

Introduction To R

R

- 데이터 분석을 위한 통계 및 그래픽스를 지원하는 자유 소프트웨어 환경
- 빅데이터 분석 기술을 통해 분석된 데이터의 의미와 가치를 시각적으로 표현하기 위한 대표적인 기술
- 벨 연구소에서 만들어진 통계 분석 언어 S를 근간으로 University of Auckland에서 Ross Ihaka와 Robert Gentleman이 만든 것이 시작
- 현재 가장 많이 사용되는 데이터 분석 도구 중 하나
- 컴퓨터 언어이자 다양한 패키지(또는 라이브러리)의 집합
 - R안에서 제약 없이 다양한 데이터 분석 수행 가능
 - 통계, 머신 러닝, 금융, 바이오 인포매틱스, 그래픽스에 이르는 다양한 통계 패키지를 무료로 제공

R

■ 적용 분야

- 통계 분석, 머신 러닝 모델링, 텍스트 마이닝, 소셜 네트워크 분석, 지도 시각화, 주식 분석, 이미지 분석, 사운드 분석 등 다양한 분야에서 활용

■ 장점

- 무료로 사용할 수 있는 오픈 소스
- 오픈 소스 생태계 → 전 세계 개발자로부터 높은 수준의 다양한 분석 도구가 생산되고 공유됨
- 많은 사용자 → 풍부한 참고 자료
- 다양하고 강력한 시각화 도구
- 프로그래밍 방식
 - » 재현성 확보
 - » 오류 감소, 상대적으로 쉬운 오류 수정
 - » 공동 작업

프로그래밍 언어 순위

■ TIOBE Index (2020.09)

Sep 2020	Sep 2019	Change	Programming Language	Ratings	Change
1	2	▲	C	15.95%	+0.74%
2	1	▼	Java	13.48%	-3.18%
3	3		Python	10.47%	+0.59%
4	4		C++	7.11%	+1.48%
5	5		C#	4.58%	+1.18%
6	6		Visual Basic	4.12%	+0.83%
7	7		JavaScript	2.54%	+0.41%
8	9	▲	PHP	2.49%	+0.62%
9	19	▲	R	2.37%	+1.33%
10	8	▼	SQL	1.76%	-0.19%

