

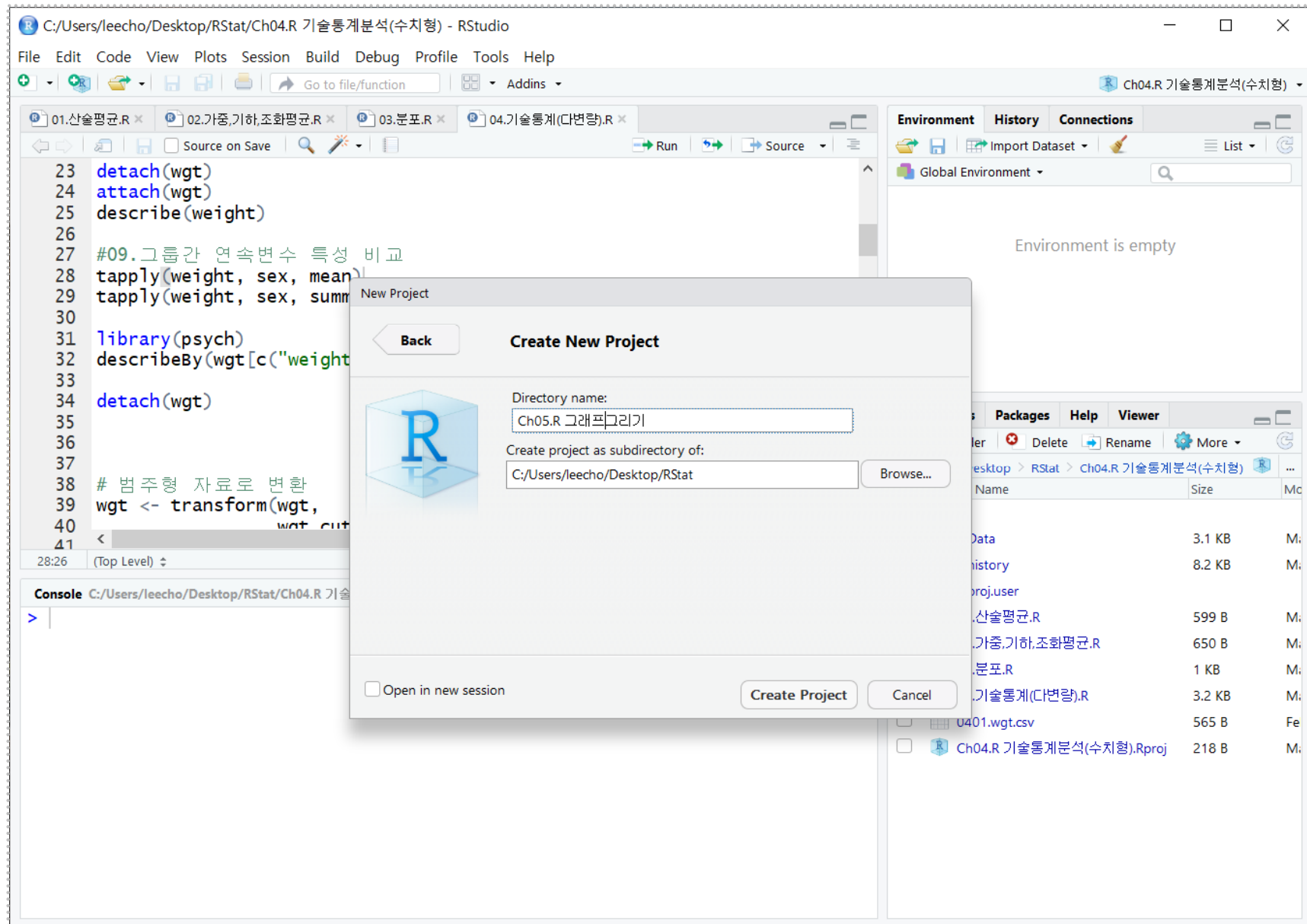
# Ch05. 그래프 그리기

# 학습목표

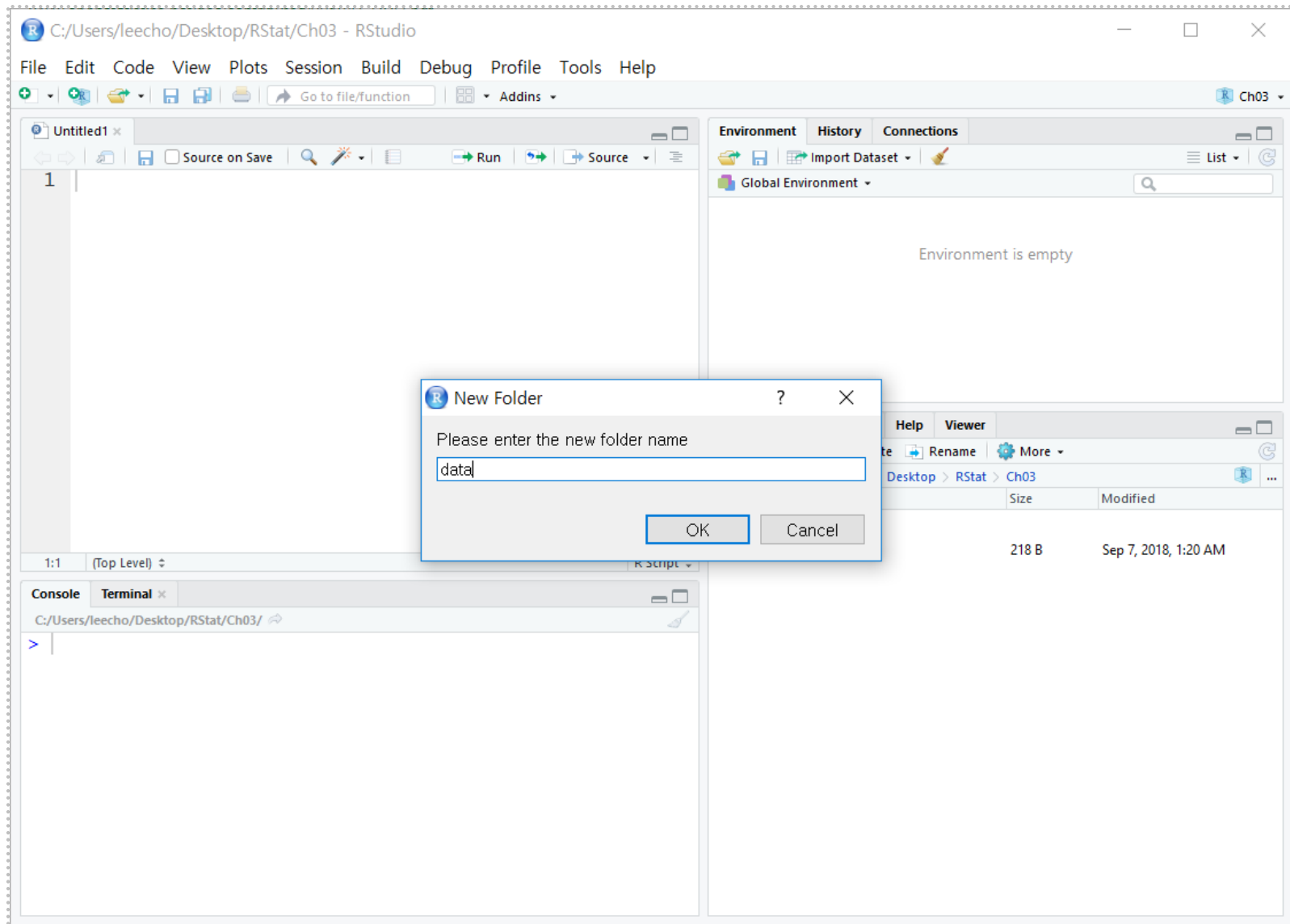
- 일변량 범주형 자료의 그래프 작성
- 다변량 범주형 자료의 그래프 작성
- 일변량 수치형 자료의 그래프 작성
- 다변량 수치형 자료의 그래프 작성

# 1. 범주형 자료(일변량)

# 프로젝트 생성



# 데이터 가져오기



# 그래프의 중요성

## ■ 그래프

- 인간이 지닌 시각적인 인지능력을 활용하여 직관적으로 그 현상을 쉽게 인식하도록 하는 방법
- 통계적인 데이터를 요약하여 시각적으로 그 특징을 나타내는 것

