

빅데이터 분석 결과 시각화

빅데이터 분석 프로그램

학습내용

- R 프로그램 설치 방법 알아보기
- RStudio UI 분석
- Rscript 실습

학습목표

- R 프로그램 설치 방법을 숙지하여 빅데이터를 분석할 수 있다.
- Rstudio UI에 대해 분석할 수 있다.
- RScript를 사용하여 프로그래밍할 수 있다.

● 빅데이터 정의 및 프로그램 종류

1. 빅데이터 정의

◆ 빅데이터란?

- 빅데이터 : 데이터를 구성하고 있는 물리적 하드웨어로부터 시작해 이를 기반으로 하는 애플리케이션과 소프트웨어를 포괄하는 거대 플랫폼
 - 단순한 데이터의 크기가 아니라 데이터의 형식과 처리 속도 등을 함께 아우름
 - 기존 방법으로는 데이터의 수집, 저장, 검색, 분석 등이 어려운 데이터를 총칭
- 빅데이터 플랫폼 : 거대한 데이터를 관리하기 위한 하드웨어 및 소프트웨어 등 모든 프로세스
 - 예) 일부 소셜 네트워크 서비스에서 발생하는 매달 300억 개의 새로운 콘텐츠

◆ 빅데이터 활용 사이트

- 구글 번역
 - 66개 나라의 언어 패턴과 단어와 문장을 분석하여 구글에서 만든 번역 시스템
 - 데이터가 많아질수록 문장의 번역률이 정확해짐
- 위키피디아
 - 누구나 자유롭게 작성하고 수정할 수 있는 다국적 온라인 백과사전
 - 위키피디아에 저장된 데이터를 활용할 수 있음
- 구글 트렌드
 - 해당 검색어를 지역, 기간, 시간 흐름에 따른 관심도를 통계 데이터로 보여줌

● 빅데이터 정의 및 프로그램 종류

2. 빅데이터 프로그램 종류

◆ 빅데이터 분석 프로그램 종류

● R 프로그램

- 통계를 위한 프로그래밍 언어
- 통계 프로그래밍 언어인 S, 함수형 언어인 스킴(Scheme)과 XlispStat의 영향을 받아 개발

Excel	R
<ul style="list-style-type: none">• 수억 건의 통계 자료를 기계적으로 처리하기 어려움• 데이터를 수집하기 어려움	<ul style="list-style-type: none">• 데이터를 통계낸 후 해석하고 의사결정 하는 데 효율적임

● R 프로그램의 장점

- 무료(오픈소스) 공개 소프트웨어
- 다른 빅데이터 분석 프로그램인 하둡SAS,SPSS와 같은 전문적인 통계 프로그램보다 배우기 쉬움
- 텍스트, 엑셀, DBMS 등 다양한 종류의 정형/비정형 데이터 이용이 가능
- 윈도우, 유닉스, 리눅스, 맥OS 등 다양한 플랫폼에서 수집/저장/분석/ 시각화가 가능하며 통계가 가능

● 하둡(Hadoop) 프로그램

- 대량의 자료를 처리할 수 있는 컴퓨터 클러스터의 분산 응용 프로그램을 지원하는 자바 소프트웨어 프레임워크

● 파이썬(Python) 프로그램

- 초보자가 배우기 쉽고 간결한 언어로 우수한 성능과 빠른 처리 속도와 데이터 분석 라이브러리를 제공함
- 데이터 분석, 금융, 자연 과학 등 다양한 분야에서 각광 받음
- 특히, 기계 학습, 자연어 처리, 패턴 인식 등 데이터 과학을 학습하기 위한 언어로 좋음

● RStudio UI 분석

1. R 프로그램 설치

◆ R 프로그램 환경설정

- R 프로그램
 - 오픈 소스로 만들어진 통계 분석 프로그램
- Java 프로그램
 - R에서 데이터분석을 할 때 자바 관련 라이브러리를 활용할 수 있도록 해주는 프로그램
- Rstudio
 - R 프로그램을 편리하게 사용할 수 있는 환경을 제공하는 프로그램

R 프로그램	Rstudio
<ul style="list-style-type: none">• DOS 상태의 인터프리터 방식의 실행을 통해 프로그래밍	<ul style="list-style-type: none">• R 프로그램 환경에서 스크립트 언어를 반복 사용할 수 있는 스크립트창, 히스토리창, 파일창, 환경설정창, 차트창 등을 제공하여 더욱 편리한 환경을 제공함

● RStudio UI 분석

1. R 프로그램 설치

◆ R 프로그램 설치

- ① <http://cran.r-project.org>에 접속해서 R 최신버전 다운로드 받음
- ② 다운 받은 R 프로그램 설치
- ③ 본인의 OS 버전에 맞는 것만 선택해서 설치
- ④ R 프로그램이 설치 완료 된 후 R.exe 실행버튼을 누르면 R 프로그램이 실행됨

◆ JDK설치하기

- ① <http://www.oracle.com> 에 접속
- ② 본인의 OS 버전에 맞는 것만 JAVA SE를 선택
- ③ JDK 설치 진행
- ④ JDK 설치가 완료되면 시스템 환경설정에서 시스템 변수를 설정