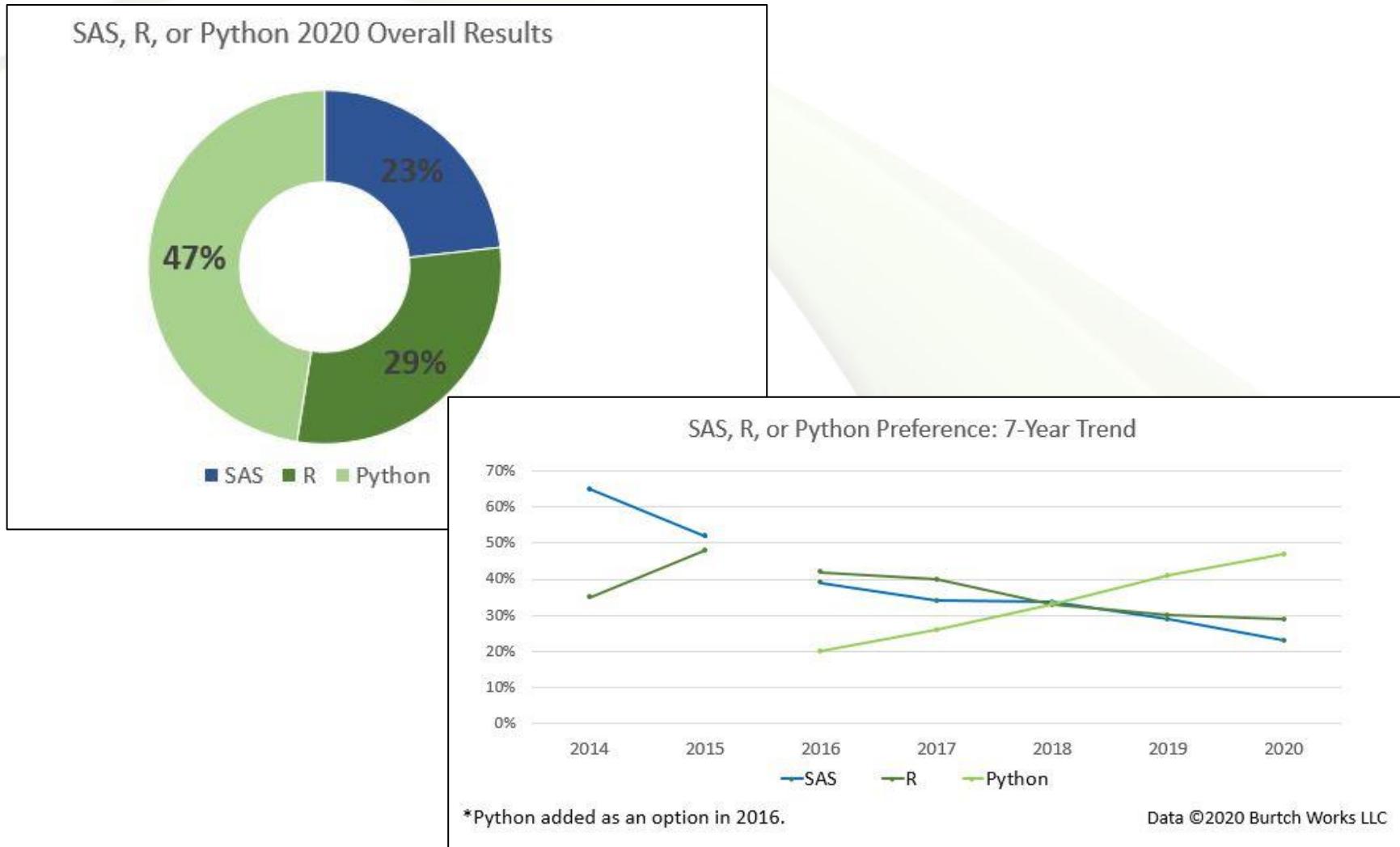
프로그래밍 언어 순위

■ pypl Index (2020.09)

Worldwide, Jan 2021 compared to a year ago:				
Rank	Change	Language	Share	Trend
1		Python	30.44 %	+1.2 %
2		Java	16.76 %	-2.0 %
3		JavaScript	8.44 %	+0.3 %
4		C#	6.53 %	-0.7 %
5	↑	C/C++	6.33 %	+0.3 %
6	↓	PHP	6.05 %	-0.2 %
7		R	3.87 %	+0.1 %
8		Objective-C	3.71 %	+1.2 %
9		Swift	2.14 %	-0.3 %
10		TypeScript	1.78 %	-0.0 %