## ■ 그래프의 문제점

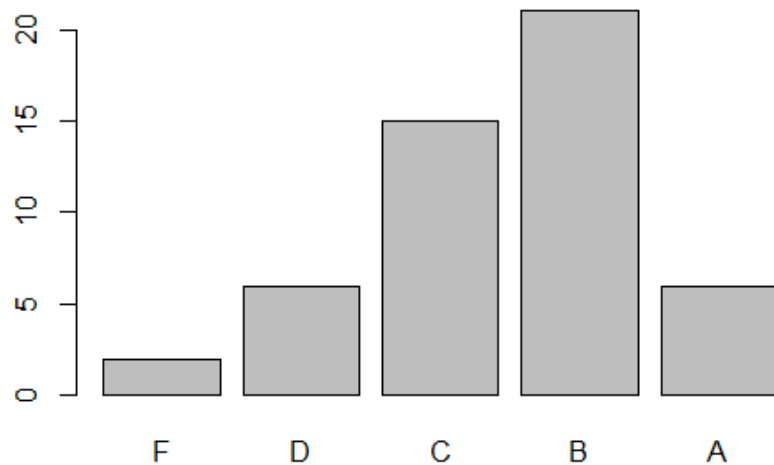
- 그래프는 자료가 가지고 있는 속성뿐만 아니라 강렬한 인상을 주게 되어 확대해석의 오류를 범할 위험이 있음

## ■ 활용

- 그래프를 통해 데이터의 특징 및 자료의 이상치를 점검하여 이후의 통계분석에 대비함

# 자료형태에 따른 그래프 그리기

- 막대그래프와 히스토그램
  - 범주형 : 막대그래프
  - 수치형 : 히스토그램



# 그래프 그리기

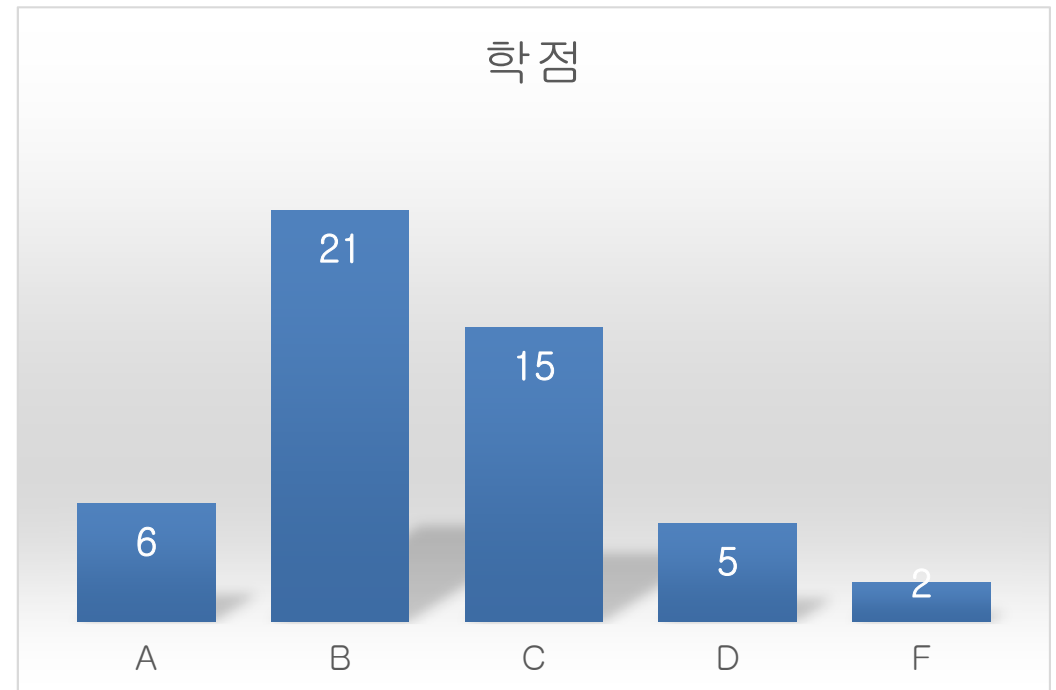
- # 범주형 변수가 1개 일 때
- barplot:
  - 막대도표,
  - 수평막대도표
  - 누적막대도표
- pie: 원도표



# 도수분포표

- 도수분포표와 막대그래프 (Bar chart)
  - 누적비율: 순서형일 경우에 사용하면 편리

학점	빈도	비율(%)	누적비율(%)
A	6	12	12
B	21	42	54
C	15	30	84
D	6	12	96
F	2	4	100
합계	50	100	100



# 막대그래프(barplot)

```
attach(freq)
```

##02. 일변량 범주형 자료

```
barplot(grade)
```

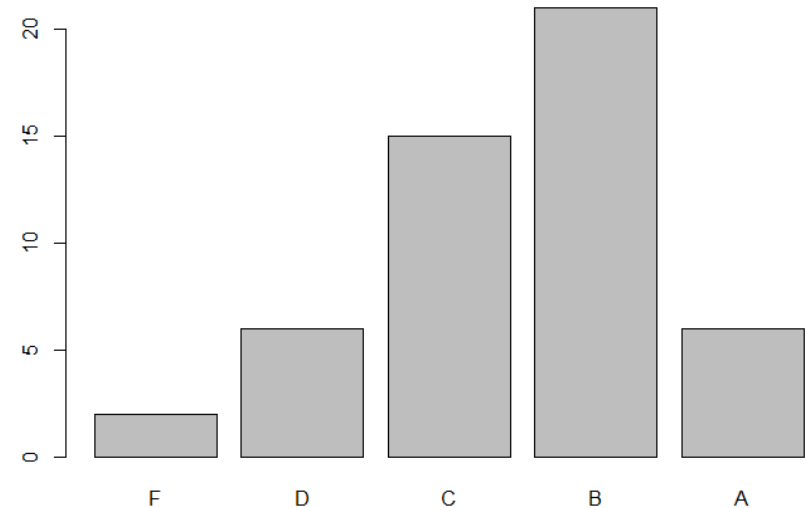
```
> barplot(grade)
Error in barplot.default(grade) :
  'height'는 반드시 벡터 또는 행렬이어야 합니다
```

#그래프 에러(table로 정리한 후에  
연결-원자료 사용 못함)

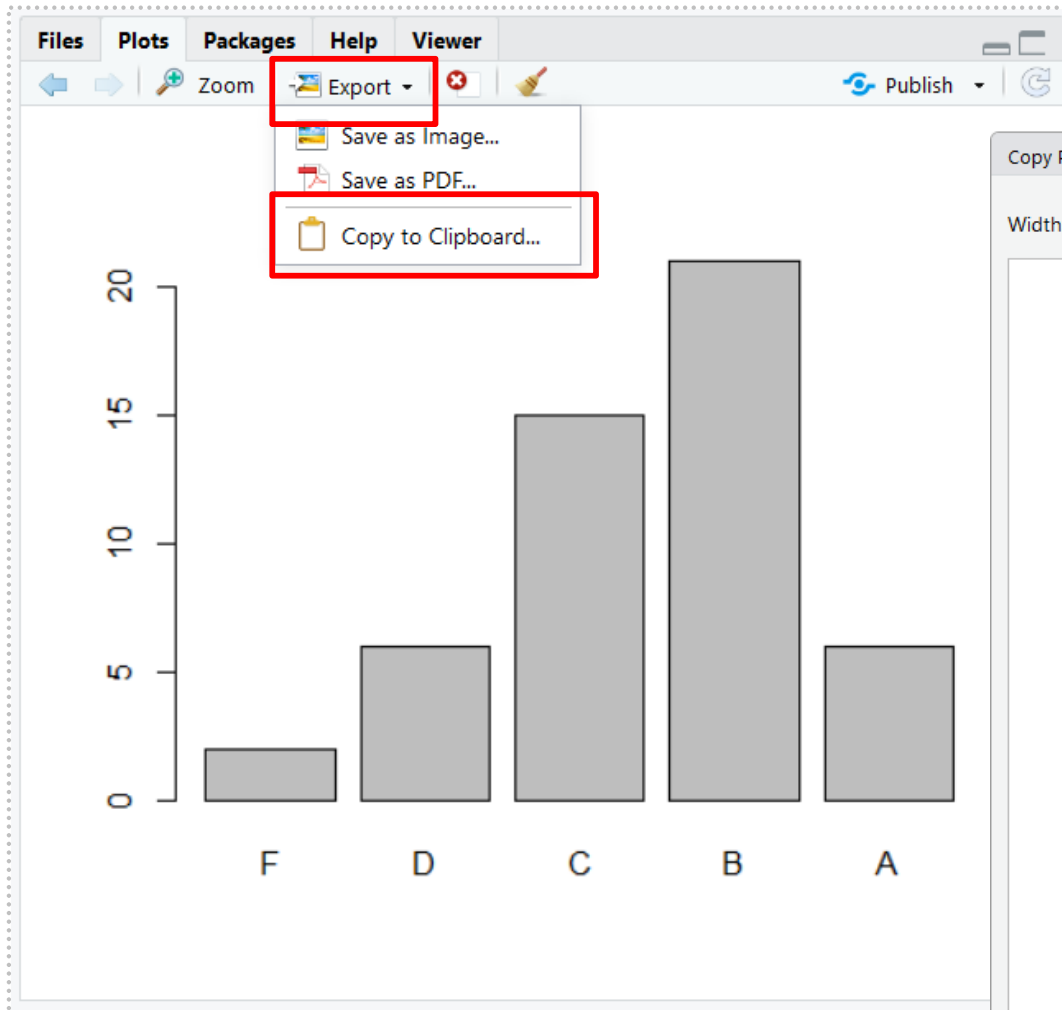
#기본 막대그래프

```
grade <- table(grade)
```

```
barplot(grade)
```



