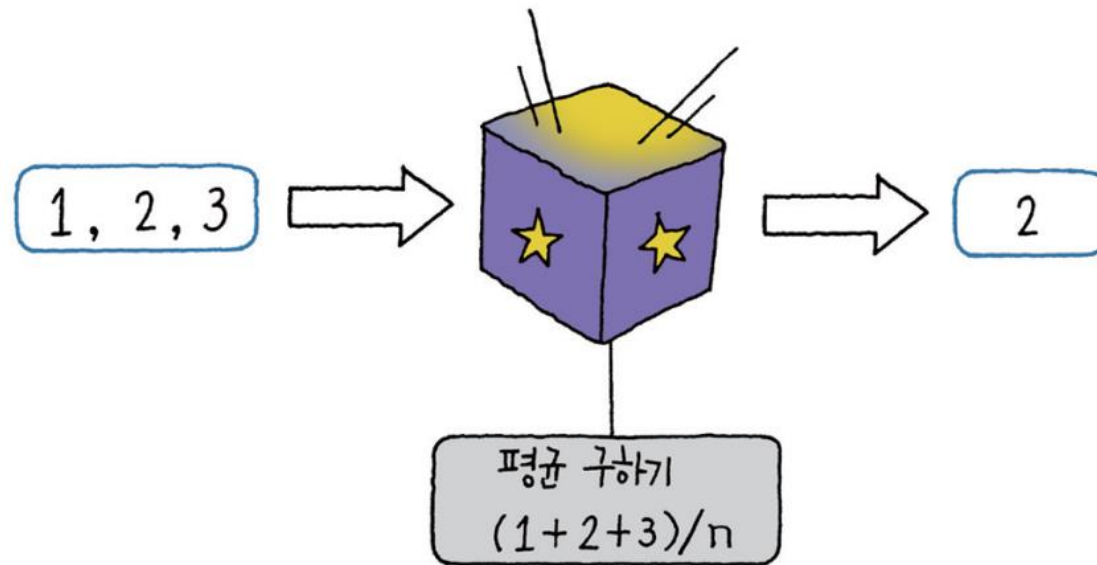


### 3. 데이터 분석을 위한 연장 챙기기



## 03-1. 변하는 수, '변수' 이해하기

### 변수(Variable)

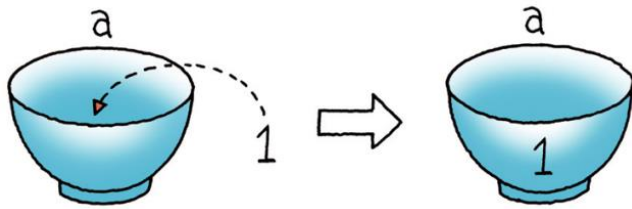
- 다양한 값을 지니고 있는 하나의 속성
- 변수는 데이터 분석의 대상



The diagram shows two boxes at the top: '변수' (Variable) on the left and '상수' (Constant) on the right. A red line connects the '변수' box to the first three columns of the table below (Income, Gender, GPA). Another red line connects the '상수' box to the fourth column of the table (Nationality).

소득	성별	학점	국적
1,000만 원	남자	3.8	대한민국
2,000만 원	남자	4.2	대한민국
3,000만 원	여자	2.6	대한민국
4,000만 원	여자	4.5	대한민국

## 변수 만들기



```
a <- 1
```

```
a
```

```
## [1] 1
```

```
b <- 2
```

```
b
```

```
## [1] 2
```

```
c <- 3
```

```
c
```

```
## [1] 3
```

```
d <- 3.5
```

```
d
```

```
## [1] 3.5
```

## 변수로 연산하기

a+b

## [1] 3

a+b+c

## [1] 6

4/b

## [1] 2

5\*b

## [1] 10

## 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)             # 1~5 까지 연속값으로 var2 생성
```

```
var2
```

```
## [1] 1 2 3 4 5
```

## seq()

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

```
var3
```

```
## [1] 1 2 3 4 5
```

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

```
var4
```

```
## [1] 1 3 5 7 9
```

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

```
var5
```

```
## [1] 1 4 7 10
```

## 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
```

## 문자로 된 변수 만들기

```
str1 <- "a"
```

```
str1
```

```
## [1] "a"
```

```
str2 <- "text"
```

```
str2
```

```
## [1] "text"
```

```
str3 <- "Hello World!"
```

```
str3
```

```
## [1] "Hello World!"
```