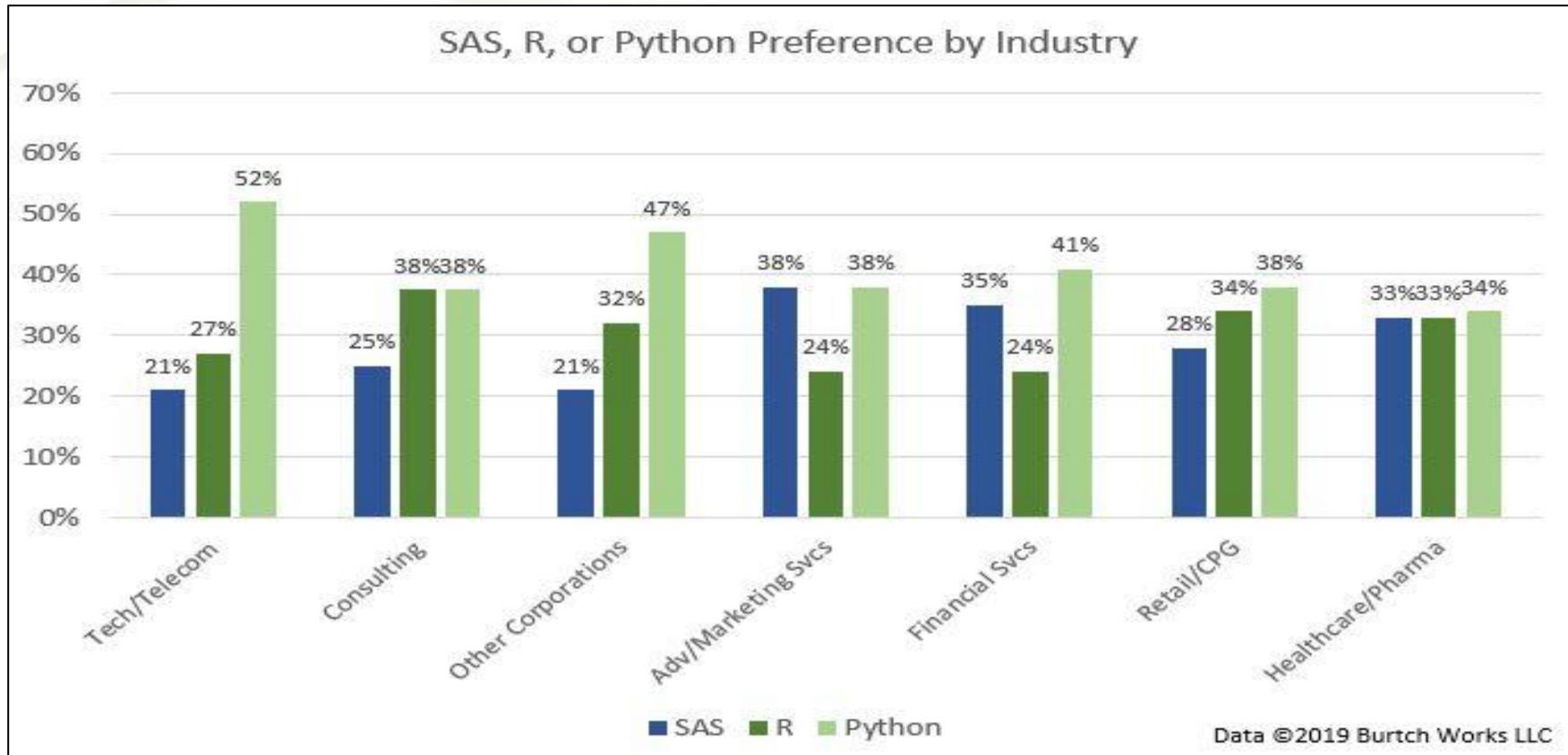
R vs Python vs SAS

- data scientists and analytics professionals survey



R vs Python vs SAS

- data scientists and analytics professionals survey



R 관련 링크

- <http://www.statmethods.net/>
- <http://www.cookbook-r.com/>
- <https://www.r-bloggers.com/>
- <http://stackoverflow.com/tags/r/info>
- <http://stackoverflow.com/questions/tagged/r-faq>
- <https://google.github.io/styleguide/Rguide.xml>

분석 환경 구축

▪ R 설치

- <https://cran.r-project.org>에서 설치 파일 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](#)

운영체제 환경에 따라 파일 다운로드

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

분석 환경 구축

▪ RStudio 설치

- <https://rstudio.com/products/rstudio/download/#download>

OS	Download	Size	SHA-256
Windows 10/8/7	 RStudio-1.4.1103.exe	156.96 MB	c3384189
macOS 10.13+	 RStudio-1.4.1103.dmg	152.77 MB	20148bd6
Ubuntu 16	 rstudio-1.4.1103-amd64.deb	119.26 MB	f0857e27
Ubuntu 18/Debian 10	 rstudio-1.4.1103-amd64.deb	120.30 MB	76864349
Fedora 19/Red Hat 7	 rstudio-1.4.1103-x86_64.rpm	138.02 MB	8fcbb2d29
Fedora 28/Red Hat 8	 rstudio-1.4.1103-x86_64.rpm	138.01 MB	e2bf11e9
Debian 9	 rstudio-1.4.1103-amd64.deb	120.45 MB	4a4d159c
OpenSUSE 15	 rstudio-1.4.1103-x86_64.rpm	122.02 MB	fdc33f7a

