

## Chapter 15. 연습문제 #1

#1-1 1. R에서 제공하는 infert 데이터셋의 education 컬럼에는 각 사람이 교육 받은 기간이 범주형 자료 형태로 저장되어 있다. infert 데이터셋의 education 컬럼 값을 잘라내어 edu 에 저장한 뒤 edu 의 값을 보이시오

```
edu <- infert[, "education"]
edu
```

```
> edu
[1] 0-5yrs 0-5yrs 0-5yrs 0-5yrs 6-11yrs 6-11yrs [133] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[7] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [139] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[13] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [145] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[19] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [151] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[25] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [157] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[31] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [163] 12+ yrs 12+ yrs 12+ yrs 0-5yrs 0-5yrs 0-5yrs
[37] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [169] 0-5yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[43] 6-11yrs 6-11yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [175] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[49] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [181] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[55] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [187] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[61] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [193] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[67] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [199] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[73] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs [205] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 12+ yrs
[79] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 0-5yrs [211] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[85] 0-5yrs 0-5yrs 0-5yrs 6-11yrs 6-11yrs 6-11yrs [217] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[91] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [223] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[97] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [229] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[103] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [235] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[109] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [241] 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
[115] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs [247] 12+ yrs 12+ yrs
[121] 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs 6-11yrs
[127] 6-11yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs 12+ yrs
Levels: 0-5yrs 6-11yrs 12+ yrs
```

#1-2 2. edu 에 있는 값들을 중복을 제거하고 보이시오

```
unique(edu)
```

```
> unique(edu)
[1] 0-5yrs 6-11yrs 12+ yrs
Levels: 0-5yrs 6-11yrs 12+ yrs
```

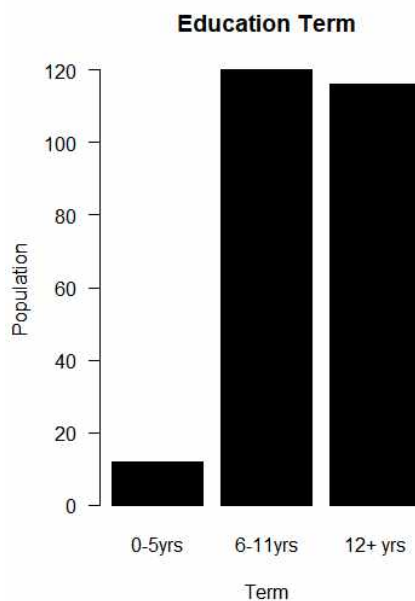
#1-3 3. edu 에 있는 값들에 대해 도수 분포표를 작성하여 보이시오

```
table(edu)
```

```
> table(edu)
edu
0-5yrs 6-11yrs 12+ yrs
    12    120    116
```

#1-4 4. edu 에 있는 값들에 대해 막대그래프를 작성하여 보이시오

```
barplot(table(edu),
        main = "Education Term",
        xlab = "Term",
        ylab = "Population",
        col = "black",
        las = 1)
```



## Chapter 15. 연습문제 #2

#2-1 1. 이 데이터를 score 벡터에 저장하시오. (과목명은 데이터 이름으로 저장하시오)

```
score <- c(90,85,73,80,85,65,78,50,68,96)
names(score) <- c("KOR","ENG","MATH","HIST","SOC","MUSIC",
                 "BIO","EARTH","PHY","ART")
```

#2-2 2. score 벡터의 내용을 보이시오

score

```
> score
  KOR  ENG  MATH  HIST  SOC  MUSIC  BIO
  90   85   73   80   85   65   78
EARTH  PHY  ART
  50   68   96
```

#2-3 3. 전체 성적의 평균은 얼마인가

mean(score)

```
> mean(score)
[1] 77
```

#2-4 4. 전체 성적의 중앙값은 얼마인가

median(score)

```
> median(score)
[1] 79
```

#2-5 5. 전체 성적의 표준편차를 보이시오

sd(score)

```
> sd(score)
[1] 13.49074
```

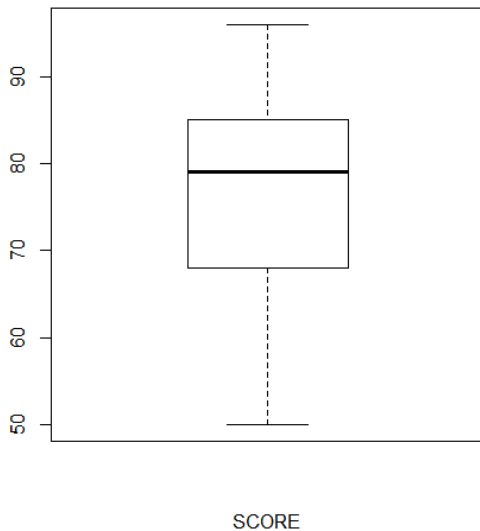
#2-6 6. 가장 성적이 높은 과목의 이름을 보이시오

```
max_score <- subset(score, score == max(score))
names(max_score)
```

```
> sc <- data.frame(score)
> rownames(subset(sc, score == max(score)))
[1] "ART"
```

