

02 R 개요, 변수와 자료형

학습목표

1. RStudio의 환경을 설정하고, 기초 사용법을 설명할 수 있다.
2. 변수와 자료형을 이해한다.
3. R 패키지를 설치할 수 있다.

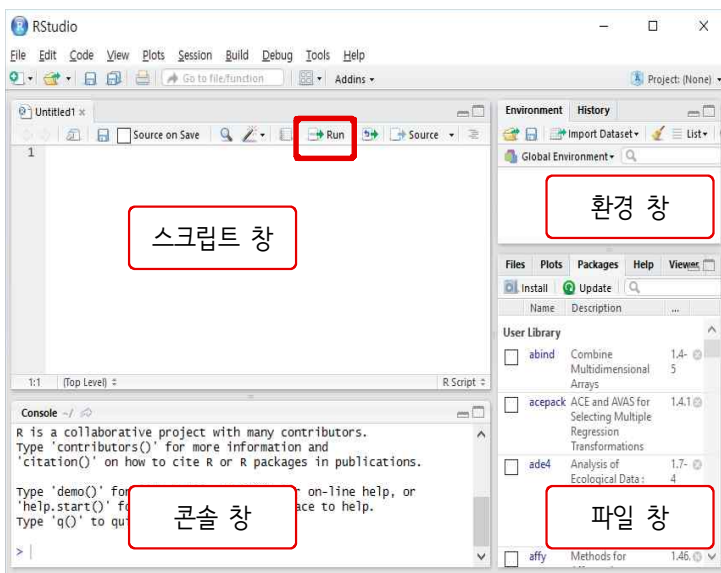
강의내용

데이터 ? - 현재는 데이터의 시대(the age of data) : 빅데이터
- 데이터로부터 유용한 정보와 통찰을 끄집어내고 합리적인 의사결정을 할 수 있다.
. 데이터는 새로운 석유다. (데이터과학자, 클라이브 험비)
. 데이터는 비즈니스를 위한 새로운 원천 재료가 되어 가고 있다(MS 부회장, 크레이그 먼디)

R이란? - 데이터의 처리, 통계 계산 및 분석, 그래픽스를 위한 프로그래밍 언어
- 인터프리터/스크립트 언어의 일종으로 무료로 사용가능한 오픈소스
- R은 기본(코어)패키지와 1만개 이상의 추가 패키지들로 구성

RStudio 사용자 친화적인 R 통합개발환경(IDE: R Integrated Development Environment)
환경설정 : [Tools]-[Global Options]
- General, Code, Appearance, Pane Layout 등 설정
- Default text encoding : UTF-8로 설정 [Code]-[Saving]탭
- Soft-warp 체크 : 자동 줄 바꿈 [Code]-[Editing]탭

R 기초 사용법 (1) R의 명령어는 대소문자 구별, 콘솔창에서 실행
(2) Script : R 명령어들로 만들어진 문서 (명령어들을 재사용할 수 있다.)
Script는 소스 창에서 작성하고 Ctrl + Enter로 실행
저장은 Ctrl + S 사용, #을 넣어 주석문 삽입
Script 새로 열기는 Ctrl + Shift + N
(3) Project : 작업의 단위로 Script나 데이터 파일이 하나의 그룹으로 저장됨
ls() 작업 중인 데이터 목록 확인
(4) Working Directory : Project의 기본 작업 디렉토리
getwd() 작업 디렉토리 확인, setwd('경로') 작업 디렉토리 변경



- 스크립트 창
R 명령문을 작성하고 실행하는 영역
- 콘솔 창
R 명령문의 실행 과정 및 결과 표시
- 환경 창
변수와 자료구조를 확인하는 영역
- 파일 창
작업폴더의 내용 확인, 그래프 표시,
패키지 설치 및 조회, 도움말 표시 등

02 R 개요, 변수와 자료형

함수

특정 기능을 수행하고 그 결과를 돌려주는 기능을 수행, `sum()`, `mean()`, `max()`, `min()`, `sqrt()` 등

변수

어떤 값을 임시로 보관해 두기 위한 저장소(저장된 값이 바뀔 수 있다는 의미에서 변수라고 함)

다양한 값을 가지고 있는 하나의 속성으로, 변수는 데이터 분석이 대상이 된다.

변수명 생성 규칙 : 문자, 숫자, 언더바(`_`), 마침표(`.`)로 구성, 하이픈(`-`) 사용 불가

문자로 시작, 대소문자 구분

```
var <- 15          v.a <- 'r'          strd <- "dogs are"
```

(`cf`) 상수 : 고정 값을 의미하며, 상수는 데이터 분석 대상이 아님

R의 자료형

기본 자료형 숫자형(**numeric**), 문자형(**character**), 논리형(**logical**), 복소수형(**complex**)

특수한 형태 **NA** 결측치(Not Available의 약자), 데이터 값이 없음을 뜻함

값이 빠져있는 경우(실수로 값을 입력하지 않았거나, 값이 어떤 이유로 관찰되지 못한 경우)

NULL 변수가 초기화 되지 않았을 때 사용, 미정(undefined) 값을 표현하는데 사용

<pre># 계산 기능 3+4*5 (1+2)+4*5 10/2 2**10 # 함수 사용 sum(1,3,5) mean(1:10) max(100, 200) sqrt(81) # 변수 생성, 변수에 값 할당 v1 <- 5 px <- 4.15 v_3 <- 1:10 v.4 <- TRUE Hallym <- "한림대학교" hallym <- "한림" v2 <- v1*30 v2 <- v2-100 sv3 <- sum(v_3)</pre>	<pre># 기본 자료형 class(10) # 숫자형(numeric) class(TRUE) # 논리형(logical) class("한림") # 문자형(character) class(1+2i) # 복소수형(complex) class(10L) # integer class(v1) # numeric class(v.4) # logical class(Hallym) # character # 특수형태의 자료형 exp(2**10) # Inf : 양의 무한대 -5/0 # -Inf : 음의 무한대 0/0 # NaN(Not a Number) : 숫자가 아님 v1[10] # NA(Not Available) 결측치 : 값이 없음 names(v2) # NULL : 값 미정</pre>
---	--

R 패키지(Package)

R 패키지는 함수, 데이터, 코드, 문서의 집합으로, R의 기능을 확장시켜 사용의 편리성을 증대시킴

R의 기본 패키지(stats, graphics, utils, datasets, base 등)를 통해 통계와 그래프 등 기본 기능 제공

CRAN 사이트(<https://cran.r-project.org/web/packages>)에 현재 14,763건 등록

tidyverse 패키지(<https://www.tidyverse.org/packages/>)

```
install.packages("ggplot2")
library(ggplot2)
#help(package="ggplot2")

val <- c("a", "a", "b", "a", "a", "c")
qplot(val)
```