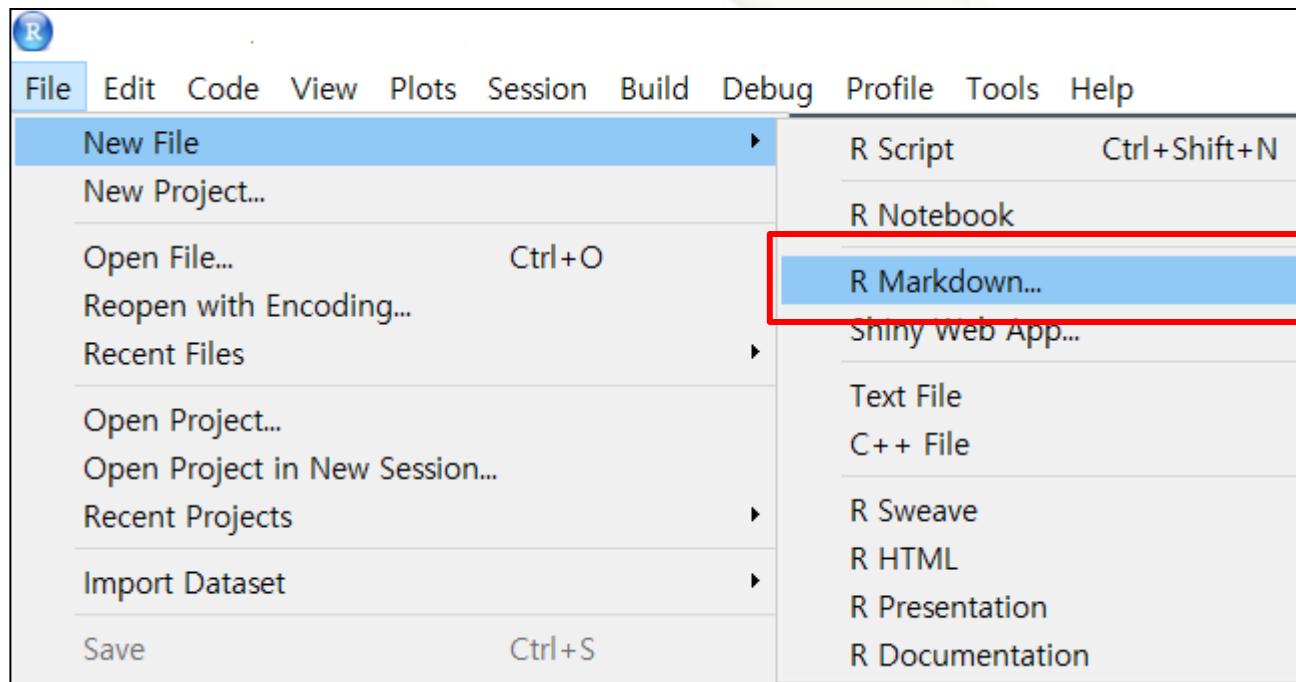
- 다운로드 완료 후 설치

설치할 컴퓨터의 운영체제에 따라 파일 다운로드

분석 환경 구축

▪ R Markdown 사용

- RStudio 실행 → 주 메뉴 → File → New File → R Markdown... 클릭
- 처음 실행하면 자동으로 마크다운 설치



분석 환경 구축

- Java 개발 환경 설치
 - 자바 프로그램을 개발하고 실행하는데 필요한 도구
 - R의 일부 패키지들이 자바를 기반으로 실행되기 때문에 경우에 따라 필요

설치 과정은 별도의 설치 안내 파일 (`jdk_installation.pdf`) 참고

명령 입력 및 실행 - RGui

- RGui 실행
- 명령 입력 및 실행 (enter)

The screenshot shows the RGui application window with the title "RGui (64-bit)". The menu bar includes "파일", "편집", "보기", "기타", "패키지를", "윈도우즈", and "도움말". The toolbar contains icons for file operations like Open, Save, Print, and Stop. The main area is titled "R Console".

