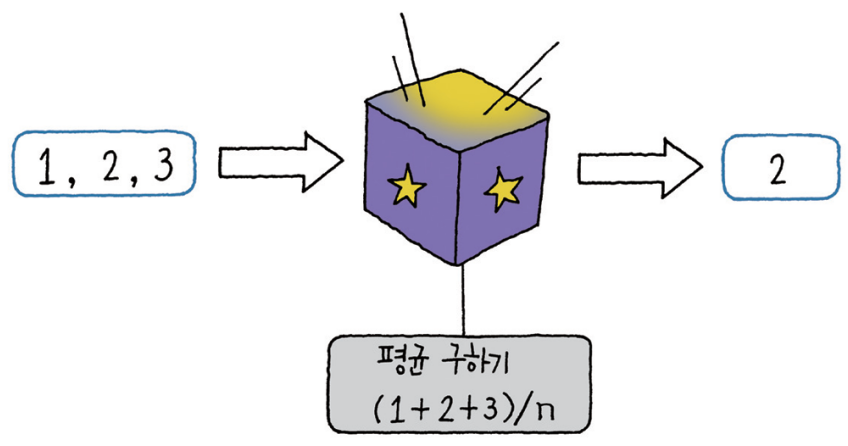
# 3. 데이터 분석을 위한 연장 챙기기

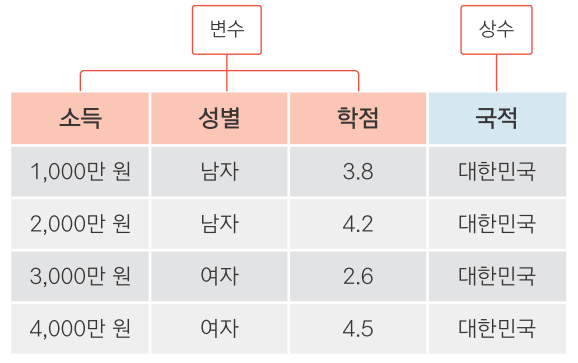


##### NP

## 03-1. 변하는 수, '변수' 이해하기

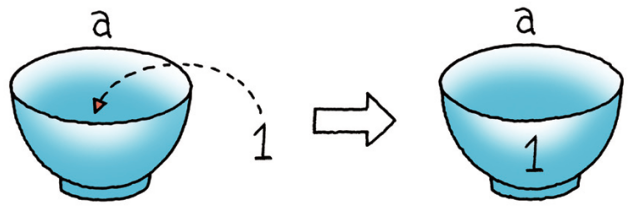
**변수(Variable)**

* 다양한 값을 지니고 있는 하나의 속성
* 변수는 데이터 분석의 대상
* 상수는 분석할 게 없다 (국적에 따른 소득차, 국적에 따는 성비)