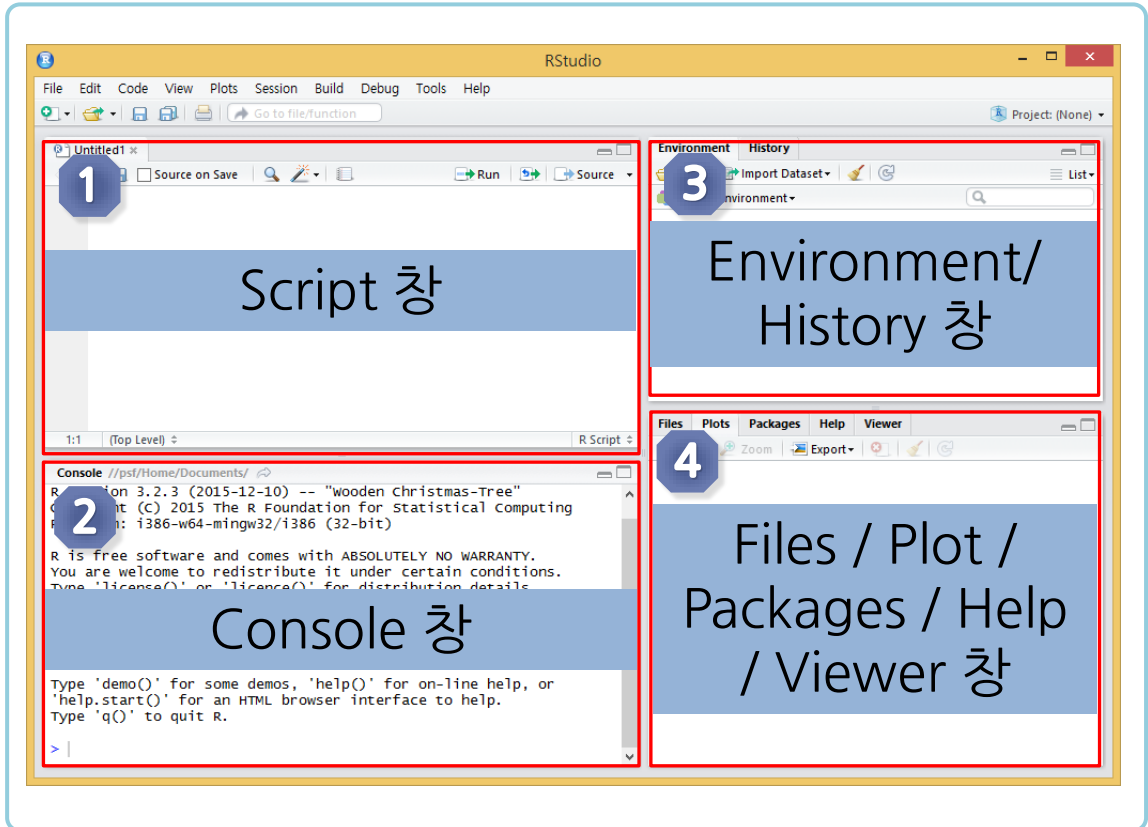
◆ Rstudio 설치하기

- ① <http://www.rstudio.com>에 접속
- ② 본인의 OS 버전에 맞는 것 선택해서 설치
- ③ 설치 버튼을 눌러 Rstudio를 설치
- ④ 설치 완료 후 rstudio.exe 실행 버튼을 눌러 실행

- RStudio UI 분석

2. RStudio 화면 구성 알아보기

- ◆ RStudio 화면 구성

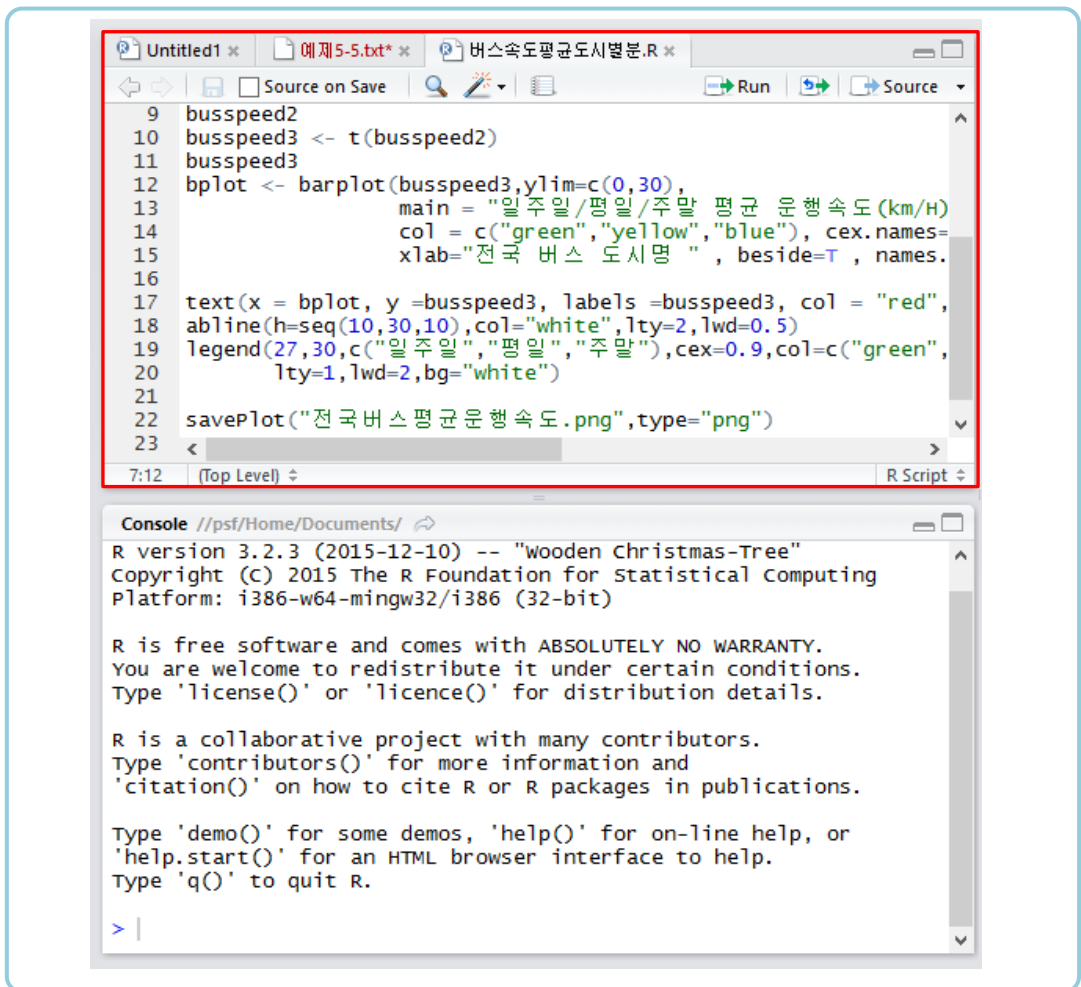


● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Script 창

- 기본 명령어를 사용하여 RScript를 실행 또는 저장할 수 있음
 - RScript 창에서 작성한 Script 실행 방법
 - Ctrl + R
 - Ctrl + Enter
 - 블록 설정 후 RStudio Script 창의 상단 메뉴인 'Run' 단추 클릭