한 줄에 입력 완성

```
> print("hello")
[1] "hello"
```

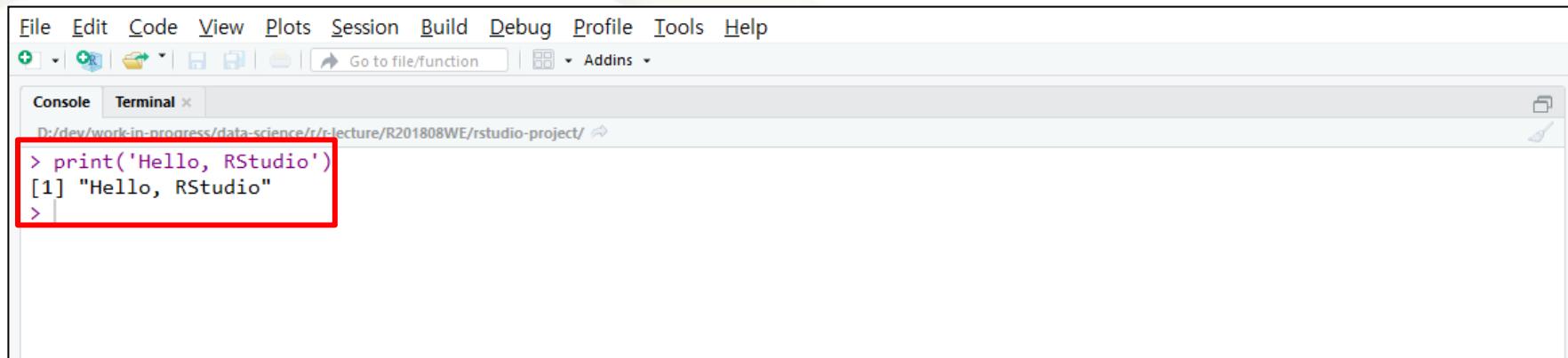
여러 줄에 입력 완성

```
> print(
+ "hello"
+ )
[1] "hello"
```

The bottom-left code block is highlighted with a red border. The right-hand code blocks are also highlighted with red borders.

명령 입력 및 실행 - RStudio

- Rstudio 실행
- Console View에서 명령어 입력 및 실행 (enter)



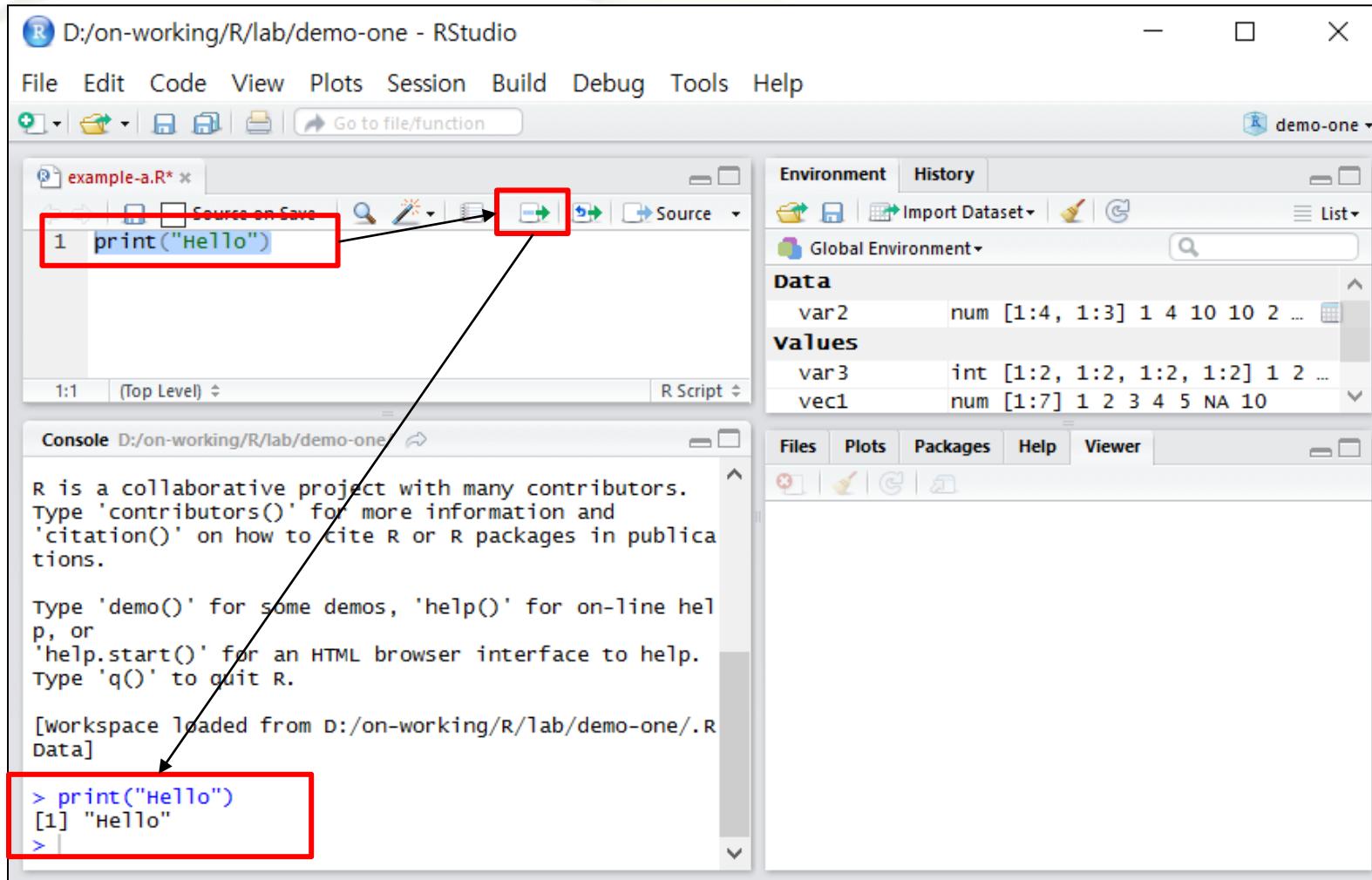
The screenshot shows the RStudio IDE with the 'Console' tab selected. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons. The console window displays the following R code and output:

```
D:/dev/work-in-progress/data-science/r/lecture/R201808WE/rstudio-project/ ↵
> print('Hello, RStudio')
[1] "Hello, RStudio"
>
```