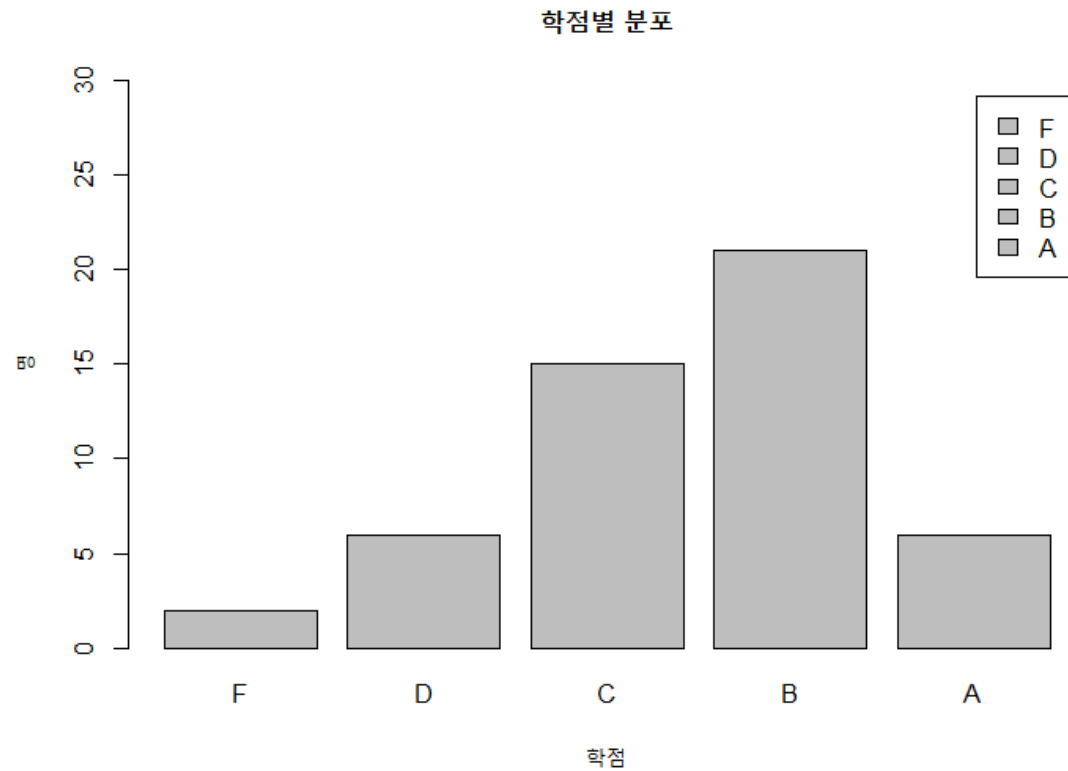
# 막대그래프(barplot)



# 막대그래프(barplot)

# 막대그래프 옵션

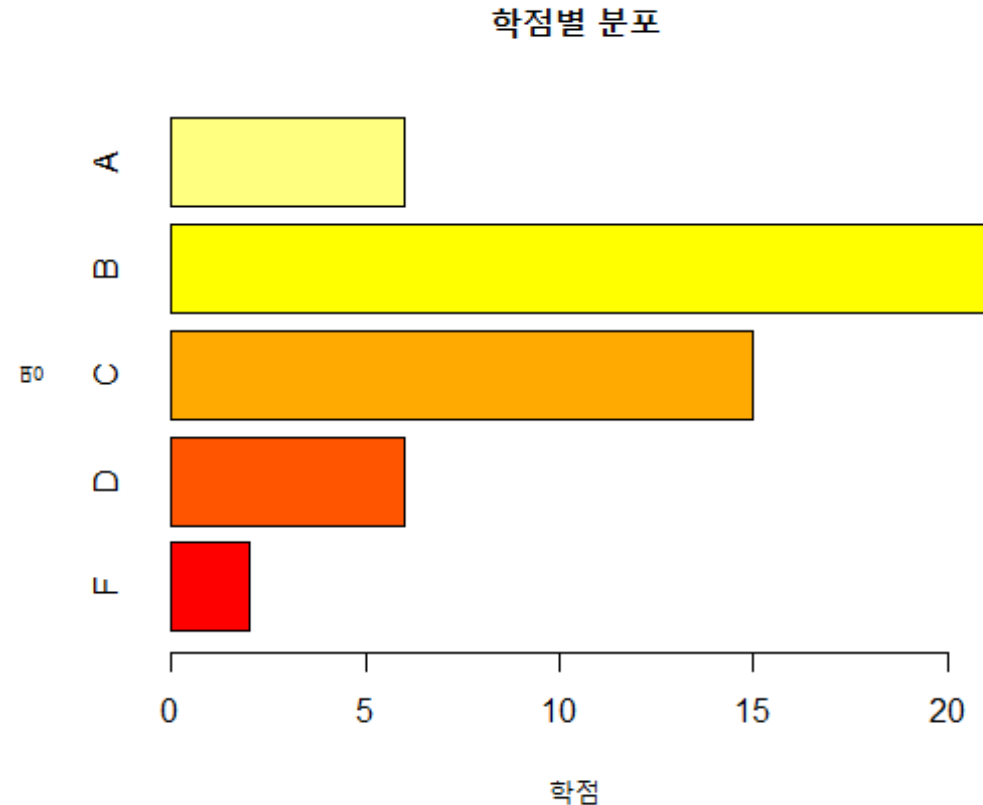
```
barplot(grade,  
        main = "학점별 분포",  
        xlab = "학점",  
        ylab = "명",  
        ylim = c(0,30),  
        legend = rownames(grade))
```



# 수평막대그래프(barplot)

#수평 막대그래프 + 컬러

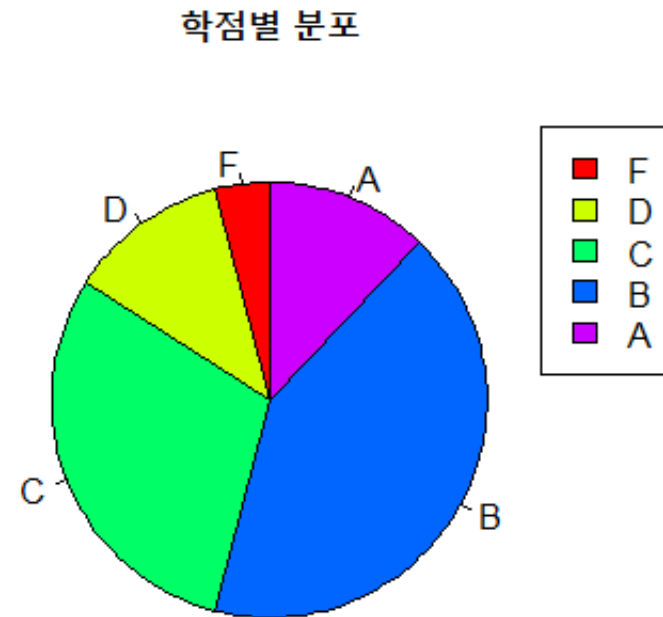
```
barplot(grade,  
        main = "학점별 분포",  
        horiz=TRUE,  
        xlab = "학점",  
        ylab = "명",  
        col=heat.colors(5))
```