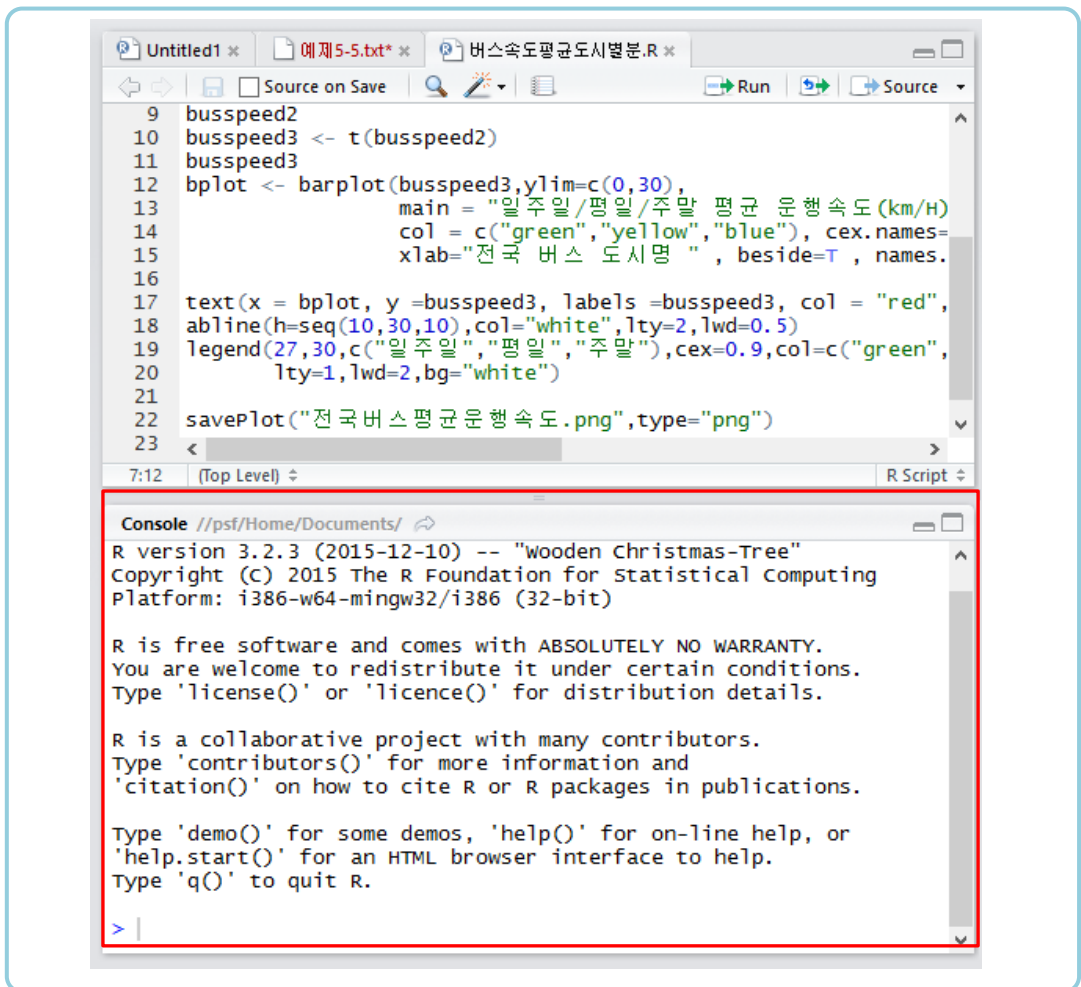


● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Console 창

- RScript를 Run할 때 한 라인씩 실행되는 화면이 보이는 곳
 - Console 창 활용
 - Interactive 하게 R 프로그램 짜고 실행
 - RScript 창 혹은 Console 창에서 작성한 프로그램의 실행 결과 보기
 - 패키지 설치, 에러/오류 메시지 등 로그 보기



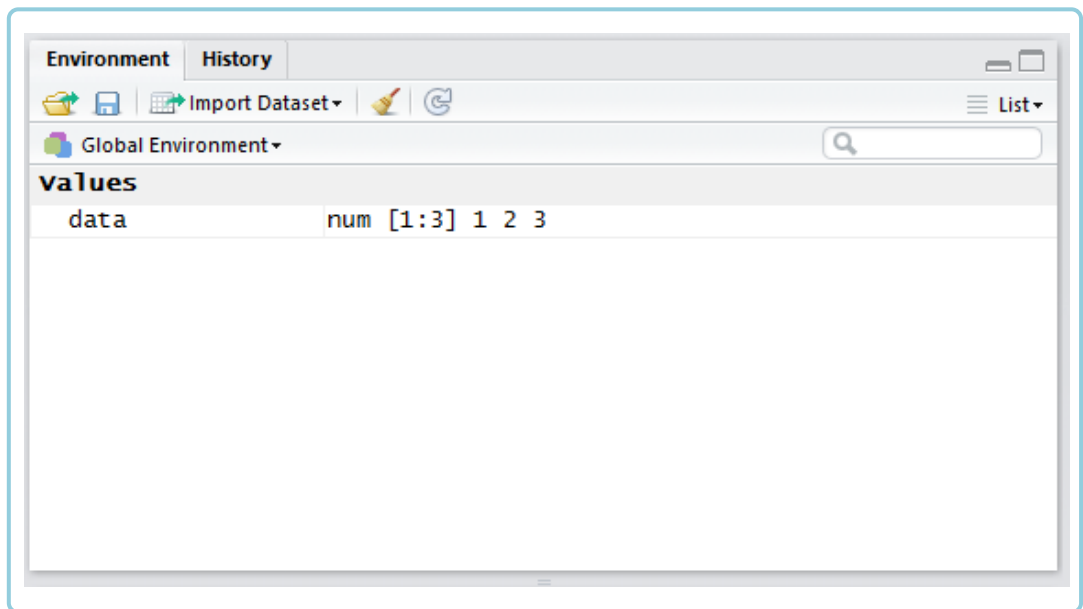
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Environment/History 창