The first two lines of the code ('> print('Hello, RStudio')' and '[1] "Hello, RStudio"') are highlighted with a red rectangular box.

명령 입력 및 실행 - RStudio

- Rstudio 실행
- File → New → Rscript → 명령 입력 → 실행버튼 클릭 (Ctrl + Enter)



배치 실행 - 파일에 저장된 스크립트 실행

- 스크립트 파일 작성

```
hello-script.R - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
print("Hello from File")
print("확인해 주세요")
```

- 스크립트 파일 실행

- source("상대경로" or "절대 경로")

example-a.R*

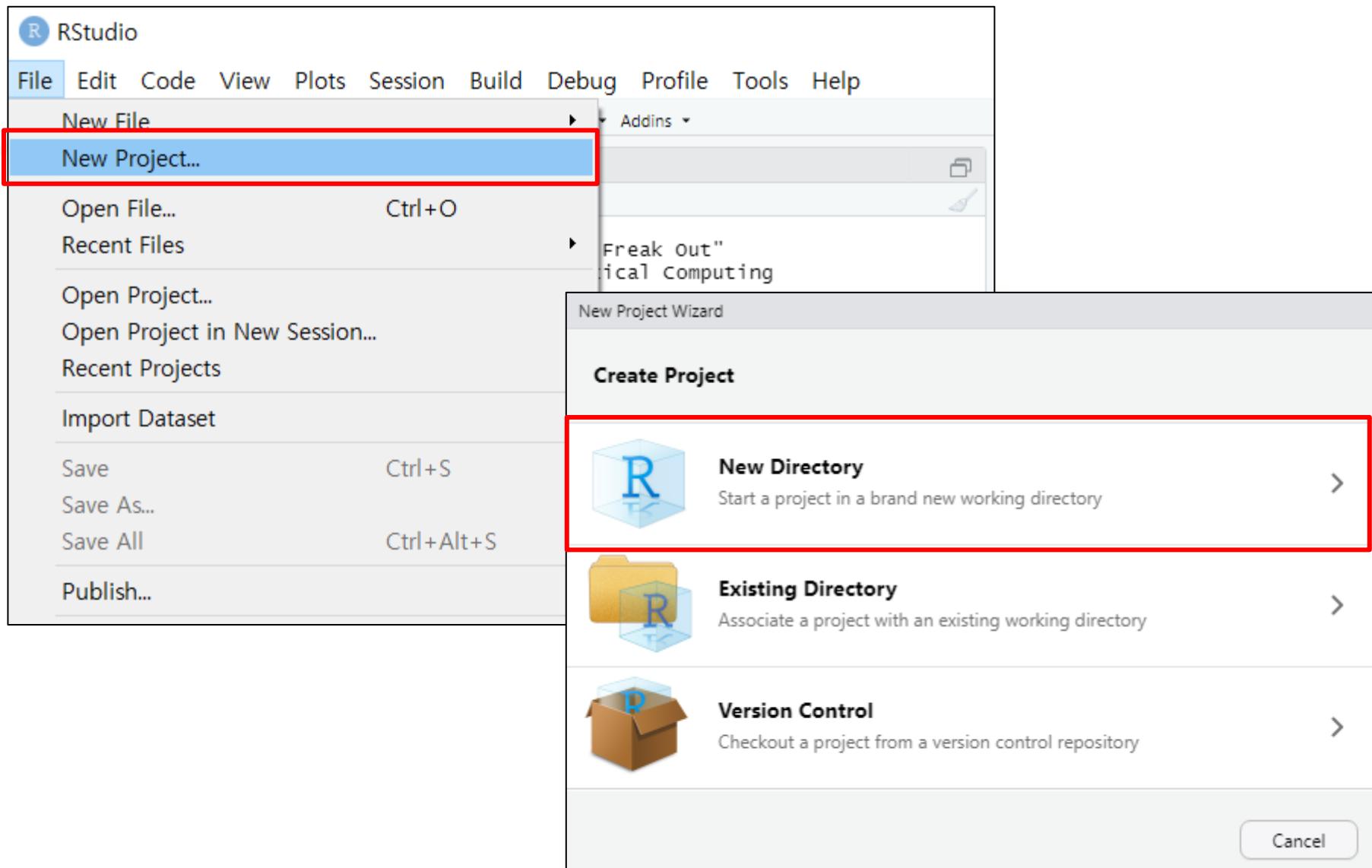
```
1 source("hello-script.R")
2 |
```

Source on Save | Source

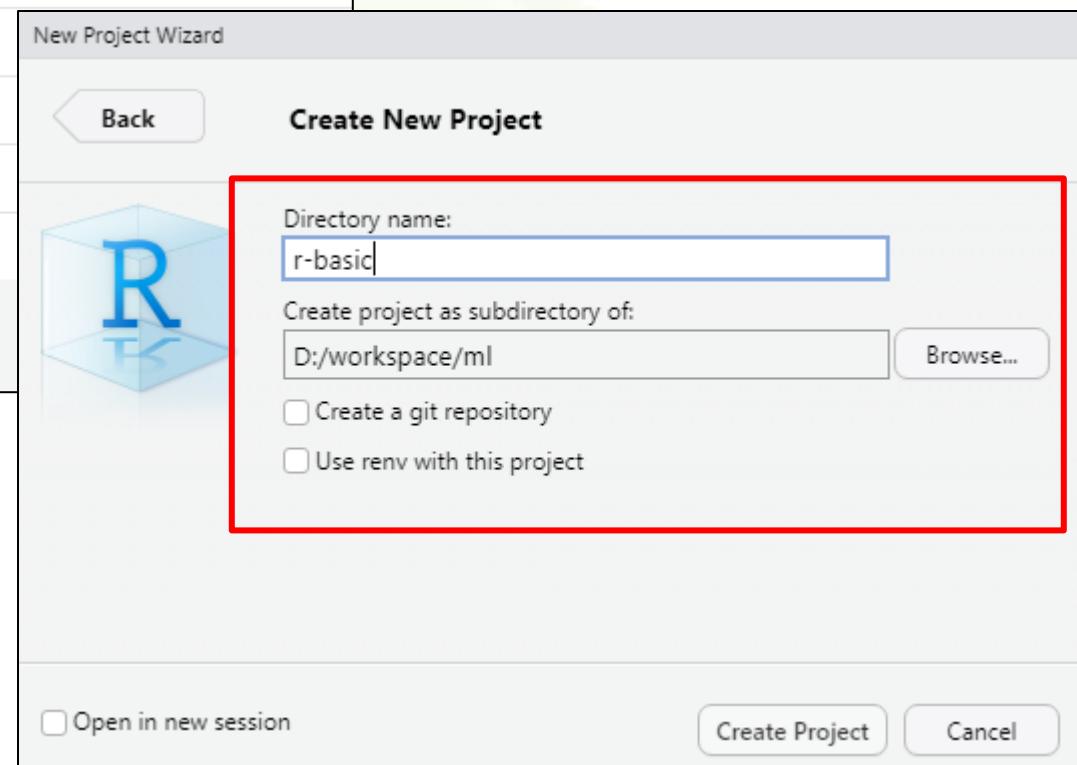
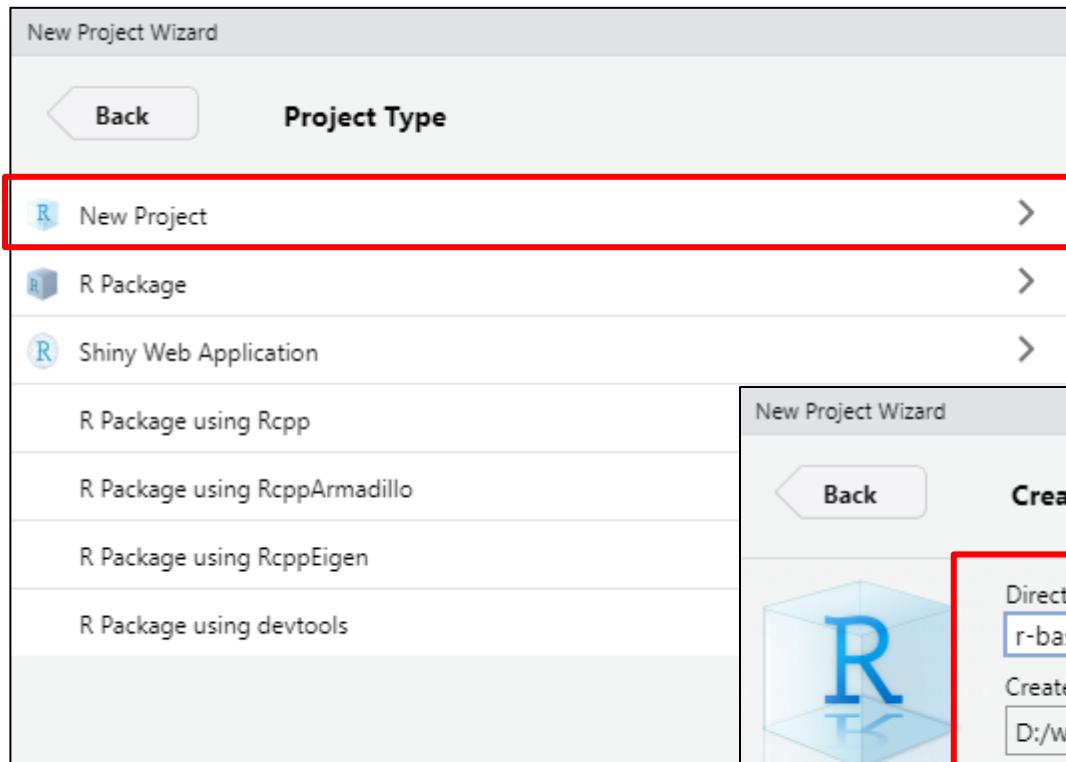
2:1 (Top Level) R Script

> source("hello-script.R")
[1] "Hello from File"
[1] "확인해 주세요"
> |

RStudio 프로젝트 사용



Rstudio 프로젝트 사용



도움말 보기

명령어

설명

`help.start()`

R 온라인 문서 열기

`help("명령어") or ?명령어`

명령어에 대한 도움말

`help.search("명령어") or ??명령어`

명령어에 관한 정보 검색

`example("명령어")`

명령에 대한 사용 사례

`RSiteSearch("명령어")`

온라인 도움말과 메일링 리스트에서 명령어 검색

`vignette()`

`vignette()`를 호출해서 설명을 볼 수 있는 목록 표시

`vignette(" 이름")`

지정된 패키지에 대한 상세 안내 문서 표시