



# 비즈니스프로그래밍

- Python과 R 기초 -

# 외부 파일 가져오기

- ▶ 데이터분석을 위해 외부에 있는 데이터를 불러와 사용하는 경우가 많다.
- ▶ xlsx파일은 엑셀로 만들어진 파일, csv파일은 쉼표로 구분된 데이터 파일
- ▶ csv파일을 불러오기 위해 `read.csv()` 함수 이용
- ▶ 파일을 불러올 때는 파일의 경로, 파일명, 파일확장자가 오타없이 정확하게 작성되었는지 확인한다.
- ▶ 엑셀파일은 `readxl` 패키지 설치 필요 ➔ 자세한 설명은 패키지 소개에서
- ▶ `readxl` 패키지에서 제공하는 `read_excel()` 함수 이용



# 3. 기초함수와 패키지



## ▶ 3.1 함수의 정의 및 기초함수

- ▶ 함수의 정의
- ▶ 수학 함수
- ▶ 기초통계량 함수
- ▶ 문자열 처리 함수
- ▶ 함수 도움말 확인
- ▶ 사용자 정의 함수



# 함수의 정의



- ▶ 함수는 반복적으로 자주 사용되는 코드를 묶어놓은 것
- ▶ 함수를 사용할 때 입력값을 전달 ➔ 인수(Arguments)
- ▶ 함수는 인수를 전달받아 처리한 후 출력값을 반환



# 수학 함수



함수	의미
abs(x)	절대값
sqrt(x)	제곱근
trunc(x)	소수점 자리 버림
round(x, digits = n)	지정된 위치에서 반올림



# 기초통계량 함수



- ▶ 기초통계: 데이터의 특성을 간단하게 서술하는 통계
- ▶ 데이터의 경향을 파악할 수 있다.

함수	의미	설명
length( )	자료의 개수	자료의 행의 개수
mean( )	평균	자료를 모두 더한 후 개수로 나눈 값
median( )	중앙값	자료를 순서대로 정렬했을 때 가운데 있는 값
max( )	최대값	자료를 순서대로 정렬했을 때 가장 큰 값
min( )	최소값	자료를 순서대로 정렬했을 때 가장 작은 값
range( )	최소값과 최대값	최대값에서 최소값의 범위
sd( )	표준편차	데이터 값이 퍼진 정도를 설명하는 통계량
summary( )	다섯수치요약(최소, 1사분위값, 중앙값, 3사분위값, 최대값)	1사분위값: 자료를 순서대로 정렬했을 때 25% 위치의 값 3사분위값: 자료를 순서대로 정렬했을 때 75% 위치의 값



# 문자열 처리 함수



함수	의미
grep(패턴, x)	패턴문자열을 포함하는 x의 원소 번호 반환
gsub(패턴, replace, x)	패턴을 가지는 문자열을 x에서 찾아 replace로 대체
substr(x, 시작, 끝)	시작과 끝을 지정하여 x에서 문자열 일부 추출
paste(., sep = "")	문자열 합치기 sep 옵션은 합칠 때 넣을 문자



# 함수 도움말 확인



- ▶ 함수에 대한 설명이 필요한 경우 ➔ ?함수명
- ▶ Help창에서 확인
- ▶ 도움말은 함수의 설명, 사용법, 인수, 예제 등으로 구성





# 사용자 정의 함수



- ▶ 함수는 이미 만들어진 함수를 호출하여 사용할 수도 있고, 필요에 따라 직접 만들어서 사용할 수도 있다.
- ▶ 함수를 만드는 방법

```
함수명 <- function(입력, 입력)
{
    특정 기능을 수행할 문장
    return(반환값)
}
```



# 3. 기초함수와 패키지



- ▶ 3.2 제어문
  - ▶ if~else문
  - ▶ ifelse( ) 함수
  - ▶ 논리식을 활용한 조건검사



# if~else문



- ▶ 조건식을 판단하는 if~else 구문
- ▶ 조건문은 조건식을 기준으로 조건이 참일 때와 거짓일 때 처리방식을 다르게 할 때 사용

```
if(조건식)
{
    참일 때 실행되는 문장
} else
{
    거짓일 때 실행되는 문장
}
```



# ifelse( ) 함수



- ▶ 조건식이 간단한 경우 ifelse( ) 함수 사용하여 표기할 수 있다.

ifelse(조건, 조건이 참일 때 실행되는 문장, 조건이 거짓일 때 실행되는 문장)

# 논리식을 활용한 조건검사

- ▶ 데이터 프레임은 `ifelse()` 함수나 `if~else` 조건문을 사용하지 않고 논리식을 이용해 조건에 맞는 데이터를 추출할 수 있다.
- ▶ 조건을 만족하는 데이터를 추출하기 위해 `데이터프레임명[조건식, ]` 으로 지정하면 조건에 맞는 행을 추출할 수 있다.



# 3. 기초함수와 패키지



## ▶ 3.2 패키지 소개

- ▶ 내장함수 이외에도 다양한 함수들을 제공하고 있어서 더욱 쉽고 간단히 통계 분석을 할 수 있다. 하지만 내장함수 이외의 함수를 사용하기 위해서는 필요한 패키지를 먼저 설치하고 불러와야 한다.
- ▶ 패키지(package)란 특수한 목적으로 만들어 놓은 함수들의 모음
- ▶ 대표적인 패키지

패키지	설명
ggplot2	데이터를 시각화
dplyr	데이터 전처리, 조작



# 3. 기초함수와 패키지



## ▶ 3.2 패키지 소개

<code>install.packages("패키지명")</code>	<code>library(패키지명)</code>
패키지를 설치하기	설치된 패키지를 가져오기

- ▶ 패키지는 한 번만 설치하면 되지만 라이브러리는 RStudio를 새로 열 때마다 다시 실행시켜야 한다.
- ▶ 패키지에 대한 도움말을 확인하고 싶다면 cheat sheet가 있는지 확인
- ▶ cheat sheet는 간단한 요약본 형태로 설명되어 있어 한눈에 함수들의 종류나 사용법 등을 확인할 수 있다.



# 3. 기초함수와 패키지



## ▶ 3.2 패키지 소개 - ggplot2로 그래프 그리기

### ▶ 1. 데이터 연결하기

```
ggplot(data, aes(x = x축데이터, y = y축데이터))
```

### ▶ 2. 그래프 모양 지정하기

```
geom_boxplot( )
```

### ▶ 3. + 기호로 그래프 연결하기