

데이터베이스 기초

15장 과제

<연습1>

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 저장한뒤 edu 의 값을 보이시오

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

(결과 생략)

2. edu 에 있는 값들을 중복을 제거하고 보이시오

```
unique(edu)
```

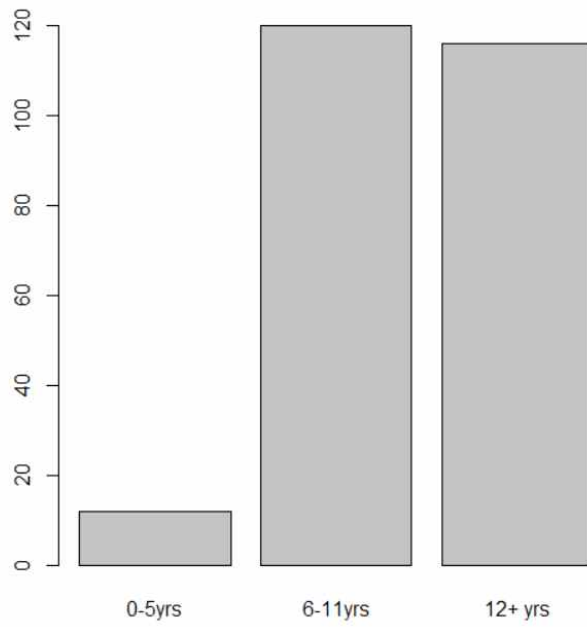
(결과 생략)

3. edu 에 있는 값들에 대해 도수 분포표를 작성하여 보이시오

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

4. edu 에 있는 값들에 대해 막대 그래프를 작성하여 보이시오

```
barplot(table(edu))
```



<연습2>

1. 이 데이터를 score 벡터에 저장하시오. (과목명은 데이터 이름으로 저장 하시오)

```
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. score 벡터의 내용을 보이시오

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

3. 전체 성적의 평균은 얼마인가

```
mean(score)
```

(결과 생략)

4. 전체 성적의 중앙값은 얼마인가

```
median(score)
```

(결과 생략)

5. 전체 성적의 표준편차를 보이시오

```
sd(score)
```

(결과 생략)

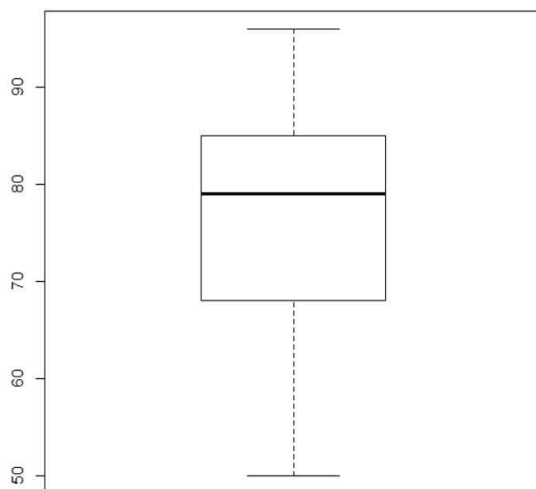
6. 가장 성적이 높은 과목의 이름을 보이시오