##### NP

### 변수 만들기



a <- 1  
a

## [1] 1

b <- 2  
b

## [1] 2

c <- 3  
c

## [1] 3

d <- 3.5  
d

## [1] 3.5

##### NP

#### 변수로 연산하기

a+b

## [1] 3

a+b+c

## [1] 6

4/b

## [1] 2

5\*b

## [1] 10

##### NP

### 여러 값으로 구성된 변수 만들기

#### c()

var1 <- c(1, 2, 5, 7, 8) # 숫자 다섯 개로 구성된 var1 생성  
var1

## [1] 1 2 5 7 8

var2 <- c(1:5) # 1~5까지 연속값으로 var2 생성  
var2

## [1] 1 2 3 4 5

##### NP

### seq()

var3 <- seq(1, 5) # 1~5까지 연속값으로 var3 생성  
var3

## [1] 1 2 3 4 5

var4 <- seq(1, 10, by = 2) # 1~10까지 2 간격 연속값으로 var4 생성  
var4

## [1] 1 3 5 7 9

var5 <- seq(1, 10, by = 3) # 1~10까지 3 간격 연속값으로 var5 생성  
var5

## [1] 1 4 7 10

##### NP

#### 연속값 변수로 연산하기

var1

## [1] 1 2 5 7 8

var1+2

## [1] 3 4 7 9 10

var1

## [1] 1 2 5 7 8

var2

## [1] 1 2 3 4 5

var1+var2

## [1] 2 4 8 11 13

##### NP

#### 문자로 된 변수 만들기

str1 <- "a"  
str1

## [1] "a"

str2 <- "text"  
str2

## [1] "text"

str3 <- "Hello World!"  
str3

## [1] "Hello World!"

##### NP

#### 연속 문자 변수 만들기

str4 <- c("a", "b", "c")  
str4

## [1] "a" "b" "c"

str5 <- c("Hello!", "World", "is", "good!")  
str5

## [1] "Hello!" "World" "is" "good!"

##### NP

#### 문자로 된 변수로는 연산할 수 없다

str1+2

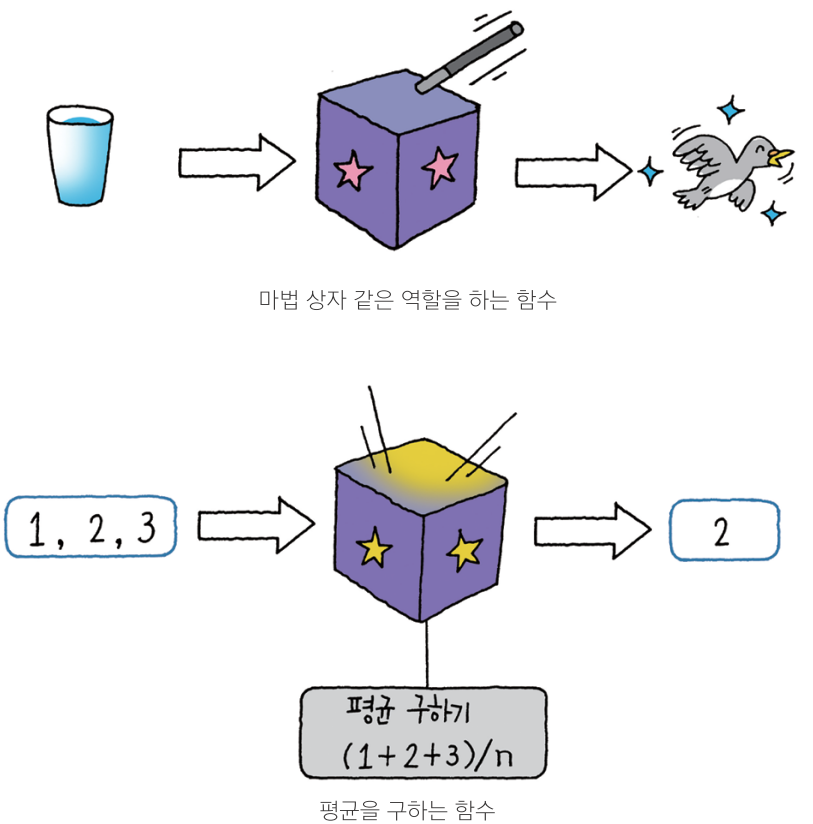
## Error in str1 + 2: non-numeric argument to binary operator

##### NP

## 03-2. 마술 상자 같은 '함수' 이해하기

**함수**

* 값을 넣으면 특정한 기능을 수행해 처음과 다른 값이 출력됨



### 숫자를 다루는 함수 이용하기

# 변수 만들기  
x <- c(1, 2, 3)  
x

## [1] 1 2 3

# 함수 적용하기  
mean(x)

## [1] 2

max(x)

## [1] 3

min(x)

## [1] 1

##### NP

### 문자를 다루는 함수 이용하기

str5

## [1] "Hello!" "World" "is" "good!"

paste(str5, collapse = ",") # 쉼표를 구분자로 str4의 단어들 하나로 합치기

## [1] "Hello!,World,is,good!"

##### NP

#### 함수의 옵션 설정하기 - 파라미터

paste(str5, collapse = " ")

## [1] "Hello! World is good!"

#### 함수의 결과물로 새 변수 만들기

x\_mean <- mean(x)  
x\_mean

## [1] 2

str5\_paste <- paste(str5, collapse = " ")  
str5\_paste

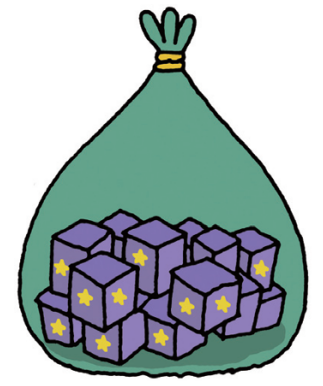
## [1] "Hello! World is good!"