# 원그래프(pie)

## #03.원그래프

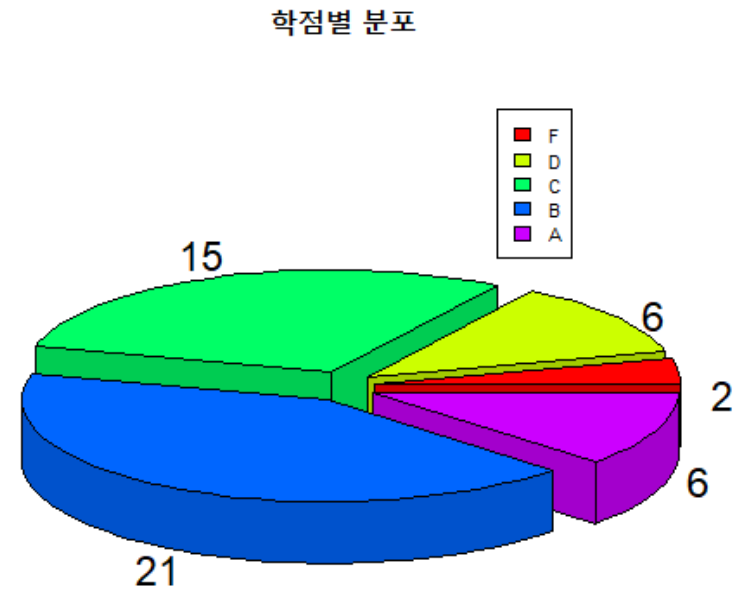
```
pie(grade,  
    main="학점별 분포",  
    init.angle=90,  
    col=rainbow(length(grade))  
)  
legend(1, 1,  
       rownames(grade),  
       cex = 1,  
       fill = rainbow(length(grade)))
```



# 원그래프(pie)

```
# 3D 원그래프 옵션
# install.packages("plotrix")
library(plotrix)

pie3D(grade,
      main="학점별 분포",
      labels=grade,
      explode=0.1,
      col=rainbow(length(grade))
)
legend(0.5, 1,
      rownames(grade),
      cex = 0.7,
      fill = rainbow(length(grade))
)
detach(freq)
```



## 2. 범주형 자료(다변량)



# 사전설계 분할표

## ■ 사전(실험)설계일 때

- 사전에 그룹의 수를 결정해서 연구할 때
- 그룹에 따른 차이를 연구할 때
- 예) 비타민과 감기에 대한 연구를 하기 위해, 비타민을 투여할 실험군과 가짜약을 투여할 대조군으로 사전에 구분하여 연구
- 실험군과 대조군에 따른 차이를 검정: 교차분석(동질성검정)
- 비율기준: 그룹별 자료수

사후

사전

그룹	감기발병		합계
	유	무	
실험군 (비타민)	17 (34.0%)	33 (66.0%)	50 (100.0%)
대조군 (Placebo)	38 (76.0%)	12 (24.0%)	50 (100.0%)
합계	55 (55.0%)	45 (45.0%)	100 (100.0%)

# 누적막대그래프(barplot)

```
## Ch05.그래프그리기
```

```
##04.다변량 범주형 자료
```

```
# 데이터 가져오기 (Ch03에서 copy)
```

```
pre <- read.csv("0302.pre.csv",  
               header=TRUE,  
               na.strings = ".")  
pre$treat <- factor(pre$treat,  
                   levels=c(1,2),  
                   labels=c("비타민", "Placebo"))  
pre$cold <- factor(pre$cold,  
                  levels=c(1,2),  
                  labels=c("Cold", "noCold"))
```

```
str(pre)
```

# 누적막대그래프(barplot)

```
## 누적막대그래프(barplot) - 2개 변수
```

```
## 데이터 table로 정리
```

```
pre <- table(treat, cold)
```

```
barplot(pre,
```

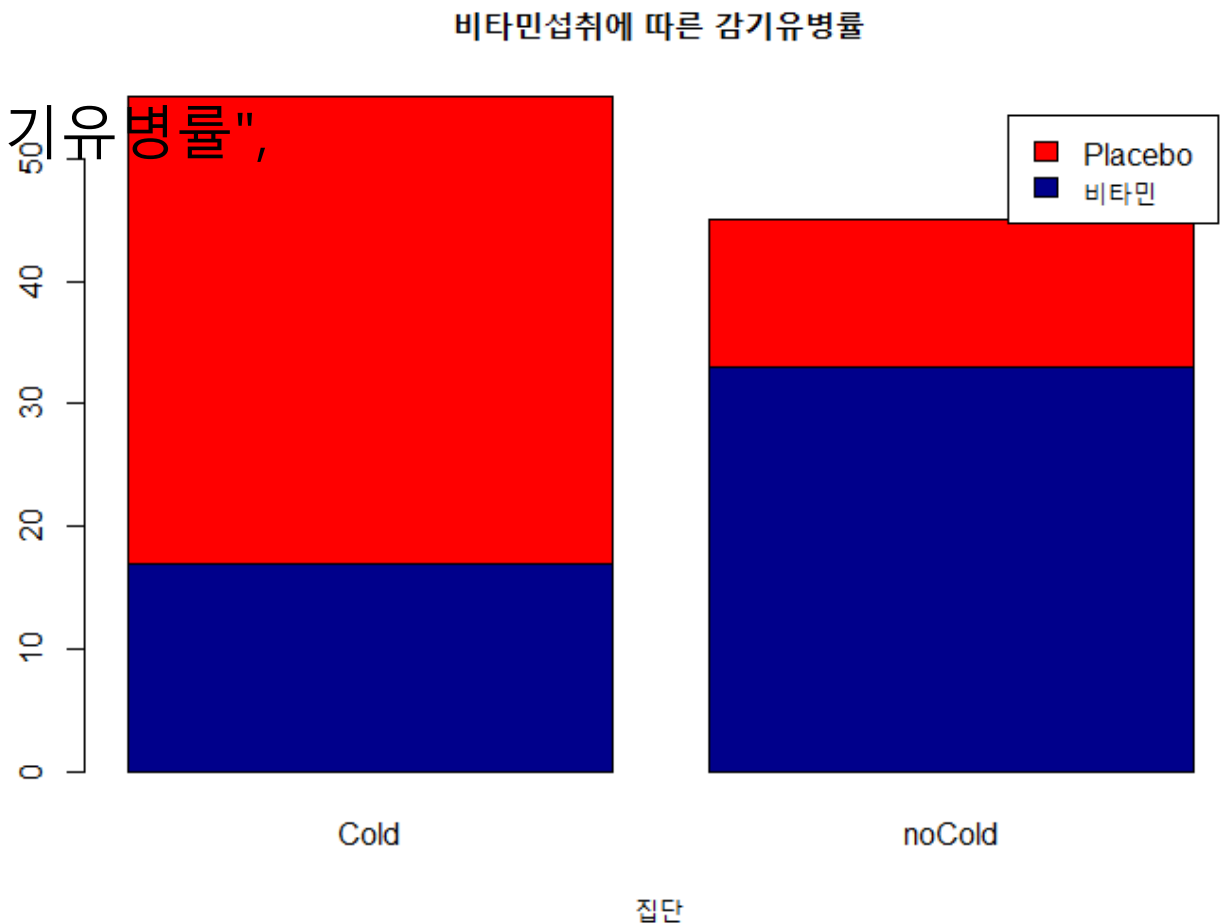
```
  main="비타민섭취에 따른 감기유병률",
```

```
  xlab = "집단",
```

```
  ylab = "%",
```

