Lesson 2.Answer

Sung Won Kang

2016년 11월 5일

0. 자료 준비(교재 pp. 37-48)

```
#data=read.csv("C:/work_2016/Rstat/Lesson 2/data/ch02.csv",header=F,na.strings=c("."))
data=read.csv(file="../data/ch02.csv",header=F,na.strings=c("."))
str(data)
```

```
## 'data.frame':    468284 obs. of  5 variables:
##  $ V1: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ V2: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V3: int  3 3 3 3 3 3 3 3 3 3 ...
##  $ V4: int  1 1 1 1 1 1 1 1 1 1 ...
##  $ V5: int  NA NA NA NA NA NA NA NA NA NA ...
```

```
data$V1=factor(data$V1,levels=c(1,2),labels=c("남자","여자"))
data$V3=factor(data$V3,levels=(1:14),
               labels=c("가구주","가구주의 배우자","자녀","자녀의배우자","가구주의 부모",
                        "배우자의 부모","손자녀및배우자","증손자녀및배우자","조부모","형제자매및배우자","형제자
                        "부모의형제자매및배우자","기타친인척","동거인"))
data$V4=factor(data$V4,levels=1:8,
               labels=c("안받음","초등학교","중학교","고등학교","대학4년제미만","대학4년제이상","석사과정","박
str(data)
```

```
## 'data.frame':    468284 obs. of  5 variables:
##  $ V1: Factor w/ 2 levels "남자","여자": 1 1 1 1 1 1 1 1 1 1 ...
##  $ V2: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V3: Factor w/ 14 levels "가구주","가구주의 배우자",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ V4: Factor w/ 8 levels "안받음","초등학교",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ V5: int  NA NA NA NA NA NA NA NA NA NA ...
```

```
save.image("data.rda")
```

1. graph

- 산점도 1.(Scatterplot)

```
str(cars)
```

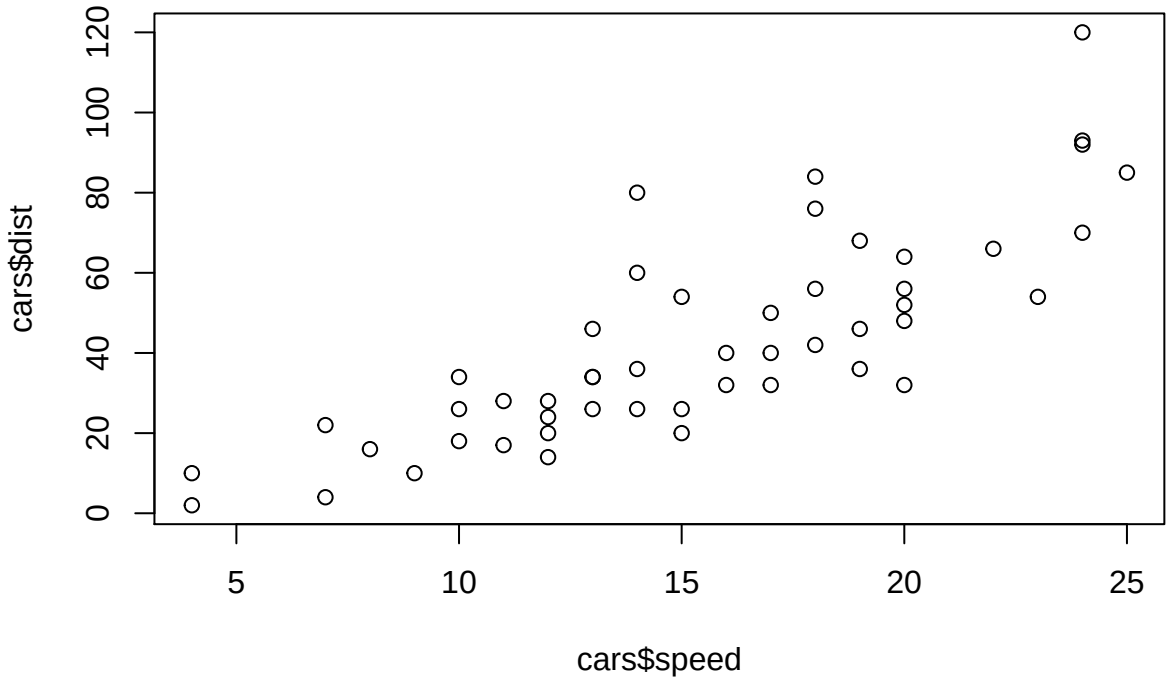
```
## 'data.frame':    50 obs. of  2 variables:
```

```
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

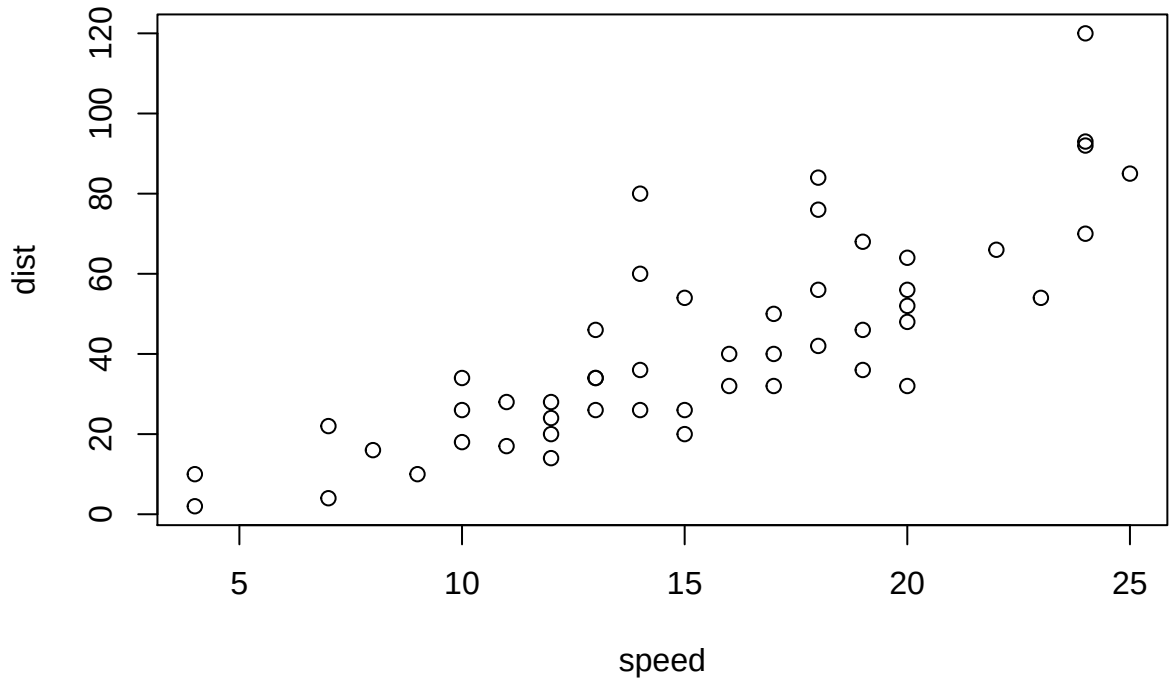
```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

