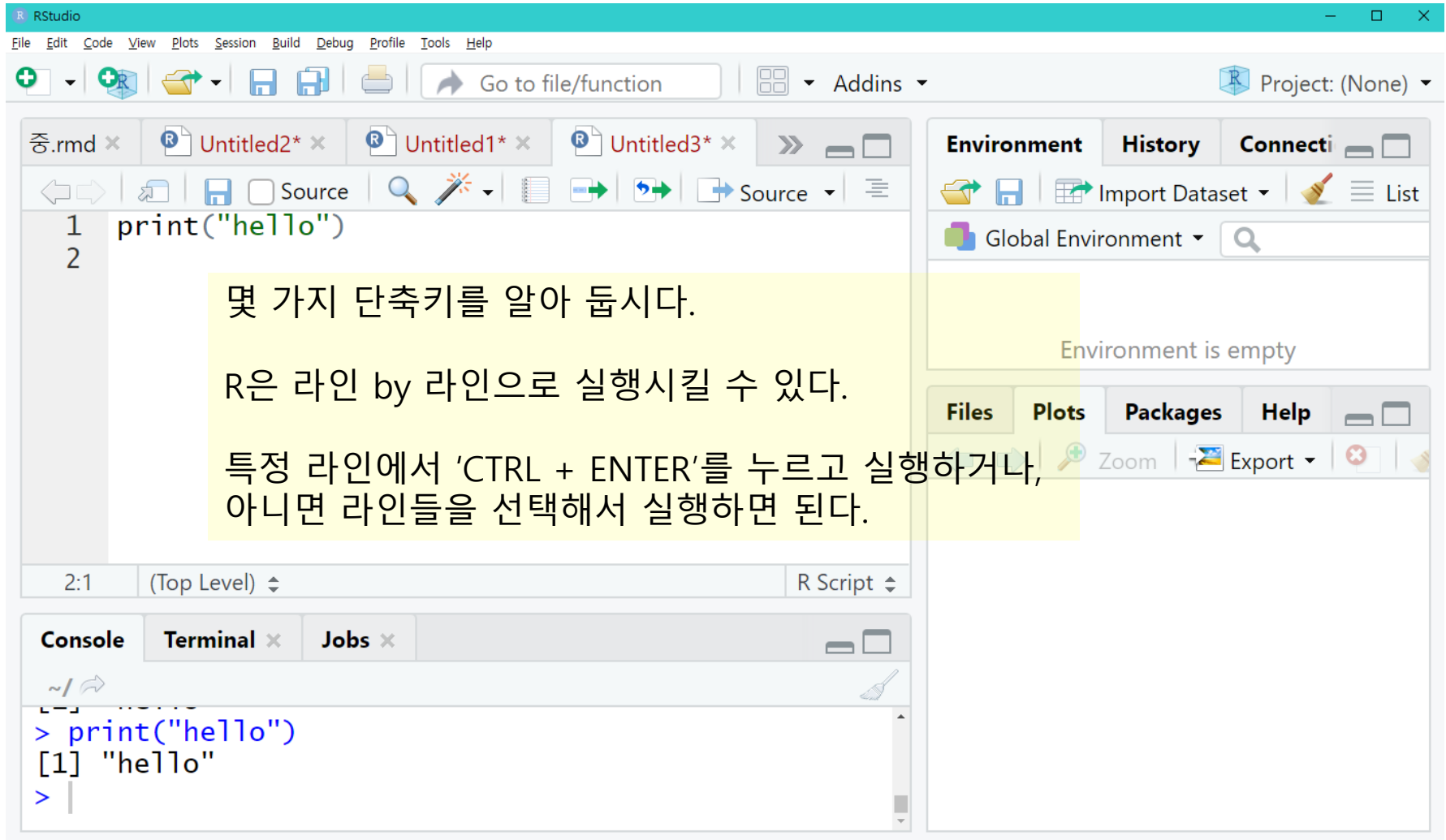
Summary

- R에 대해서 배워 봤습니다.

- R은 통계 패키지 도구 일 뿐만 아니라, 일반적인 Objective oriented programming 언어로도 쓰여집니다.

  - 특정 케이스(데이터)에 대한 실험 코드를 배포하는 것은 얼마든지 가능합니다.

  - 하지만, R로 naïve algorithm을 배포하는 것은 흔치 않습니다.

  - for, while loop과 같은 코딩을 하는 순간부터 무척 느려질 수 있습니다.

    → 이건 R, Python 공통의 문제 입니다.

    (Fibonacci number 문제 풀 수 있는 한계가 일반 PC에서 20을 넘지 못함,

    Java, C#은 30대, C는 40 이상)

- R Engine 설치, R 개발을 위한 R Studio, R Tools 설치했는지 확인하세요.

- 차주부터는 R Studio를 활용해서 R 스크립트 개발을 시작합니다.