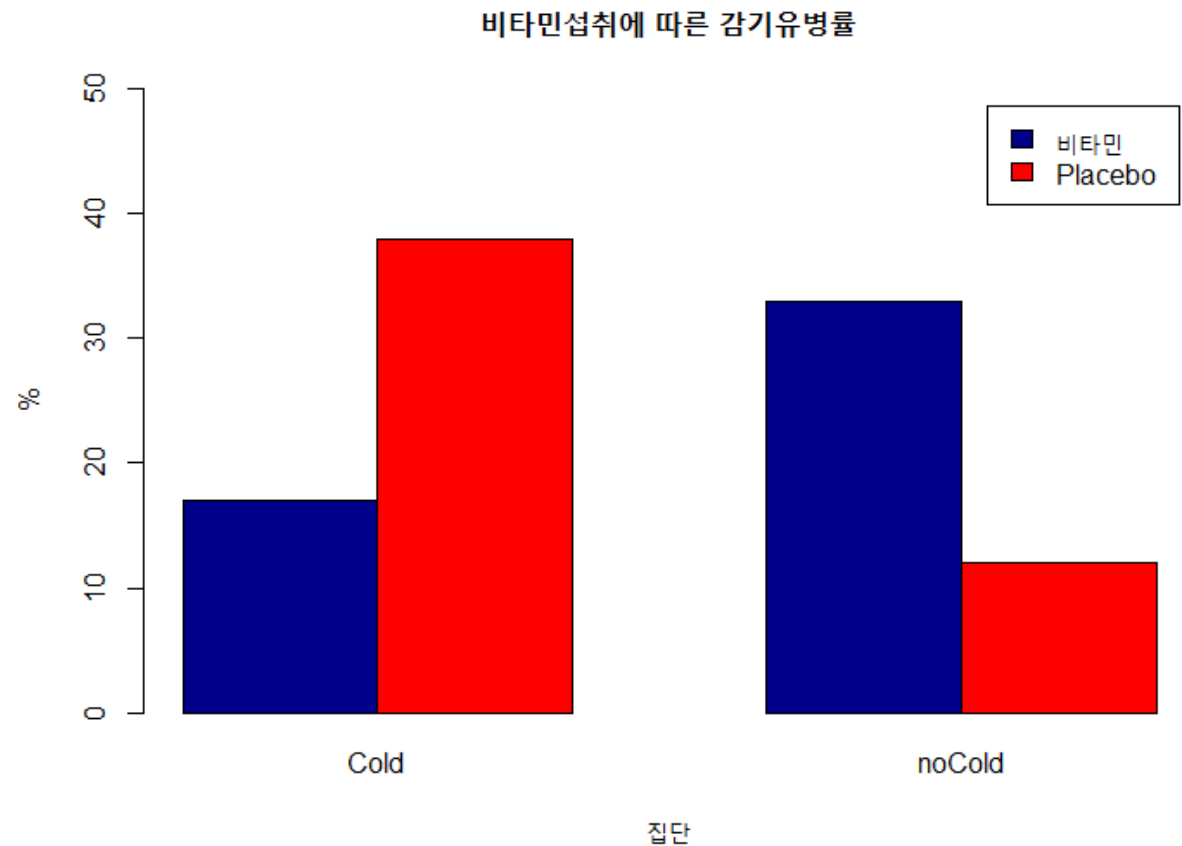
```
  col=c("darkblue","red"),
```

```
  legend = rownames(pre))
```



# 누적막대그래프(barplot)

```
barplot(pre,  
  main="비타민섭취에 따른 감기유병률",  
  xlab = "집단",  
  ylab = "%",  
  ylim = c(0,50),  
  col=c("darkblue","red"),  
  legend = rownames(pre),  
  beside=TRUE)
```



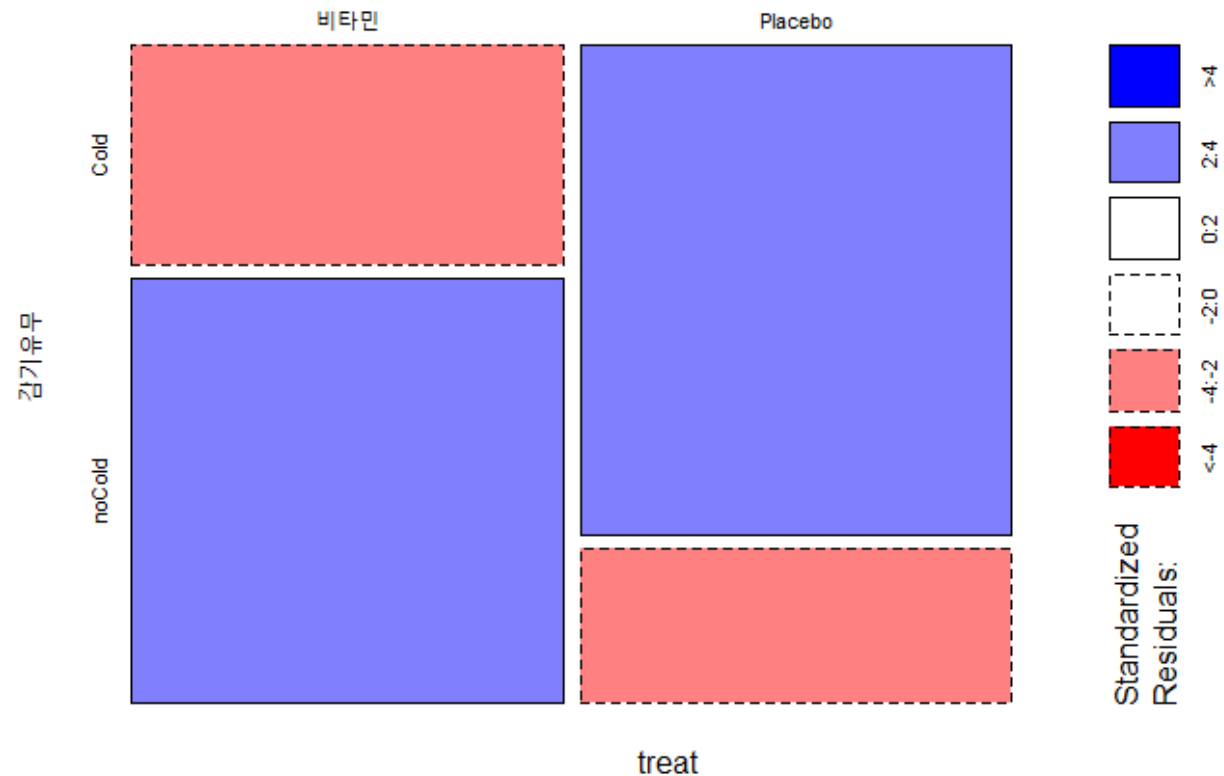
# 모자이크 그래프

## #05.모자이크 그래프

```
mosaicplot(pre, shade=TRUE,  
           xlab = "treat", ylab = "감기유무",  
           main = "비타민섭취에 따른 감기유병률")
```

비타민섭취에 따른 감기유병률

```
detach(pre)
```



### 3. 수치형 자료(일변량)

# 히스토그램 그리기

```
## Ch05.그래프그리기
```

```
##06.일변량 수치형 자료
```

```
# 데이터 가져오기 (Ch04에서 copy)
```

```
wgt <- read.csv("0401.wgt.csv",  
               header=TRUE,  
               na.strings = ".")  
  
wgt$sex <- factor(wgt$sex,  
                 levels=c(1,2),  
                 labels=c("남자","여자")  
)  
str(wgt$sex)
```