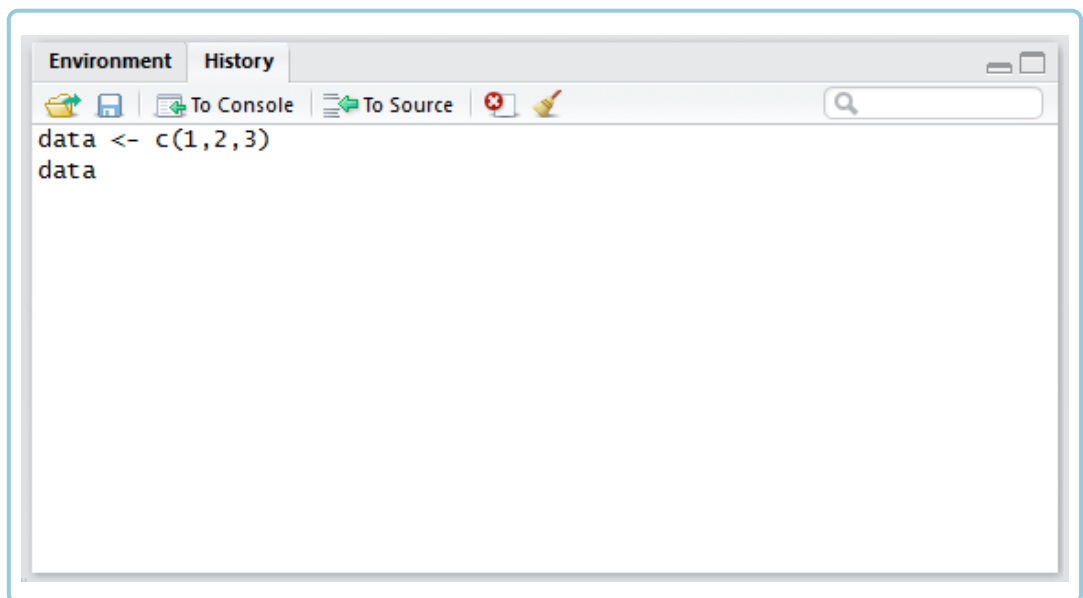
- Environment 탭

- 데이터셋의 이름, 관측치 개수, 변수 개수가 나옴



- History 탭

- RStudio 사용 History를 볼 수 있음



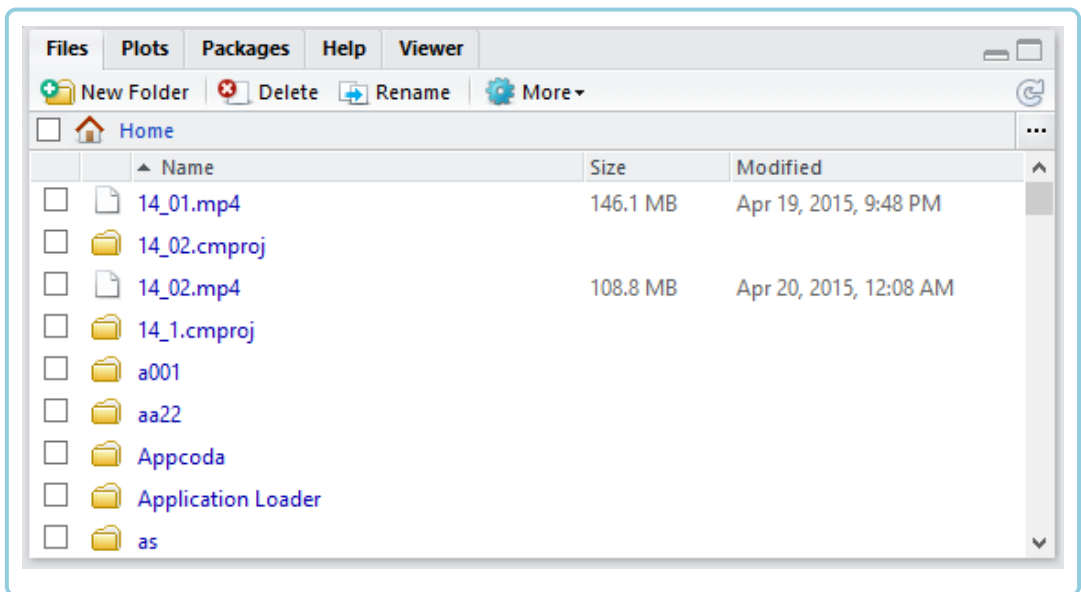
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Files / Plots / Packages / Help / Viewer 창

● Files 탭

- Windows의 탐색기처럼 활용
- 찾기, 신규 폴더 만들기, 파일 삭제하기, 이름 수정하기, 작업경로(Working directory) 설정하기 등이 가능



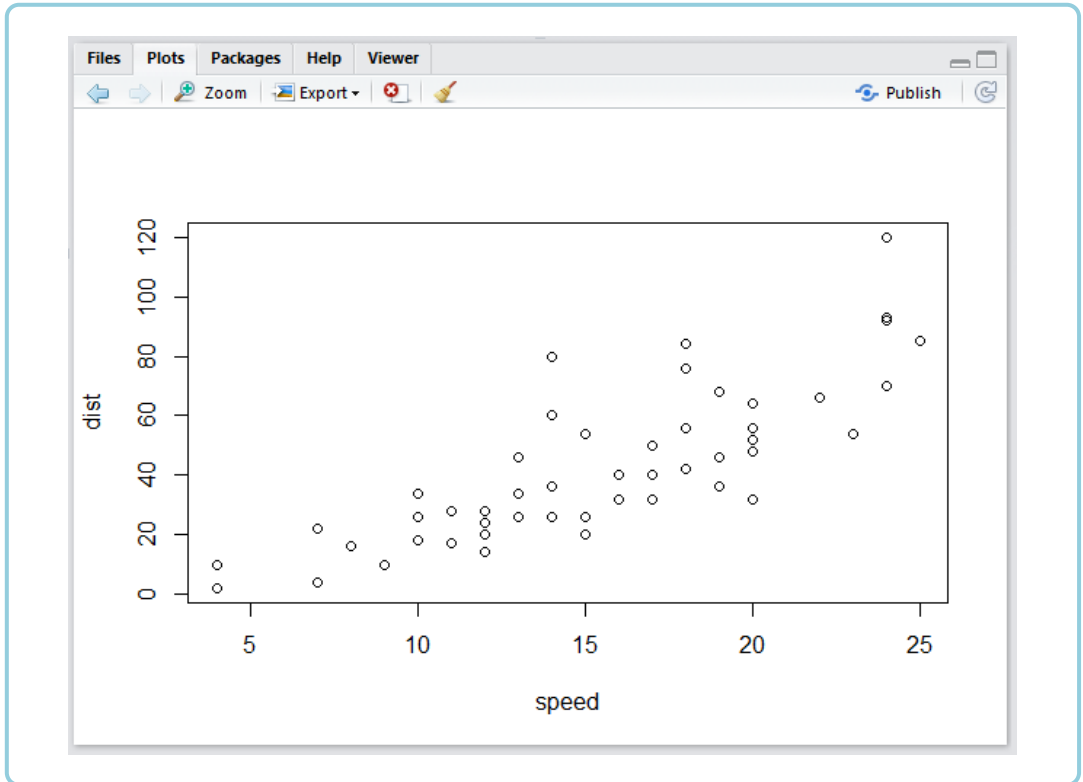
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Files / Plots / Packages / Help / Viewer 창

● Plots 탭

- 데이터의 시각화를 위한 공간
- 다양한 그림출력이 Plots 창을 이용하여 출력됨



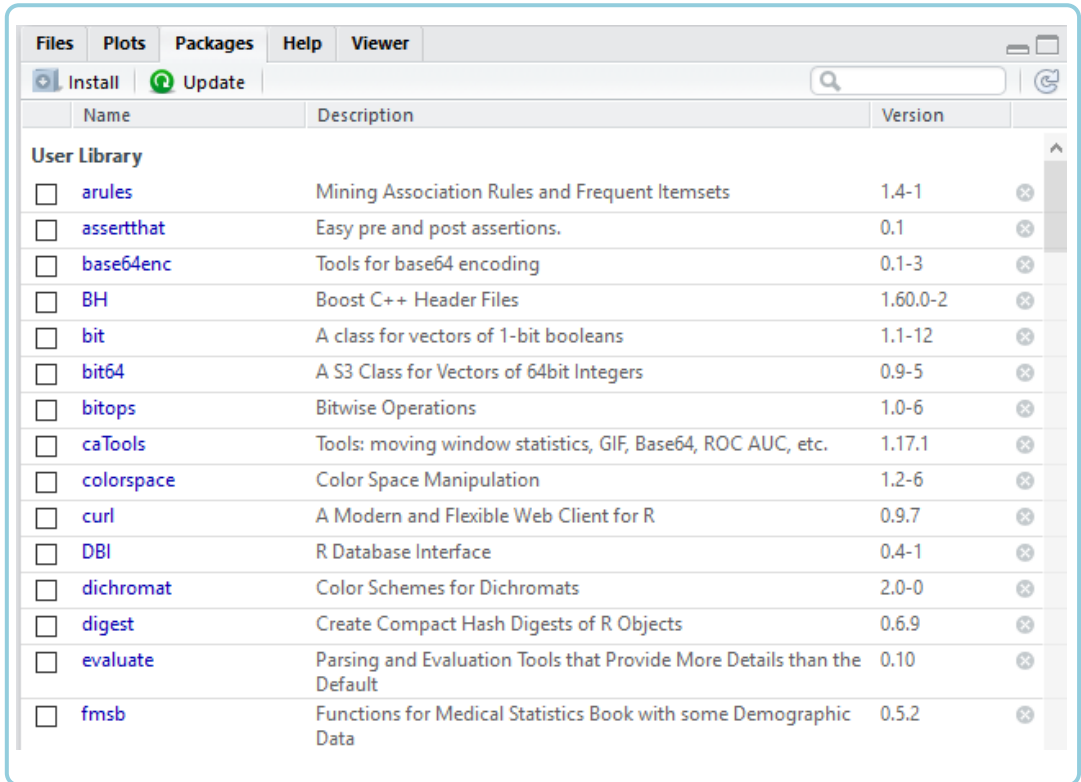
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Files / Plots / Packages / Help / Viewer 창