## 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!"
```

문자로 된 변수로는 연산할 수 없다

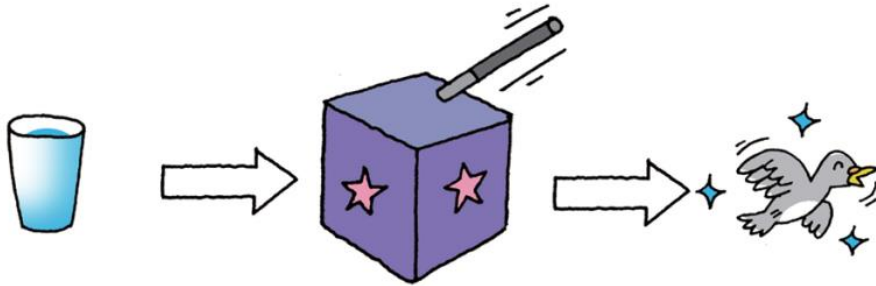
```
str1+2
```

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

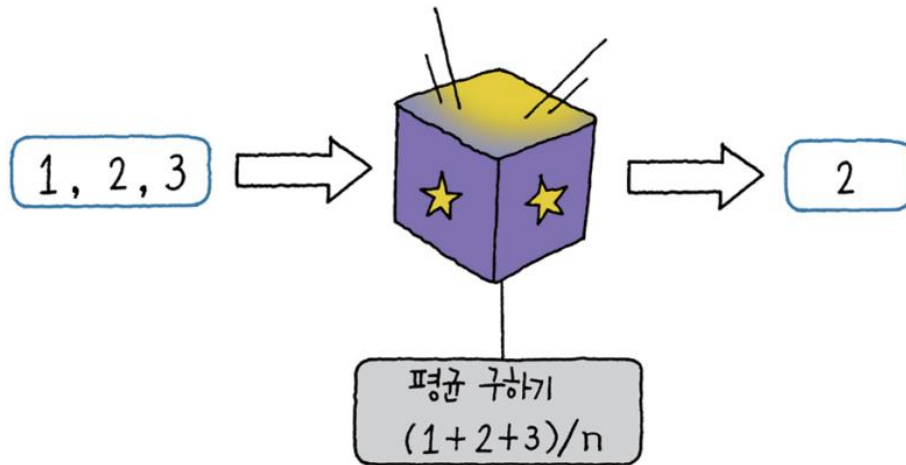
## 03-2. 마술 상자 같은 '함수' 이해하기

### 함수

- 값을 넣으면 특정한 기능을 수행해 처음과 다른 값이 출력됨



마법 상자 같은 역할을 하는 함수



평균을 구하는 함수

## 숫자를 다루는 함수 사용하기

*# 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
```

## 문자를 다루는 함수 사용하기

```
str5
```

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

```
paste(str5, collapse = ",") # 쉽표를 구분자로 str4 의 단어들 하나로 합치기
```

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

## 함수의 옵션 설정하기 - 파라미터

```
paste(str5, collapse = " ")  
## [1] "Hello! World is good!"
```

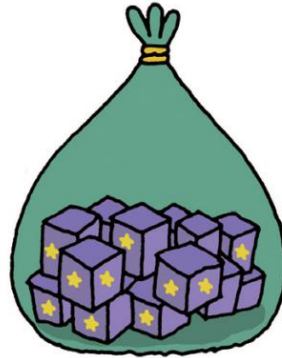
## 함수의 결과물로 새 변수 만들기

```
x_mean <- mean(x)  
x_mean  
## [1] 2  
  
str5_paste <- paste(str5, collapse = " ")  
str5_paste  
## [1] "Hello! World is good!"
```