# 히스토그램 그리기

```
attach(wgt)
```

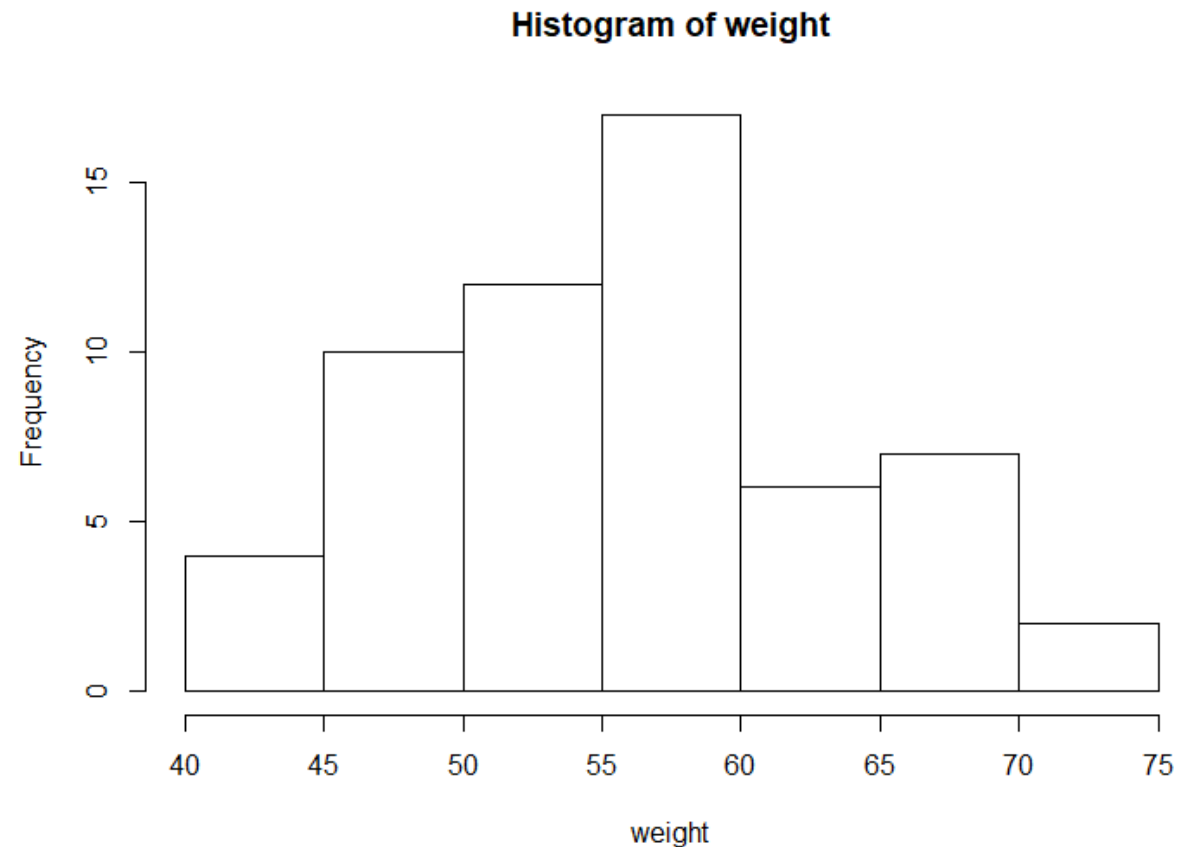
```
# 몸무게 히스토그램 및 분포
```

```
hist(weight)
```

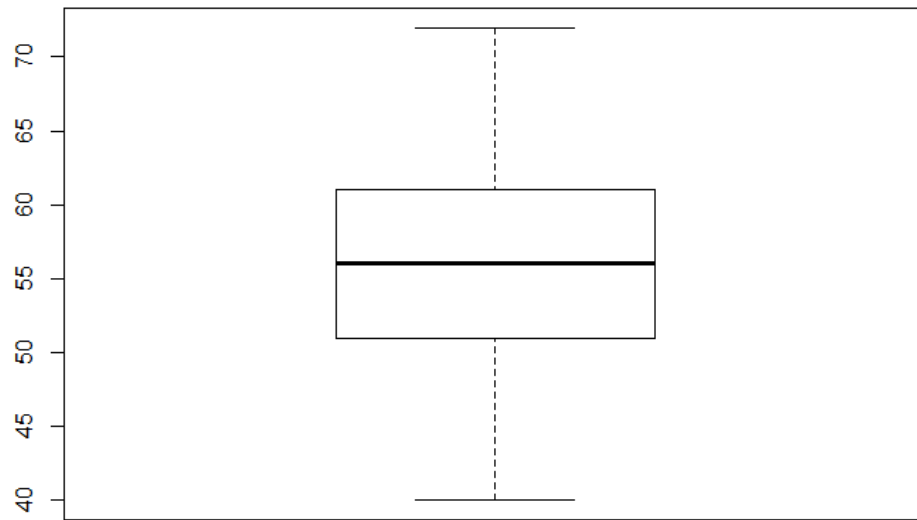
```
boxplot(weight)
```

```
stem(weight)
```

```
detach(wgt)
```







```
> stem(weight)
```

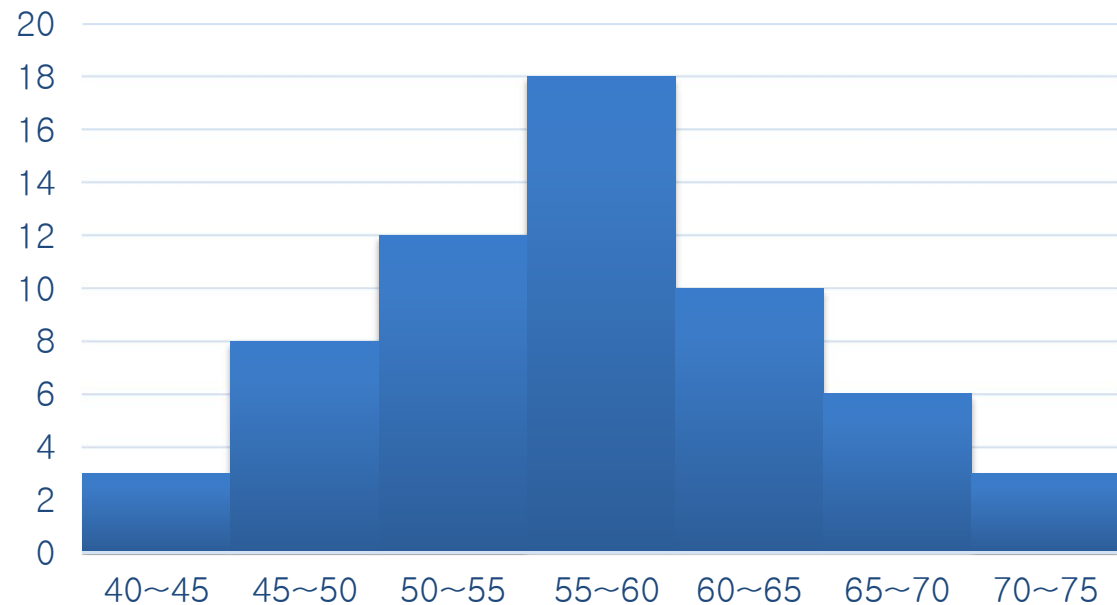
The decimal point is 1 digit(s) to the right of the |

```
4 | 014
4 | 56788899
5 | 000122223334
5 | 555666667777888999
6 | 00122223
6 | 677889
7 | 012
```

# 수치형 자료의 범주화

- 범주형 자료로 변환 후 정리
- 수치형 자료의 그래프는 히스토그램으로 정리

몸무게	빈도
40~45미만	3
45~50미만	8
50~55미만	12
55~60미만	18
60~65미만	10
65~70미만	6
70~75미만	3
합계	60



# 범주형 자료로 변환

- # 범주형 자료로 변환
- `wgt <- transform(wgt,`
- `wgt.cut = cut(weight,`
- `breaks=c(0,45,50,55,60,65,70,100),`
- `right=FALSE,`
- `labels=c("~40미만", "45~50미만","50~55미만","55~60미만",`
- `"60~65미만","65~70미만","70이상~"))`

- # `right=FALSE` →  $a \leq x < b$
- # `right=TRUE` →  $a < x \leq b$

```
> wgt
```

	id	weight	sex	wgt.cut
1	1	40	남자	~40미만
2	2	50	남자	50~55미만
3	3	56	남자	55~60미만
4	4	51	남자	50~55미만
5	5	55	남자	55~60미만
6	6	61	남자	60~65미만
7	7	70	남자	70이상~
8	8	44	남자	~40미만
9	9	66	남자	65~70미만

# 범주형 자료로 변환

```
# 범주형 자료로 변환 (ifelse이용)
wgt<- transform(wgt,
  wgt.if = ifelse(weight < 45, "~45",
    ifelse(weight >= 45 & weight < 50, "45~50미만",
      ifelse(weight >= 50 & weight < 55, "50~55미만",
        ifelse(weight >= 55 & weight < 60, "55~60미만",
          ifelse(weight >= 60 & weight < 65, "60~65미만",
            ifelse(weight >= 65 & weight < 70, "65~70미만", "70~")
          )
        )
      )
    )
  )
)
```