```
names(score[score==max(score)])
```

(결과 생략)

7. 성적에 대한 boxplot 을 그리시오. 이상치에 해당하는 과목이 있으면 제시하시오

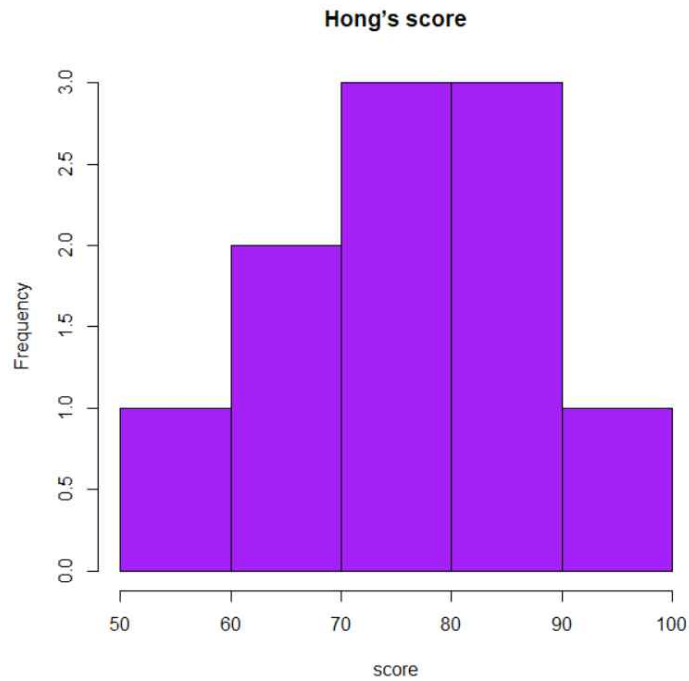
```
boxplot(score)
```



이상치 없음

8. 성적에 대한 histogram 을 그리되 다음조건을 만족하도록 하시오 (그래프 title : Hong's score, 막대색: 보라색)

```
hist(score, main="Hong's score", col="purple")
```



<연습3>

- mtcars 데이터셋을 이용하여 다음 문제를 해결하시오

1. 중량(wt) 의 평균값, 중앙값, 절사평균값(절사범위:15%), 표준편차 를 구 하시오

평균값 : `mean(mtcars[,"wt"])`

중앙값 : `median(mtcars[, "wt"])`

절사평균값 : `mean(mtcars[, "wt"],trim=0.15)`

표준편차 : `sd(mtcars[, "wt"])`

2. 중량(wt)에 대해 `summary()` 함수의 적용 결과를 보이시오

`summary(mtcars[, "wt"])`

```
summary(mtcars[ , "wt"])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.513   2.581   3.325   3.217   3.610   5.424
```

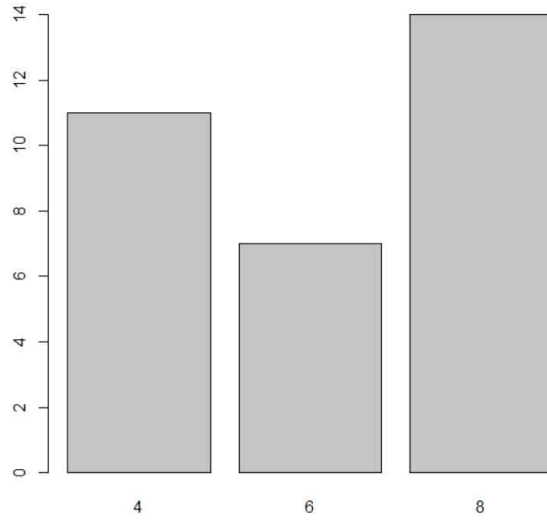
3. 실린더수(cyl) 에 대해 도수분포표를 구하시오

`table(mtcars[, "cyl"])`