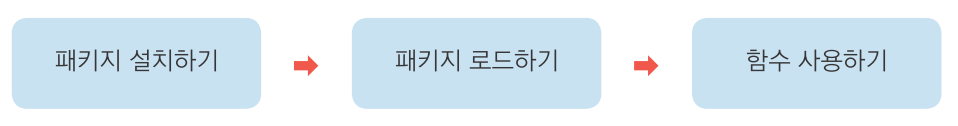
##### NP

## 03-3. 함수 꾸러미, '패키지' 이해하기

**패키지(packages)**

* 함수가 여러 개 들어 있는 꾸러미
* 하나의 패키지 안에 다양한 함수가 들어있음
* 함수를 사용하려면 패키지 설치 먼저 해야함





### ggplot2 패키지 설치하기, 로드하기

install.packages("ggplot2") # ggplot2 패키지 설치  
library(ggplot2) # ggplot2 패키지 로드

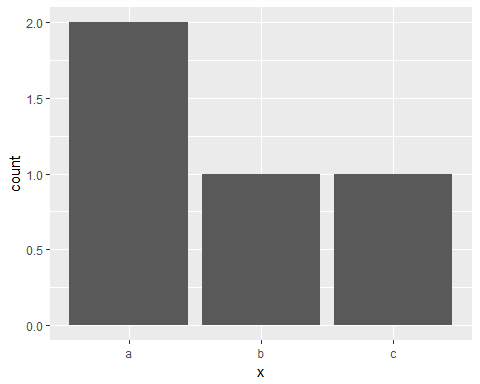
##### NP

#### 함수 사용하기

# 여러 문자로 구성된 변수 생성  
x <- c("a", "a", "b", "c")  
x

## [1] "a" "a" "b" "c"

# 빈도 그래프 출력  
qplot(x)



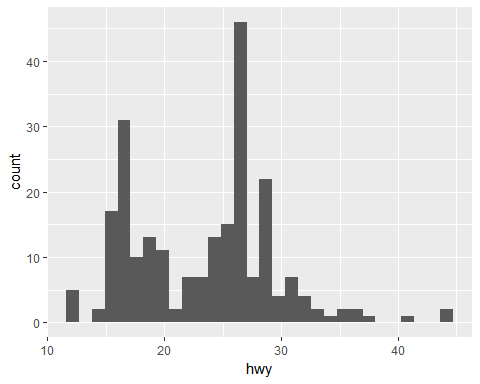
##### NP

#### ggplot2의 mpg 데이터로 그래프 만들기 : 파라미터에 따라 그래프가 변하는 것만 확인

mpg(Mile Per Gallon) : 미국 환경 보호국에서 공개한 자료, ggplot2 페키지에 예제 데이터로 들어 있음

* 1999~2008년 사이 미국에서 출시된 자동차 234종의 연비 정보 제공
* hwy : 자동차가 고속도로에서 1갤런에 몇 마일을 가는 지 나타낸 변수

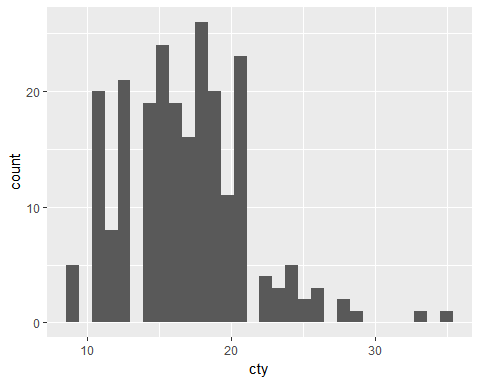
# data에 mpg, x축에 hwy 변수 지정하여 그래프 생성 : 고속도로 연비별 빈도 막대 그래프  
qplot(data = mpg, x = hwy)