# 범주형 자료로 변환

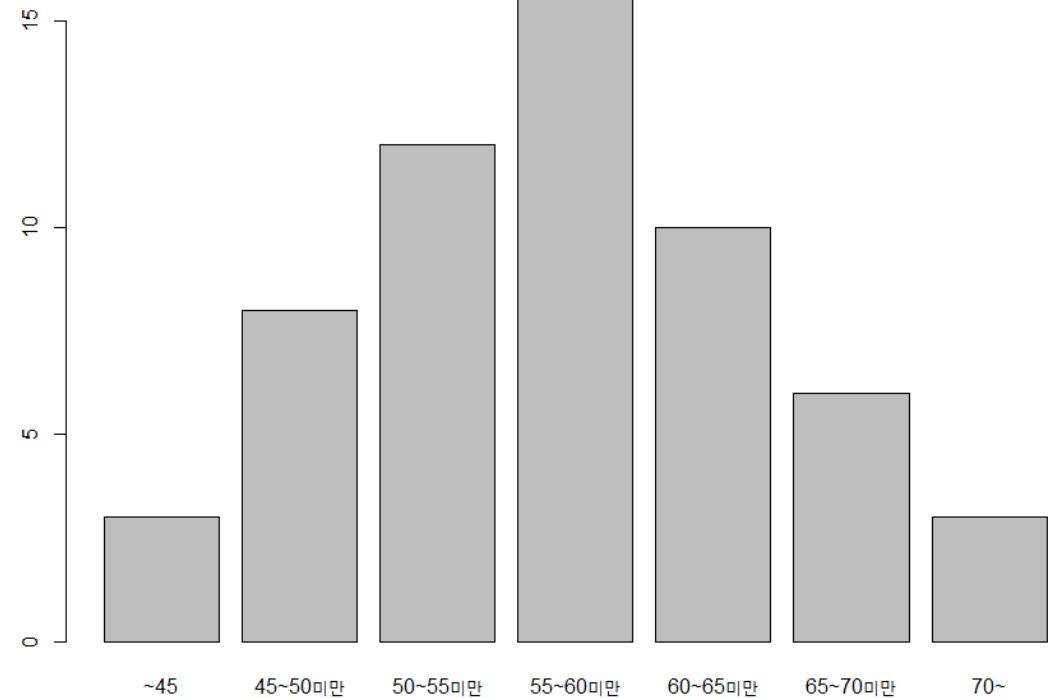
```
detach(wgt)
```

# 데이터프레임을 새롭게 수정했을 경우에는 detach한후에 다시 attach해줘야 함

```
attach(wgt)
```

```
table(wgt.cut)
```

```
barplot(table(wgt.cut))
```

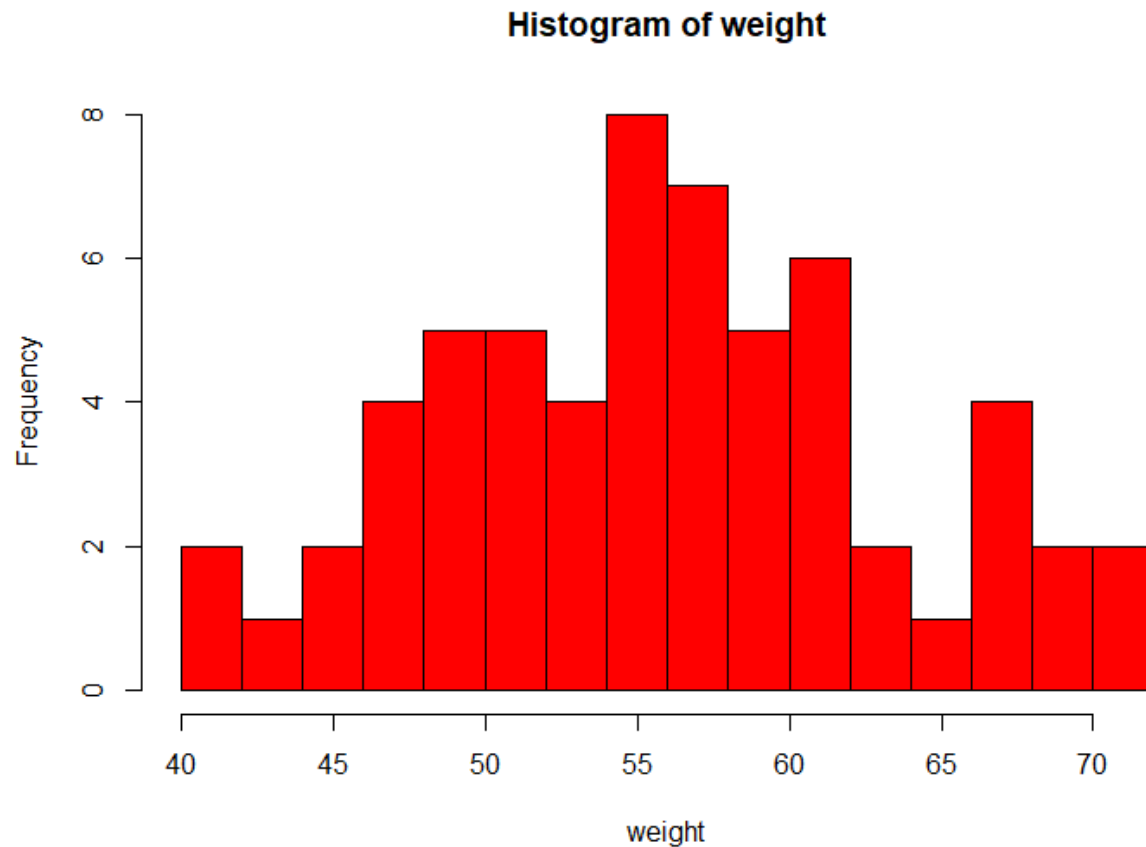


# 범주형 자료로 변환

```
# 히스토그램 bin 설정
```

```
hist(weight,  
      breaks=15,  
      col="red",  
      min(weight),  
      max(weight))
```

```
detach(wgt)
```



## 4. 수치형 변수(다변량)

# 그룹별 수치자료 비교

## ■ 그룹간 수치자료 비교

- 범주형 자료 + 수치형 자료
- 통계값: 표본크기, 평균, 표준편차

sex	N	평균	표준 편차	최소값	최대값
남	30	55.50	7.93	40	72
여	30	57.37	6.99	41	71
총계	60	56.43	7.47	40	72

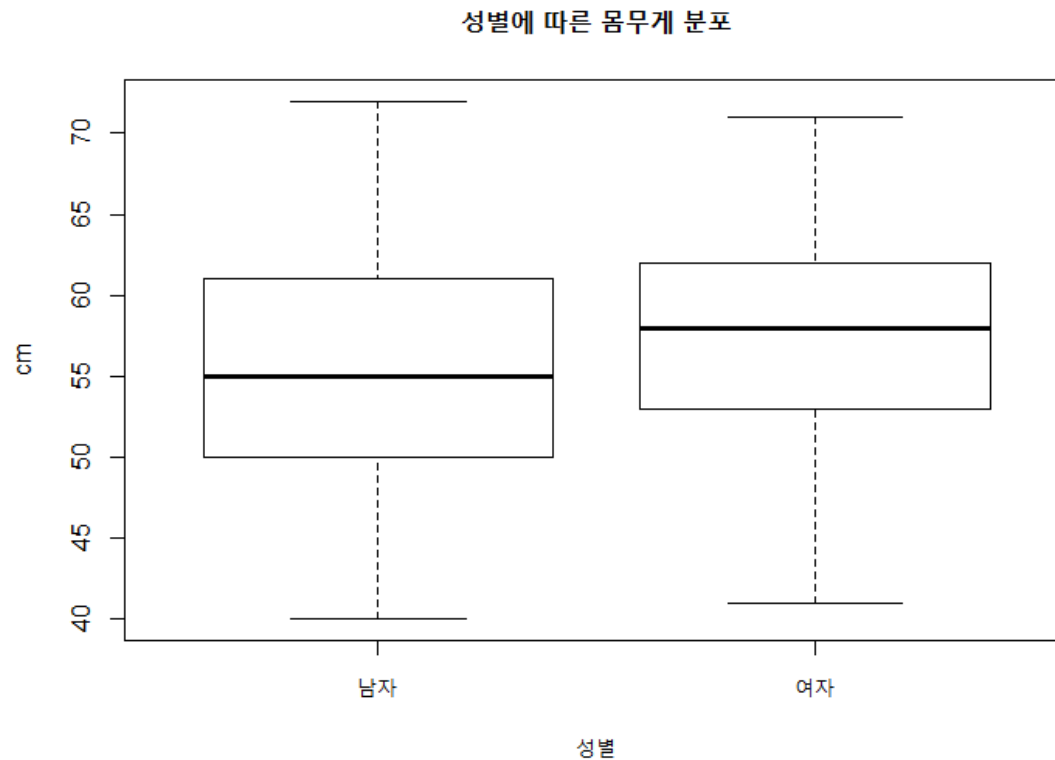
id	weight	sex
1	40	남
2	50	남
3	56	여
4	51	여
5	55	남
6	61	남
7	70	남
8	44	여
9	66	남
10	60	여
11	56	여
12	72	남
13	46	남
14	63	여
15	56	남
16	52	여
17	48	여