● Packages 탭

- R 프로그램은 다양한 사용자들이 원하는 함수들을 패키지라는 형태로 배포
- 이 패키지를 관리하는 창



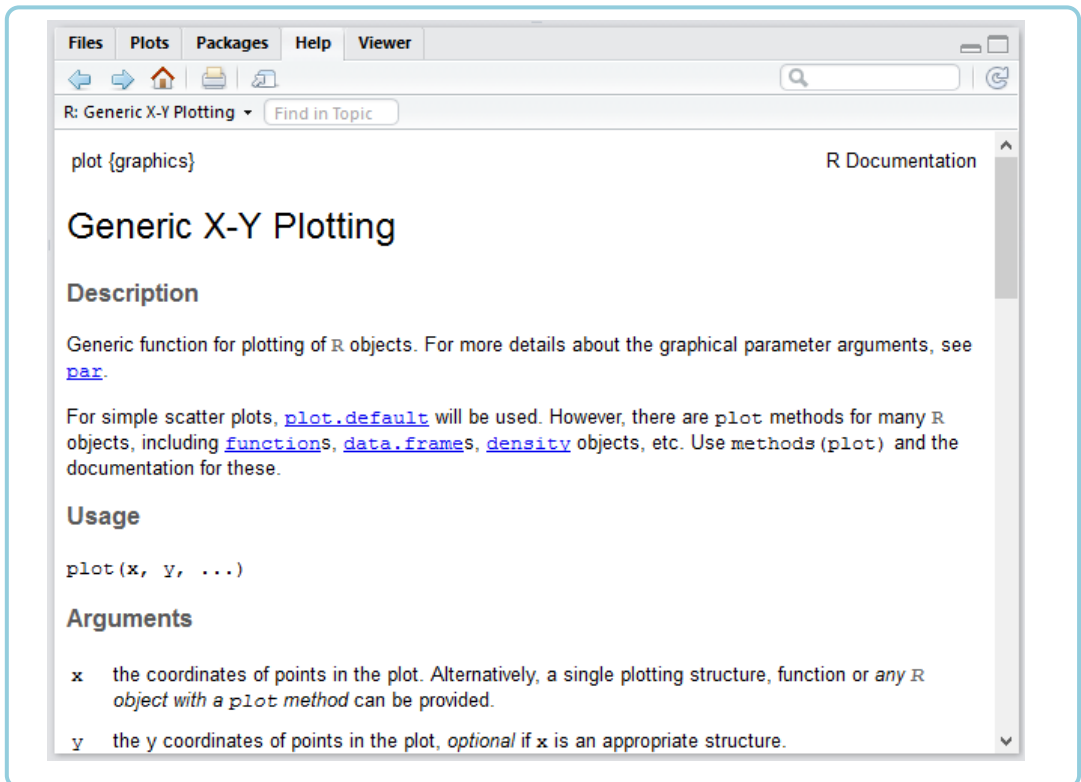
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Files / Plots / Packages / Help / Viewer 창