#2-7 7. 성적에 대한 boxplot을 그리시오. 이상 치에 해당하는 과목이 있으면 제시하시오

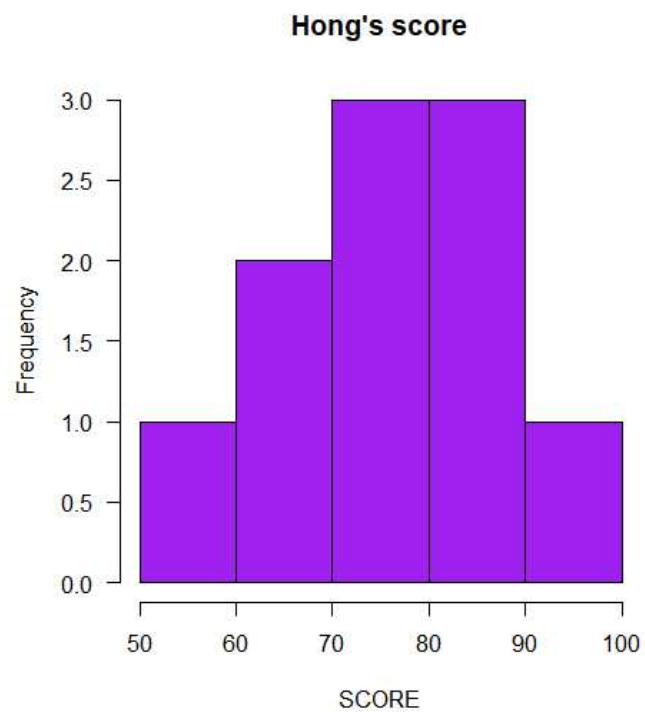
```
boxplot(score, xlab = "SCORE")
```



- 이상 치는 존재하지 않는다

#2-8 8. 성적에 대한 histogram을 그리되 다음조건을 만족하도록 하시오 (그래프 title : Hong's score, 막대색: 보라색)

```
hist(score,  
      main = "Hong's score",  
      col = "purple",  
      xlab = "SCORE",  
      las = 1)
```



## Chapter 15. 연습문제 #3

#3-1 1. 중량(wt) 의 평균값, 중앙값, 절사평균값(절사범위:15%), 표준편차를 구하시오

```
wt <- (mtcars$wt)
mean(wt)
median(wt)
mean(wt, trim=0.15)
sd(wt)
```

```
> wt <- (mtcars$wt)
> mean(wt)
[1] 3.21725
> median(wt)
[1] 3.325
> mean(wt, trim=0.15)
[1] 3.165208
> sd(wt)
[1] 0.9784574
```

#3-2 2. 중량(wt)에 대해 summary() 함수의 적용 결과를 보이시오

```
summary(wt)
```

```
> summary(wt)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.513   2.581   3.325   3.217   3.610   5.424
```

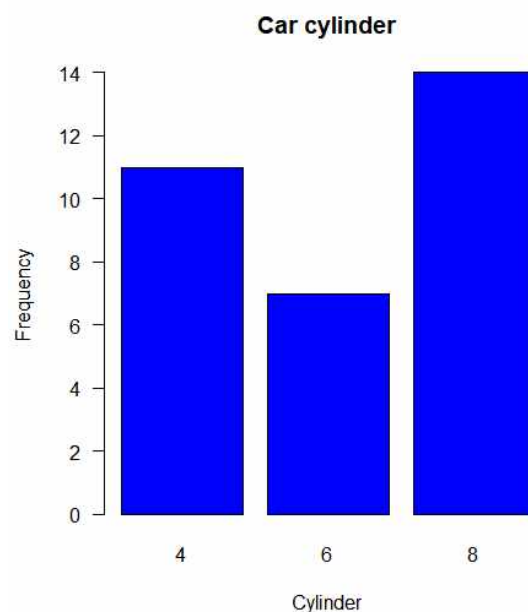
#3-3 3. 실린더수(cyl) 에 대해 도수분포표를 구하시오

```
cyl <- (mtcars$cyl)
table(cyl)
```

```
> cyl <- (mtcars$cyl)
> table(cyl)
cyl
 4  6  8
11  7 14
```