# 그룹간 비교

# 그룹간 연속변수 특성 비교

```
boxplot(weight~sex, data=wgt,  
        main="성별에 따른 몸무게 분포",  
        xlab="성별",  
        ylab="cm")
```



# 여러 연속자료 비교

```
# 여러 연속자료 비교
```

```
load("game.RData")
```

```
str(game)
```

```
attach(game)
```

```
o1 <- mean(game$o1)
```

```
o2 <- mean(game$o2)
```

```
fb1 <- mean(game$fb1)
```

```
fb2 <- mean(game$fb2)
```

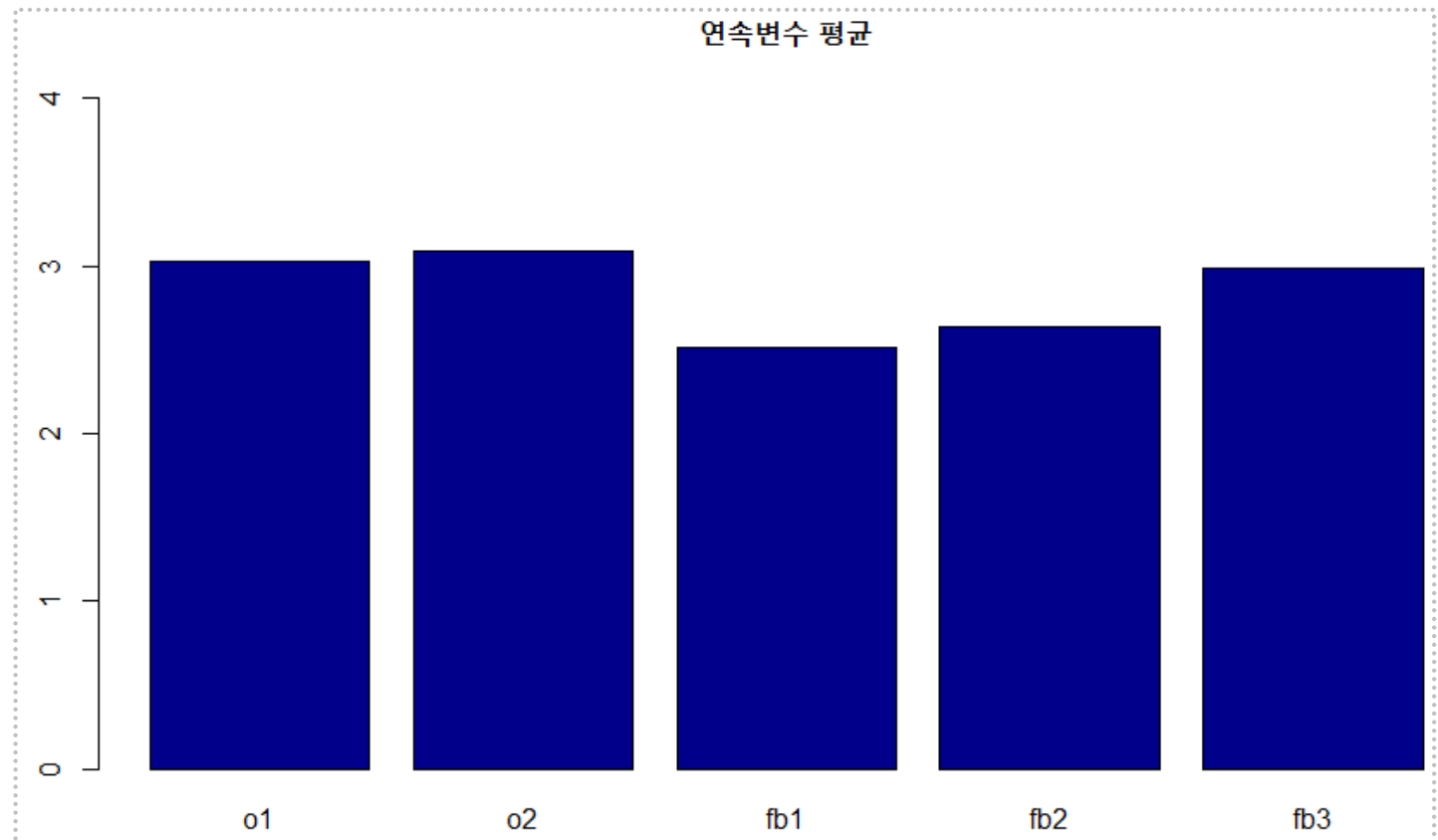
```
fb3 <- mean(game$fb3)
```

```
game.t <- cbind(o1, o2, fb1, fb2, fb3)
```

```
game.t
```

# 여러 연속자료 비교

```
barplot (game.t,  
        col=c("darkblue"),  
        main = "연속변수 평균",  
        names.arg=c(colnames(game.t)),  
        ylim=c(0,4))
```



## 5. 고급 그래프 그리기

# 고급그래프 그리기(ggplot2)

## ■ #그래프 그리기

- barplot : table로 정리한 후에 연결
- ggplot2 : 원데이터 사용

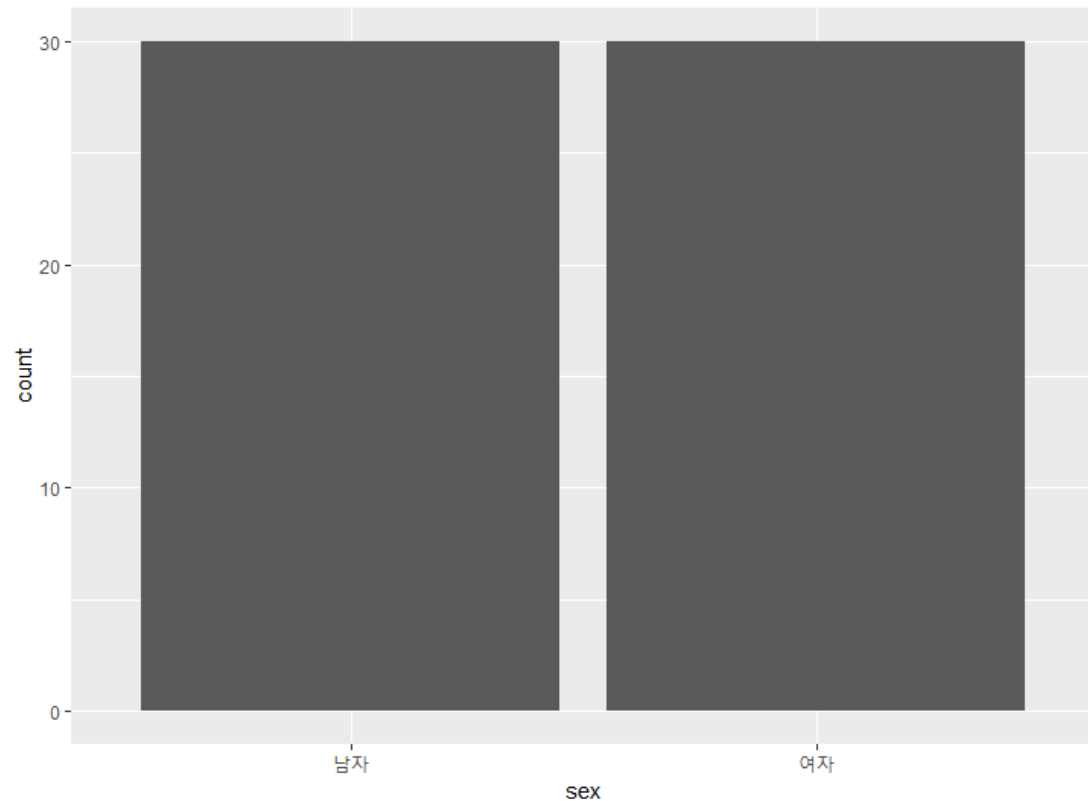
##09.고급 그래프 그리기

```
# install.packages("ggplot2")
```

```
library(ggplot2)
```

## 막대그래프

```
ggplot(wgt, aes(x=sex)) +  
  geom_bar()
```



# 고급그래프 그리기(ggplot2)

## 그룹간 상자도표(범주+수치형 자료)

```
ggplot(wgt, aes(x=sex, y=weight)) +  
  geom_boxplot(position="dodge") +  
  ggtitle("성별 체중 상자도표")
```