## 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



패키지 설치하기



패키지 로드하기



함수 사용하기

## ggplot2 패키지 설치하기, 로드하기

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



## 함수 사용하기

*# 여러 문자로 구성된 변수 생성*

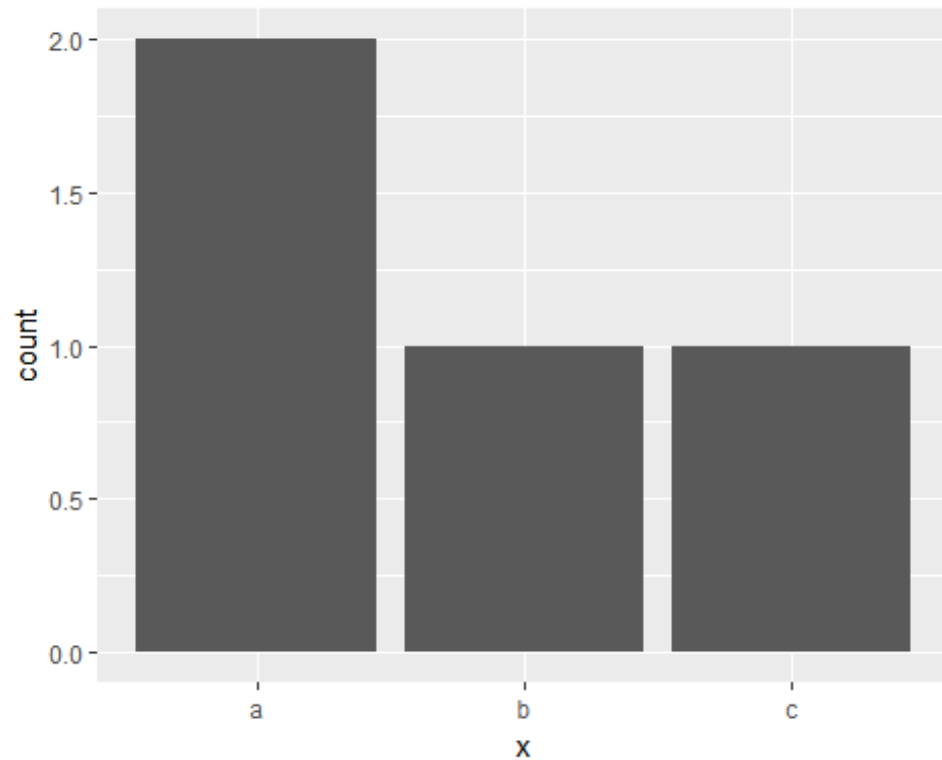
```
x <- c("a", "a", "b", "c")
```

```
x
```