● Help 탭

- R 프로그램의 모든 함수들은 Help 문서 지원



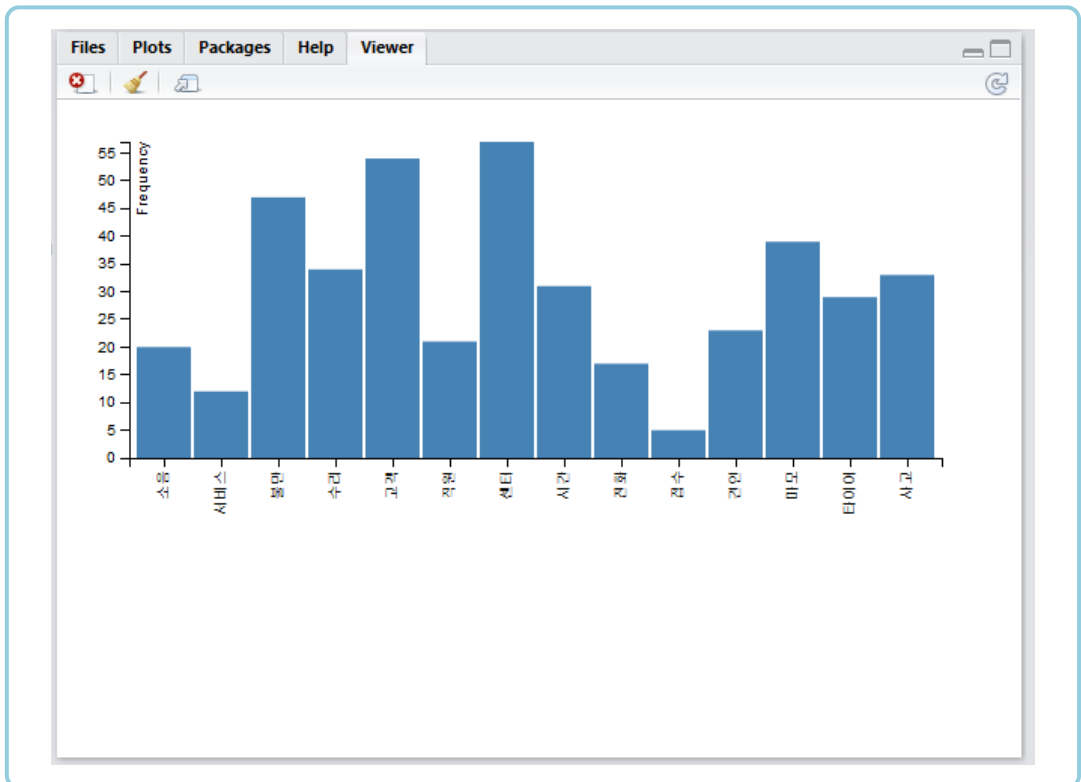
● RStudio UI 분석

2. RStudio 화면 구성 알아보기

◆ Files / Plots / Packages / Help / Viewer 창

● Viewer 탭

- 파이선과 같은 언어를 통해 데이터를 시각화할 때 보여주는 창

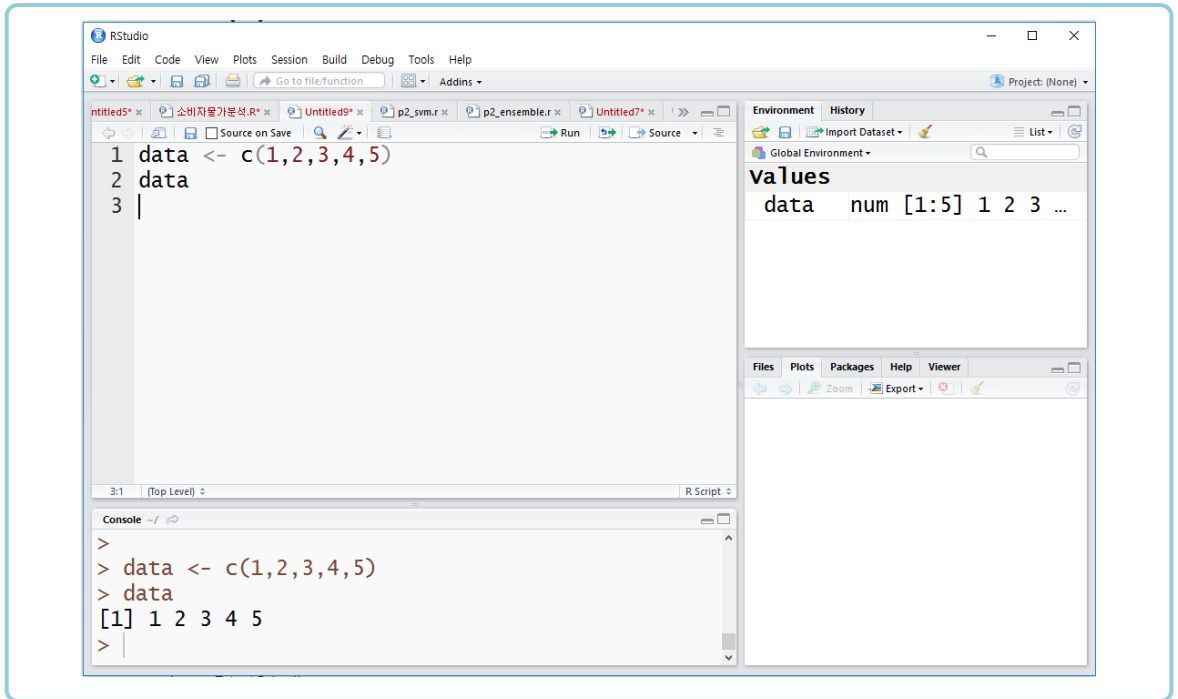


● RScript 실습

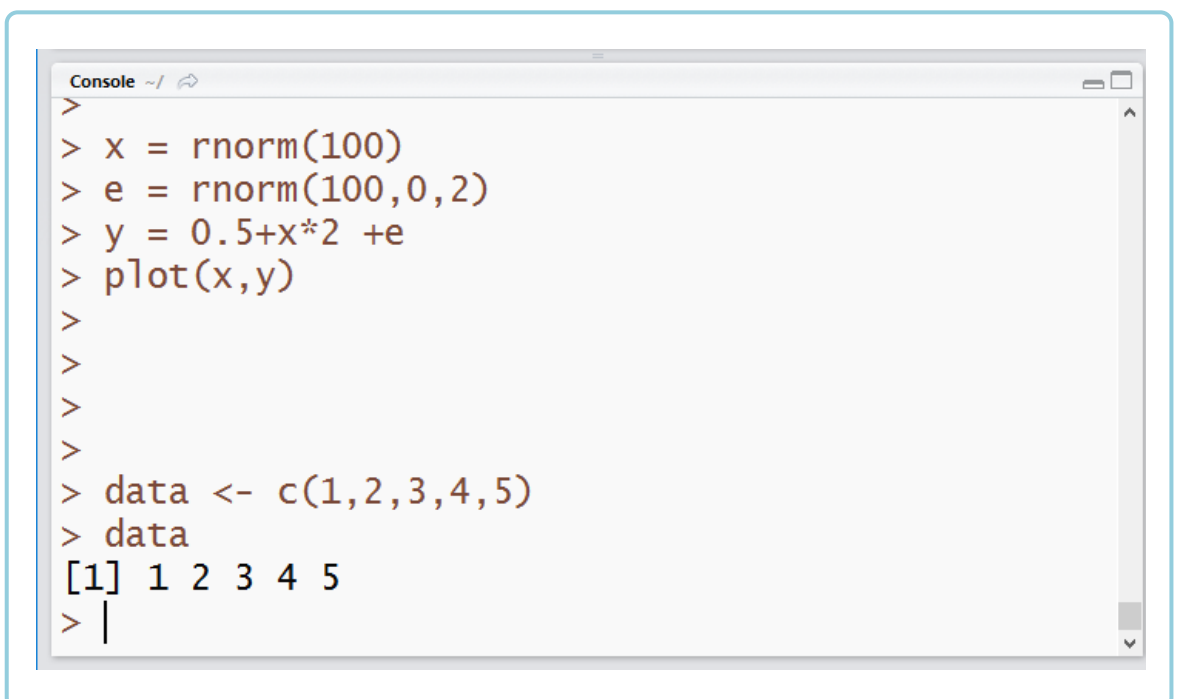
1. Rscript 언어 특징

① 인터프리터 형태의 언어

- Console 창에 직접 코딩해도 되지만 Script 창에서 코딩을 함으로써 전체 코드를 한꺼번에 제어할 수 있음



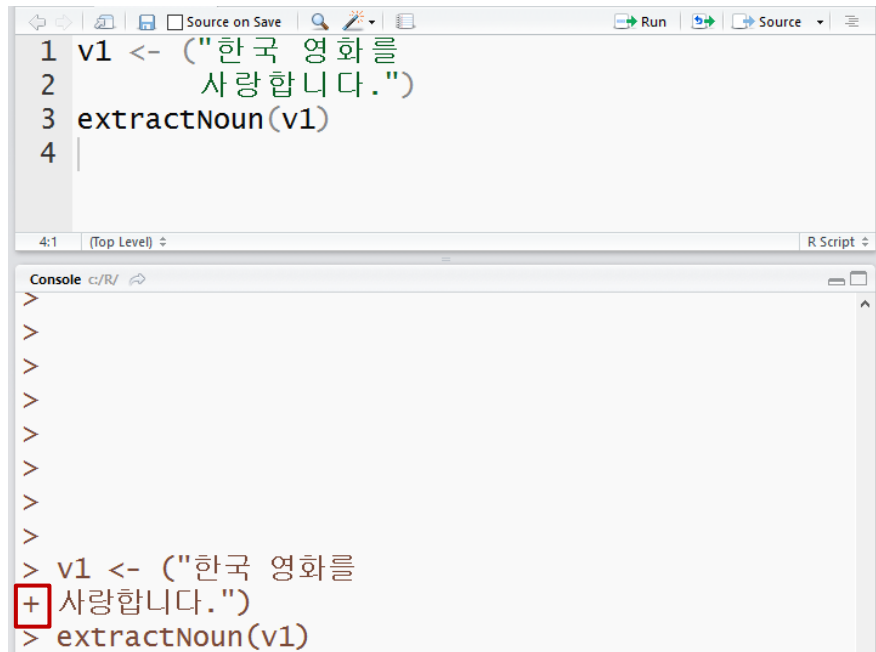
② ‘>’는 명령 프롬프트이고 명령을 실행하려면 명령 입력 후 Enter 누름



● RScript 실습

1. Rscript 언어 특징

- ③ 여러 줄에 명령 입력 시 다음 줄의 가장 왼쪽에 '+' 기호가 자동으로 생성 됨



The screenshot shows the RStudio interface. The editor window contains the following R code:

```
1 v1 <- ("한국 영화를  
2      사랑합니다.")  
3 extractNoun(v1)  
4 |
```

The console window shows the execution of the code, with the prompt character '>' appearing on each line. The third line of the console shows the prompt character '>' followed by a red box containing a '+' sign, indicating that RStudio automatically adds this character to the start of subsequent lines of input.