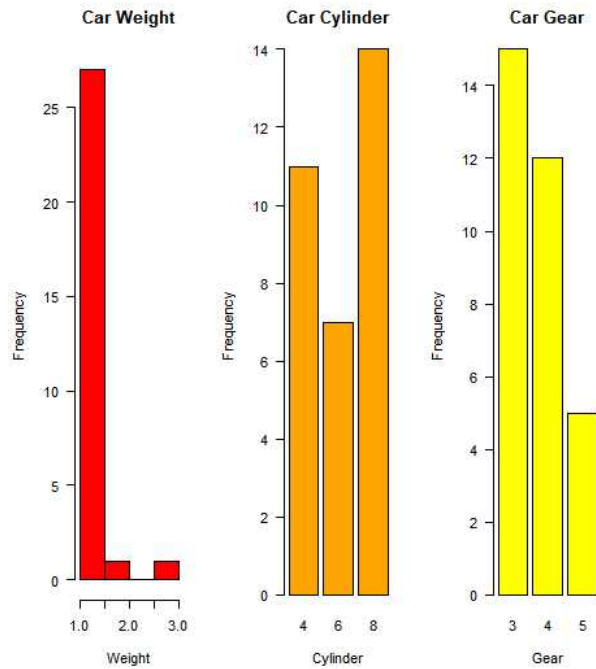
#3-4 4. 앞에서 구한 도수분포표를 막대그래프로 그려 보시오

```
barplot(table(cyl),
        main = "Car cylinder",
        xlab = "Cylinder",
        ylab = "Frequency",
        col = "blue",
        las = 1)
```



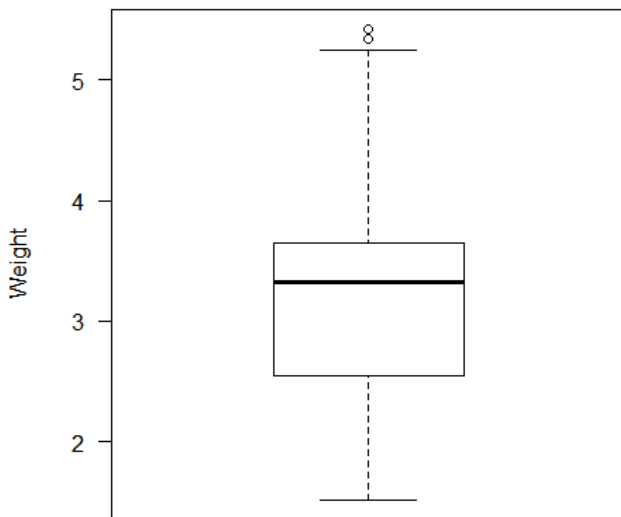
#3-5 5. 중량(wt)의 히스토그램, 실린더(cyl), 기어(gear) 에 대한 막대 그래프를 한 화면에 보이게 작성하시오

```
par(mfrow = c(1,3))
hist(table(mtcars$wt),
      main = "Car Weight",
      col = "red",
      xlab = "Weight",
      las = 1)
barplot(table(mtcars$cyl),
        main = "Car Cylinder",
        col = "orange",
        xlab = "Cylinder",
        ylab = "Frequency",
        las = 1)
barplot(table(mtcars$gear),
        main = "Car Gear",
        col = "yellow",
        xlab = "Gear",
        ylab = "Frequency",
        las = 1)
par(mfrow = c(1,1))
```



#3-6 6. 중량(wt)에 대해 boxplot 을 그려 보시오. Boxplot 으로 부터 관찰할 수 있는 정보를 적으시오

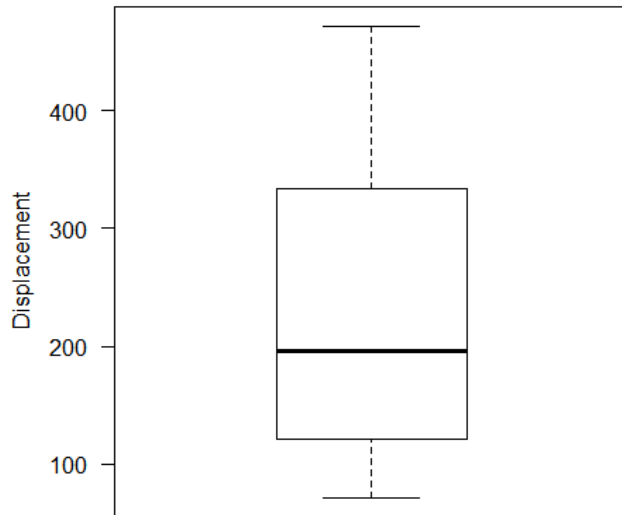
```
boxplot(mtcars$wt,
        ylab = "Weight",
        las = 1)
```



- 값보다 큰 이상 치에 해당하는 값이 두 개가 있다
- 제1사분위수는 약 2.6으로 나타났다
- 중앙값은 약 3.4로 나타났다
- 제3사분위수는 약 3.7로 나타났다
- 최솟값은 약 1.5로 나타났다
- 최댓값은 약 5.4로 나타났다

#3-7 7. 배기량(displacement)에 대해 boxplot 을 그려 보시오. Boxplot 으로 부터 관찰할 수 있는 정보를 적으시오

```
boxplot(mtcars$displacement,  
        ylab = "Displacement",  
        las = 1)
```



- 이상 치에 해당하는 값은 나타나지 않았다
- 제1사분위수는 약 120으로 나타났다
- 중앙값은 약 200으로 나타났다
- 제3사분위수는 약 330으로 나타났다
- 최솟값은 약 70으로 나타났다
- 최댓값은 약 470으로 나타났다