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

*# 빈도 그래프 출력*

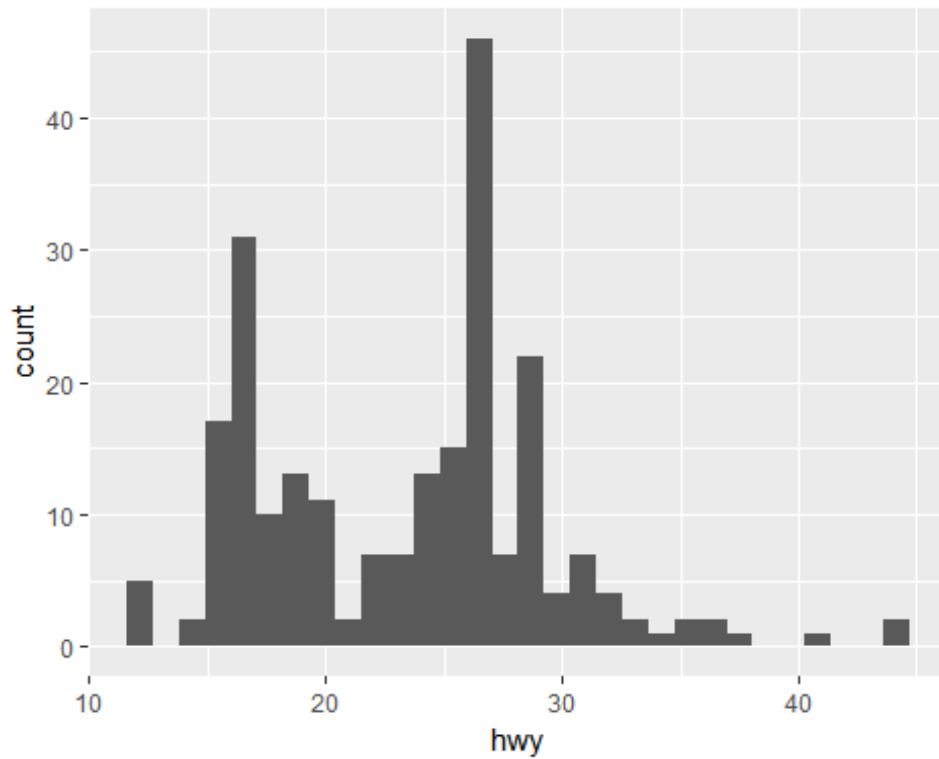
```
qplot(x)
```



## ggplot2의 mpg 데이터로 그래프 만들기

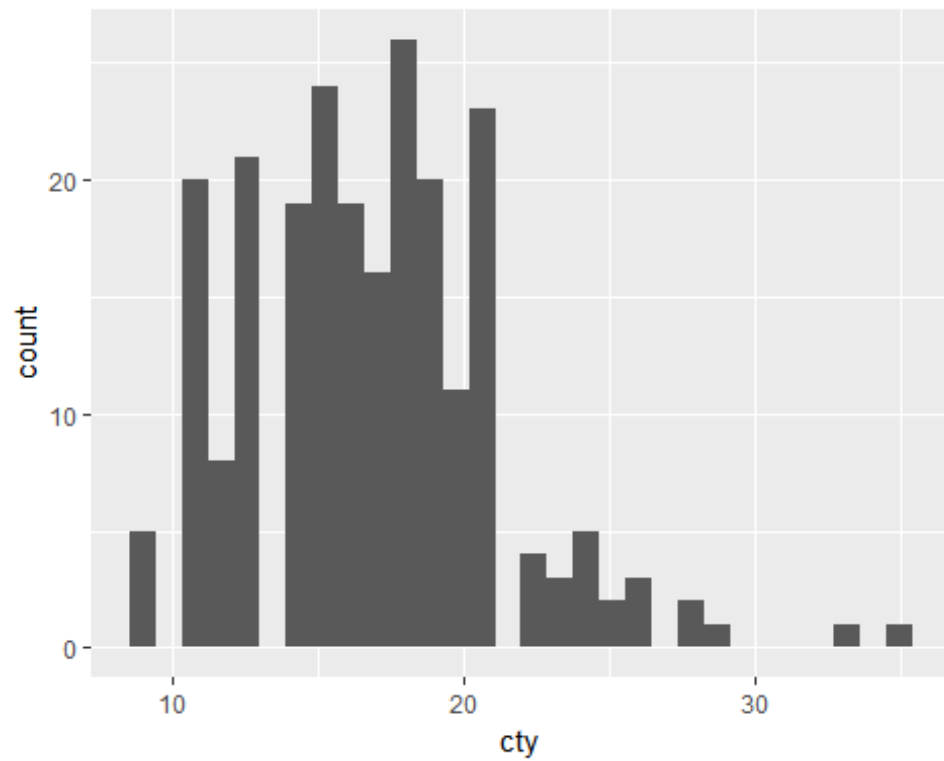
*# data 에 mpg, x 축에 hwy 변수 지정하여 그래프 생성*

```
qplot(data = mpg, x = hwy)
```



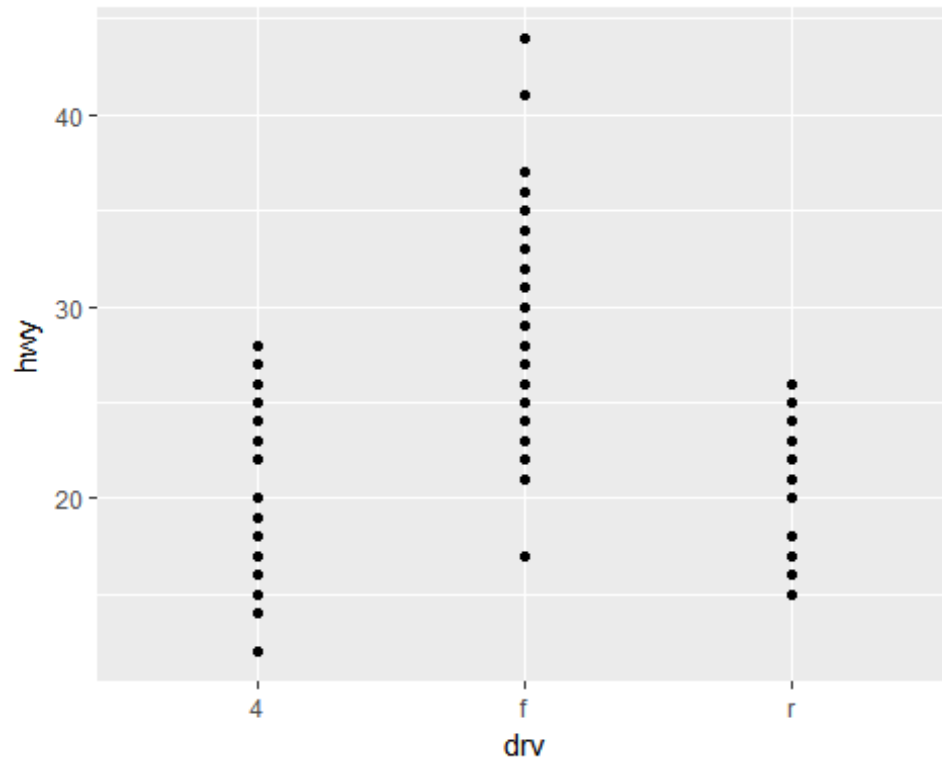
## qplot() 파라미터 바꿔보기

```
# x 축 cty  
qplot(data = mpg, x = cty)
```