```
>  
>  
>  
>  
>  
>  
>  
>  
>  
> v1 <- ("한국 영화를  
+ 사랑합니다.")  
> extractNoun(v1)
```

● RScript 실습

1. Rscript 언어 특징

④ 대소문자 구분

- 소문자는 오류가 나고 대문자는 오류가 나지 않음

```
1 library(KoNLP)
2 v1 <- ("한국 영화를 사랑합니다.")
3 extractnoun(v1)
4 extractNoun(v1)
5
```

Console C:/R/ ↗

다음의 패키지를 주석합니다: KoNLP

The following object is masked from 'package:tau':

is.ascii

```
> v1 <- ("한국 영화를 사랑합니다.")
> extractnoun(v1)
Error: could not find function "extractnoun"
> extractNoun(v1)
[1] "한국" "영화" "사랑"
>
```

⑤ '#'은 주석기호

```
1 library(KoNLP)
2 v1 <- ("한국 영화를 사랑합니다.")
3 extractnoun(v1)
4 extractNoun(v1) # 명사를 추출하는 함수
5
```

Console C:/R/ ↗

다음의 패키지를 주석합니다: KoNLP

The following object is masked from 'package:tau':

is.ascii

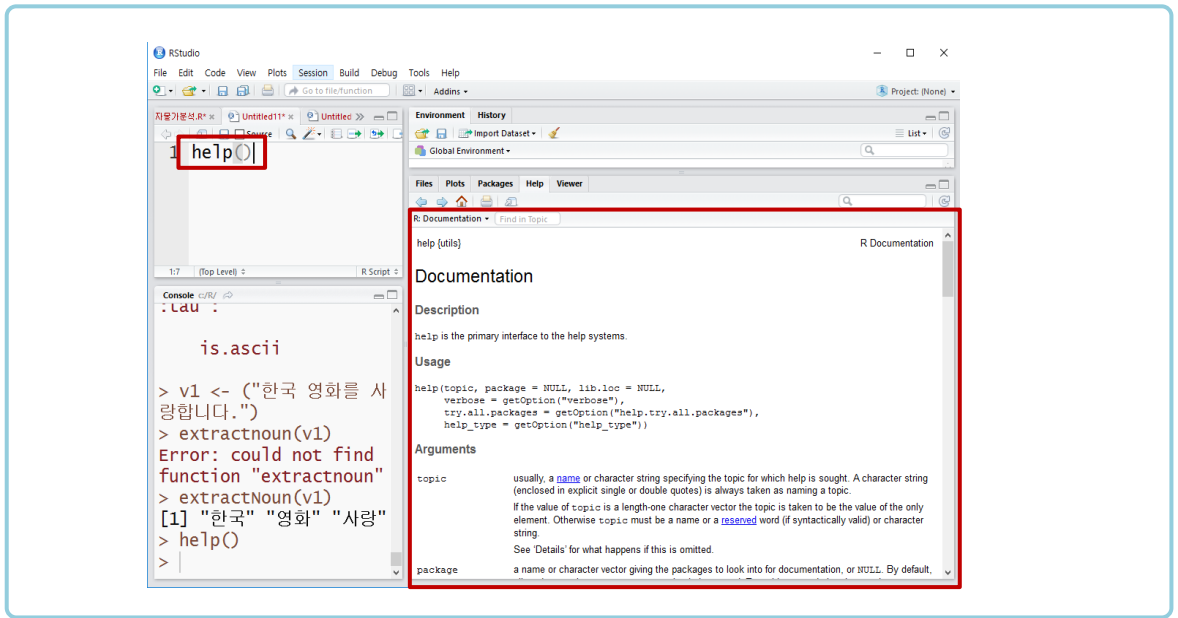
```
> v1 <- ("한국 영화를 사랑합니다.")
> extractnoun(v1)
Error: could not find function "extractnoun"
> extractNoun(v1)
[1] "한국" "영화" "사랑"
>
```

● RScript 실습

2. RScript 주요 함수

◆ help()