```
 4    6    8
11    7   14
```

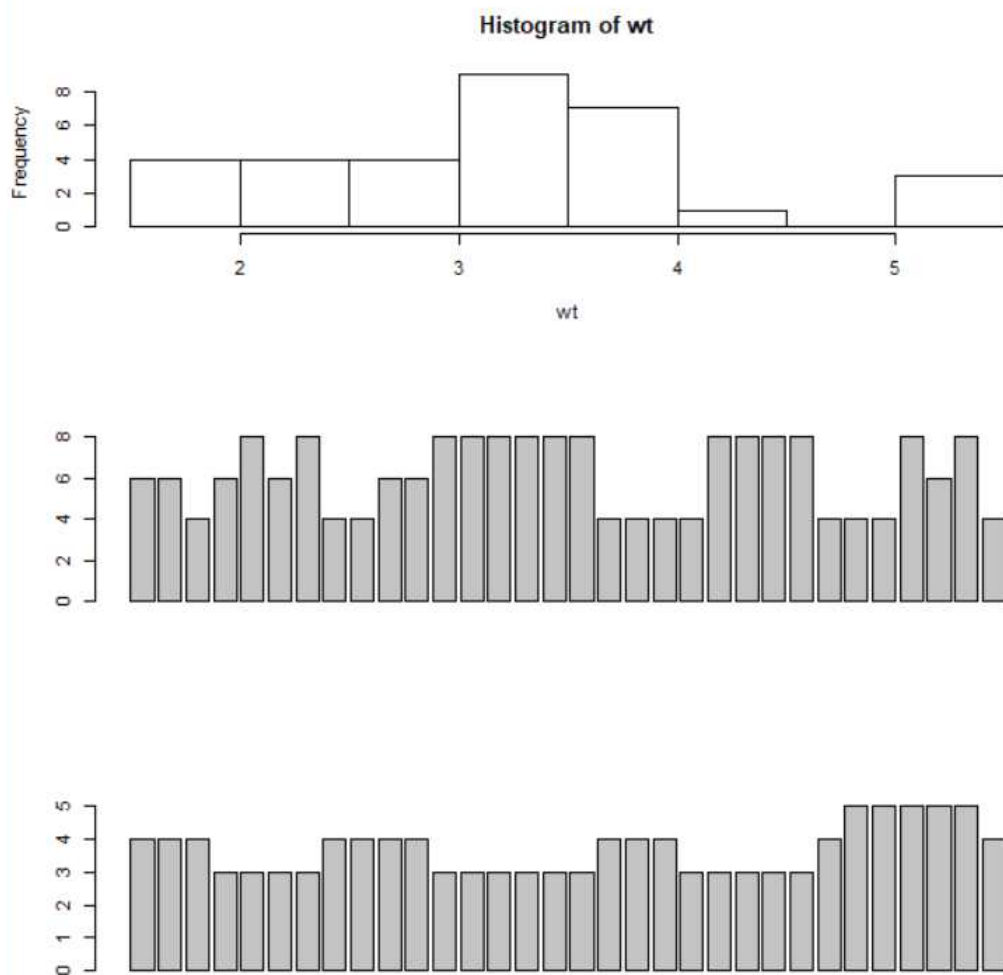
4. 앞에서 구한 도수분포표를 막대그래프로 그려 보시오

```
barplot(table(mtcars[ , "cyl"]))
```



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

```
par(mfrow = c(3, 1))  
hist(mtcars$wt)  
barplot(mtcars$cyl)  
barplot(mtcars$gear)  
par(mfrow = c(1, 1))
```



또는

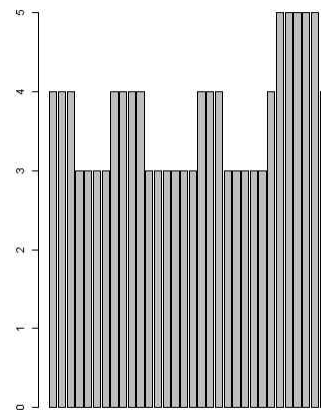
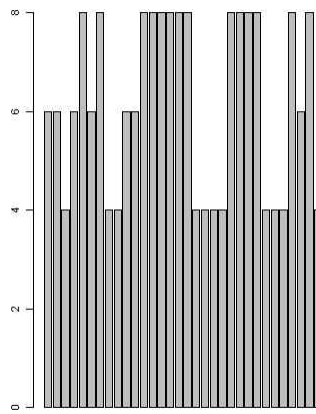
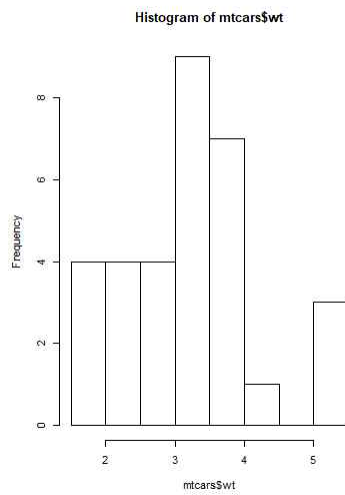
```
par(mfrow = c(1, 3))
```