##### NP

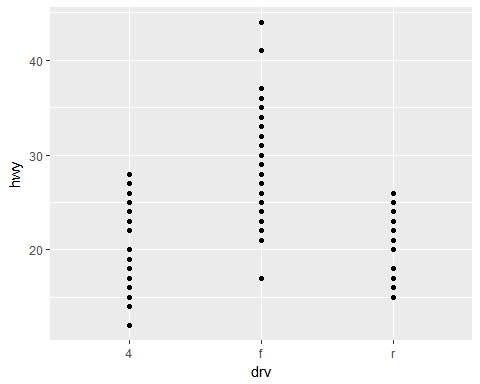
#### qplot() 파라미터 바꿔보기

# x축 cty  
qplot(data = mpg, x = cty)



##### NP

# x축 drv, y축 hwy  
qplot(data = mpg, x = drv, y = hwy)



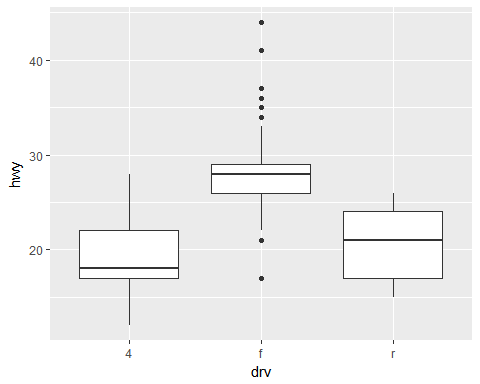
##### NP

# x축 drv, y축 hwy, 선 그래프 형태  
qplot(data = mpg, x = drv, y = hwy, geom = "line")



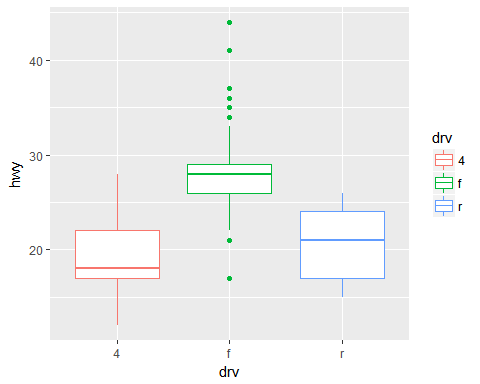
##### NP

# x축 drv, y축 hwy, 상자 그림 형태  
qplot(data = mpg, x = drv, y = hwy, geom = "boxplot")



##### NP

# x축 drv, y축 hwy, 상자 그림 형태, drv별 색 표현  
qplot(data = mpg, x = drv, y = hwy, geom = "boxplot", colour = drv)



##### NP

#### 함수의 기능이 궁금할 땐 Help 함수를 활용해 보세요

?qplot

##### NP

### 혼자서 해보기

#### Q1. 시험 점수 변수 만들고 출력하기

다섯 명의 학생이 시험을 봤습니다. 학생 다섯 명의 시험 점수를 담고 있는 변수를 만들어 출력해 보세요. 각 학생의 시험 점수는 다음과 같습니다.

80, 60, 70, 50, 90

#### Q2. 전체 평균 구하기

앞 문제에서 만든 변수를 이용해서 이 학생들의 전체 평균 점수를 구해보세요.

#### Q3. 전체 평균 변수 만들고 출력하기

전체 평균 점수를 담고 있는 새 변수를 만들어 출력해 보세요. 앞 문제를 풀 때 사용한 코드를 응용하면 됩니다.

##### NP

### 정답

#### Q1. 시험 점수 변수 만들고 출력하기

score <- c(80, 60, 70, 50, 90)  
score

## [1] 80 60 70 50 90

#### Q2. 전체 평균 구하기

mean(score)

## [1] 70

#### Q3. 전체 평균 변수 만들고 출력하기

mean\_score <- mean(score)  
mean\_score

## [1] 70