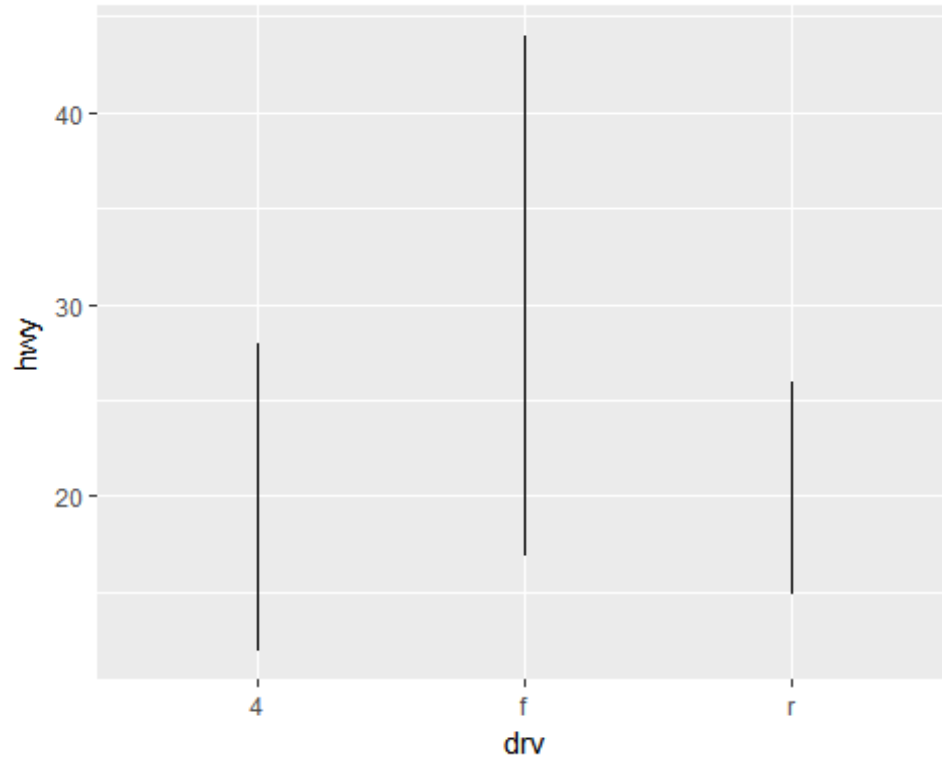
```
# x ≙ drv, y ≙ hwy
```

```
qplot(data = mpg, x = drv, y = hwy)
```