```
hist(mtcars$wt)
```

```
barplot(mtcars$cyl)
```

```
barplot(mtcars$gear)
```

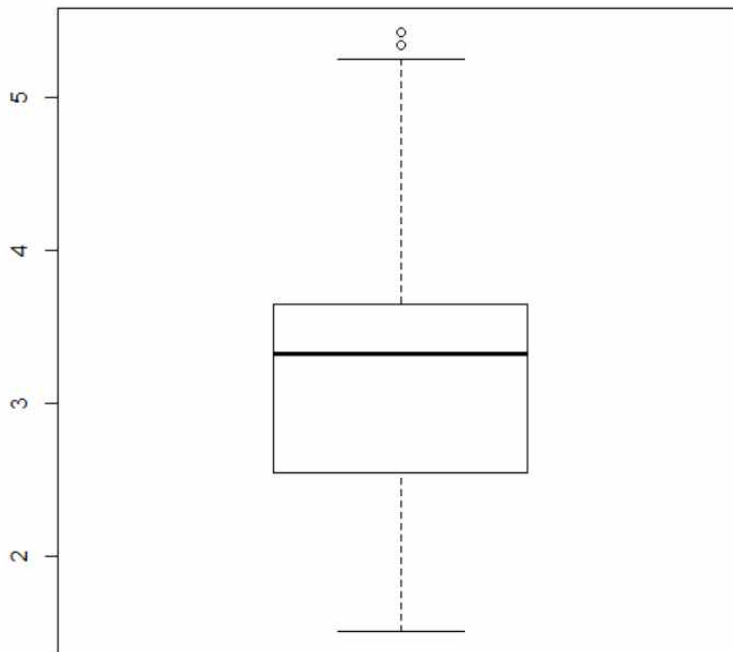
```
par(mfrow = c(1, 1))
```



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

```
boxplot(mtcars[ , "wt"])
```

또는 `boxplot(mtcars$wt)`



최소값 : 1.513

최대값 : 5.424

중앙값 : 3.325

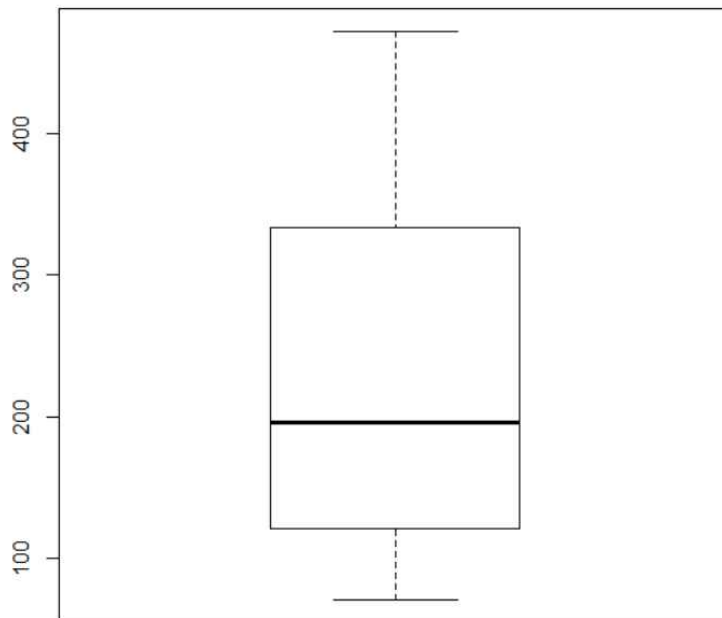
전체 데이터의 50%가 존재하는 범위 : 2.581 ~ 3.610

이상치 존재

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

```
boxplot(mtcars[ , "displacement"])
```

또는 `boxplot(mtcars$displacement)`



최소값 : 71.1

최대값 : 472.0

중앙값 : 196.3

전체 데이터의 50%가 존재하는 범위 : 120.8 ~ 326.0