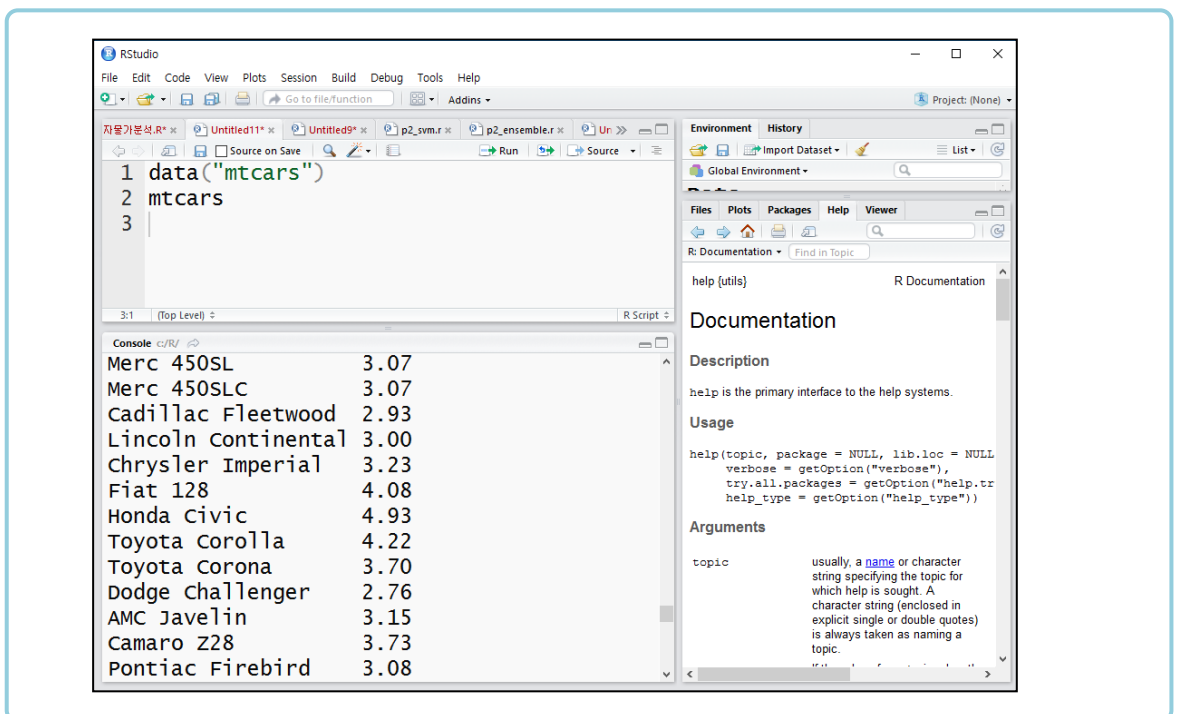
- 잘 모르는 함수에 대해 검색 또는 도움말을 요청할 수 있음



◆ data()

- 데이터셋을 가져오기 위해서 `data()` 함수를 호출

예) `data(mtcars)` → 내장되어 있는 `mtcars`라는 데이터를 가져옴



● RScript 실습

2. RScript 주요 함수

◆ head()

- 앞 부분의 일부 데이터만 불러 올 수 있음
 - 기본으로 6개의 데이터를 불러낼 수 있음
 - Head함수에서 개수를 설정할 수 있음

```
> head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0
Valiant	18.1	6	225	105	2.76	3.460	20.22	1

	am	gear	carb
Mazda RX4	1	4	4
Mazda RX4 Wag	1	4	4
Datsun 710	1	4	1
Hornet 4 Drive	0	3	1
Hornet Sportabout	0	3	2
Valiant	0	3	1

```
> head(mtcars,5)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0

	am	gear	carb
Mazda RX4	1	4	4
Mazda RX4 Wag	1	4	4
Datsun 710	1	4	1
Hornet 4 Drive	0	3	1
Hornet Sportabout	0	3	2

● RScript 실습

2. RScript 주요 함수

◆ summary()

- 데이터에 대한 요약 정보를 나타냄

```
> summary(mtcars)
```

mpg	cyl	disp
Min. :10.40	Min. :4.000	Min. : 71.1
1st Qu.:15.43	1st Qu.:4.000	1st Qu.:120.8
Median :19.20	Median :6.000	Median :196.3
Mean :20.09	Mean :6.188	Mean :230.7
3rd Qu.:22.80	3rd Qu.:8.000	3rd Qu.:326.0
Max. :33.90	Max. :8.000	Max. :472.0

hp	drat	wt
Min. : 52.0	Min. :2.760	Min. :1.513
1st Qu.: 96.5	1st Qu.:3.080	1st Qu.:2.581
Median :123.0	Median :3.695	Median :3.325
Mean :146.7	Mean :3.597	Mean :3.217
3rd Qu.:180.0	3rd Qu.:3.920	3rd Qu.:3.610
Max. :335.0	Max. :4.930	Max. :5.424

qsec	vs	am
Min. :14.50	Min. :0.0000	Min. :0.0000
1st Qu.:16.89	1st Qu.:0.0000	1st Qu.:0.0000
Median :17.71	Median :0.0000	Median :0.0000
Mean :17.85	Mean :0.4375	Mean :0.4062
3rd Qu.:18.90	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :22.90	Max. :1.0000	Max. :1.0000

gear	carb
Min. :3.000	Min. :1.000
1st Qu.:3.000	1st Qu.:2.000
Median :4.000	Median :2.000
Mean :3.688	Mean :2.812
3rd Qu.:4.000	3rd Qu.:4.000
Max. :5.000	Max. :8.000

Min.	최소값
Max.	최대값
Mean	산술평균
Median	중앙값
1st Qu. (Lower Quartile)	하한 사분위수
3rd Qu. (Upper Quartile)	상한 사분위수

1. 빅데이터 정의 및 프로그램 종류

■ 빅데이터란?

- 데이터를 구성하고 있는 물리적 하드웨어로부터 시작해 이를 기반으로 하는 애플리케이션과 소프트웨어를 포괄하는 거대 플랫폼

■ 빅데이터 플랫폼

- 거대한 데이터를 관리하기 위한 하드웨어 및 소프트웨어 등 모든 프로세스

■ 빅데이터 분석 프로그램 종류

- R 분석 프로그램
- 하둡(Hadoop) 프로그램
- 파이썬(Python) 프로그램

2. RStudio UI 분석

■ R 프로그램 환경설정

- R 프로그램 : 오픈 소스로 만들어진 통계 분석 프로그램
- Java 프로그램 : R에서 데이터분석을 할 때 자바 관련 라이브러리를 활용할 수 있게 해주는 프로그램
- Rstudio : 스튜디오는 R 프로그램 환경에서 스크립트 언어를 반복 사용할 수 있는 스크립트창, 히스토리창, 파일창, 환경설정창, 차트창등을 제공하여 더욱 편리한 환경을 제공

■ RStudio 메인 화면

- Script 창 : 기본 명령어를 사용하여 RScript를 실행 또는 저장할 수 있음
- Console 창 : RScript를 Run할 때 한 라인씩 실행되는 화면이 보이는 곳
- Environment/History 창 : 데이터셋의 이름, 관측치 개수, 변수 개수가 나오고 RStudio 사용 History를 볼 수 있는 곳
- Files / Plot / Packages / Help / Viewer 창 : Windows의 탐색기처럼 활용

3. RStudio UI 분석

■ RScript 언어 특징

- 인터프리터 형태의 언어
- ‘>’는 명령 프롬프트이고 명령을 실행하려면 명령 입력 후 Enter 누름
- 여러 줄에 명령 입력 시 다음 줄의 가장 왼쪽에 ‘+’기호가 자동으로 생성 됨
- 대소문자 구분
- ‘#’은 주석기호

■ RScript 주요 함수

- `help()` : 잘 모르는 함수에 대해 검색 또는 도움말을 요청할 수 있음
- `data()` : 데이터셋을 가져오기 위해서 `data()` 함수를 호출
- `head()` : 앞 부분의 일부 데이터만 불러 올 수 있음
- `summary()` : 데이터에 대한 요약 정보를 나타냄