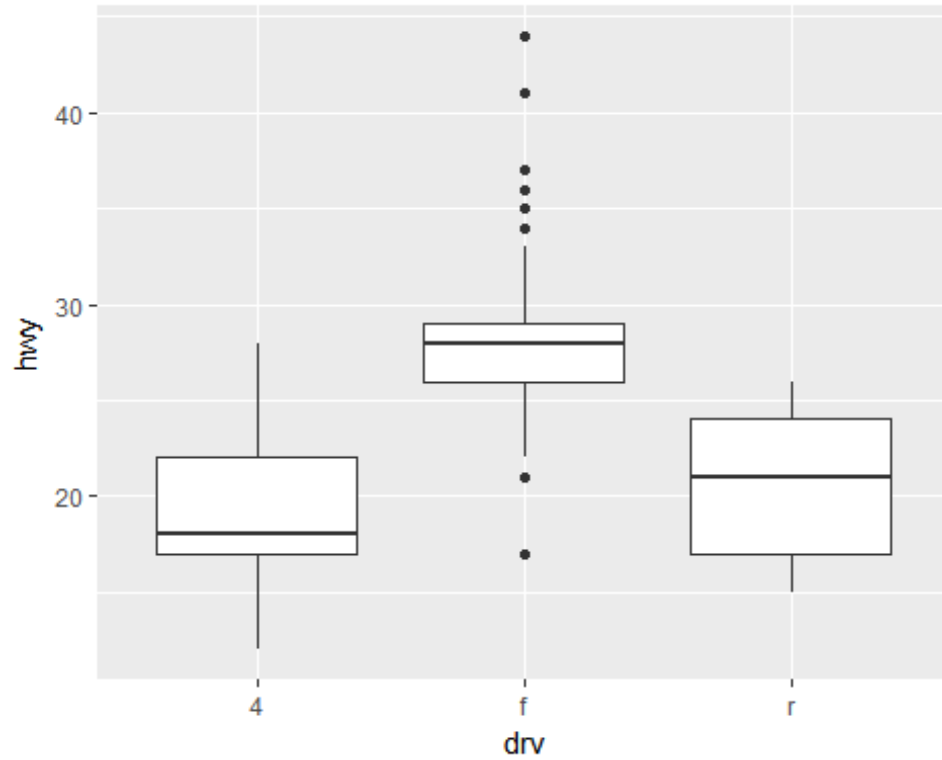
# x 축 drv, y 축 hwy, 선 그래프 형태

```
qplot(data = mpg, x = drv, y = hwy, geom = "line")
```



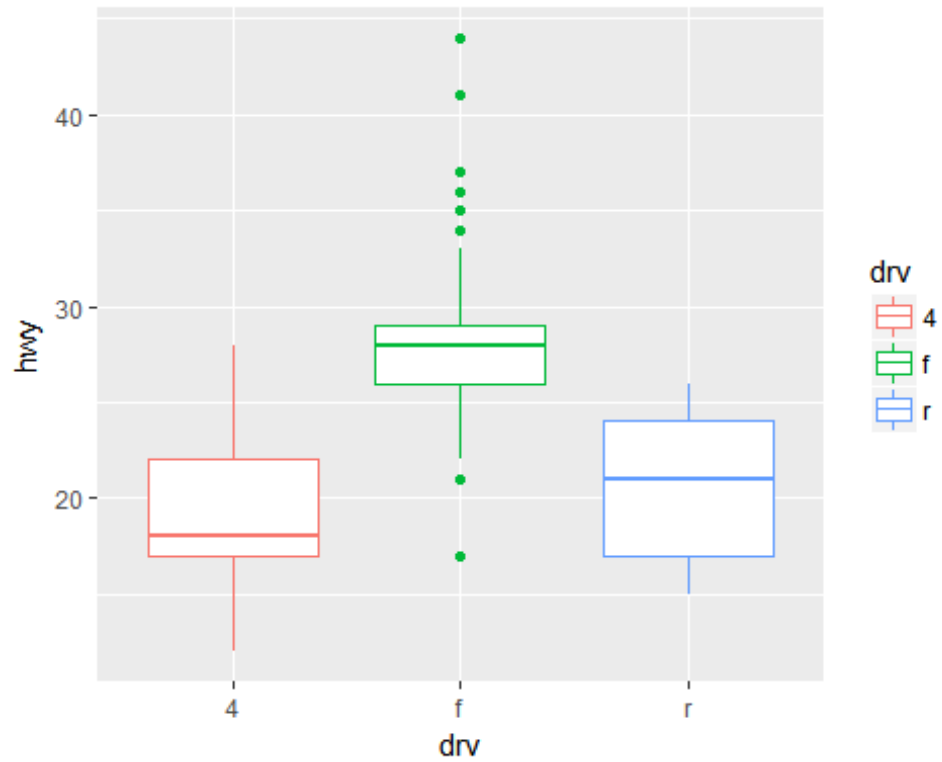
# x 축 drv, y 축 hwy, 상자 그림 형태

```
qplot(data = mpg, x = drv, y = hwy, geom = "boxplot")
```



# x 축 drv, y 축 hwy, 상자 그림 형태, drv 별 색 표현

```
qplot(data = mpg, x = drv, y = hwy, geom = "boxplot", colour = drv)
```



함수의 기능이 궁금할 땐 Help 함수를 활용해 보세요

```
?qplot
```



# 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.

## 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
```