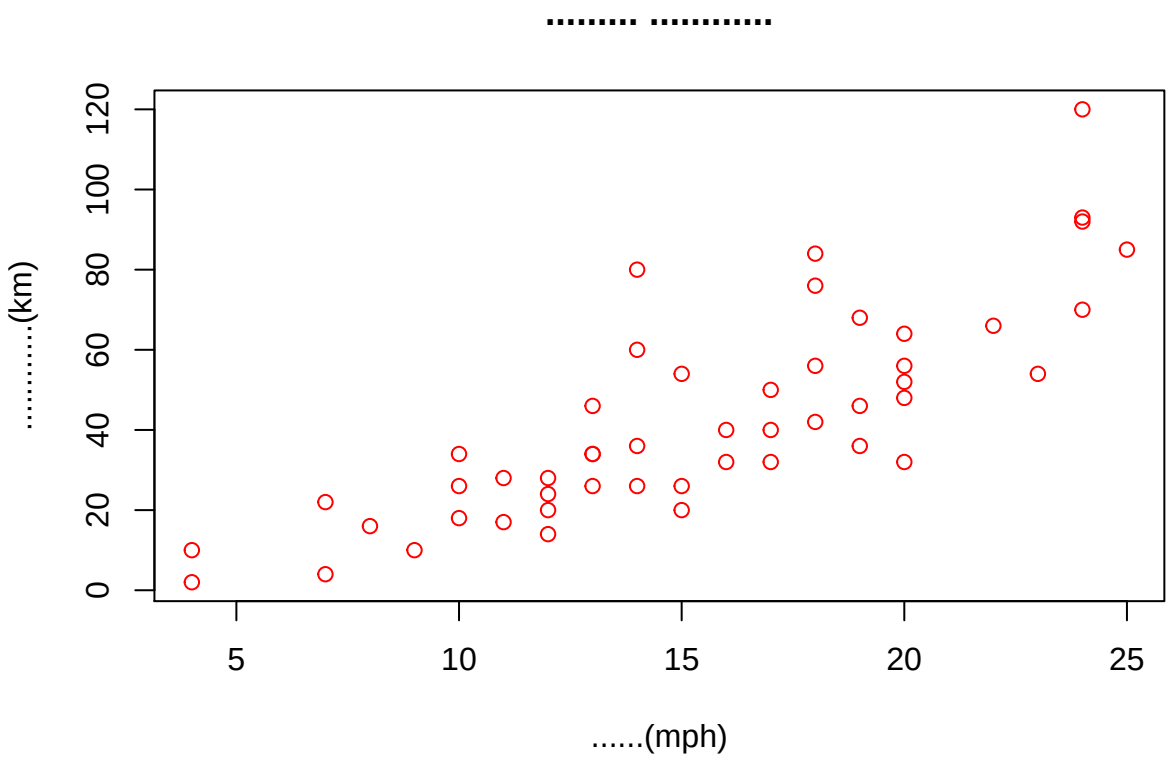
```
#그냥 보는 그래프
plot(cars$speed,cars$dist)
```




```
plot(cars)
```




```
#남에게 보여주는 그래프
plot(cars$speed,cars$dist, main="속도와 제동거리", xlab="속도(mph)",ylab="제동거리(km)",pch=1,col="red")
```




```
# pch =1 : 동그라미로 표시
# col ="red": 표시 색깔은 빨간색
```

- 산점도 2.(Scatterplot): 시계열(time series)

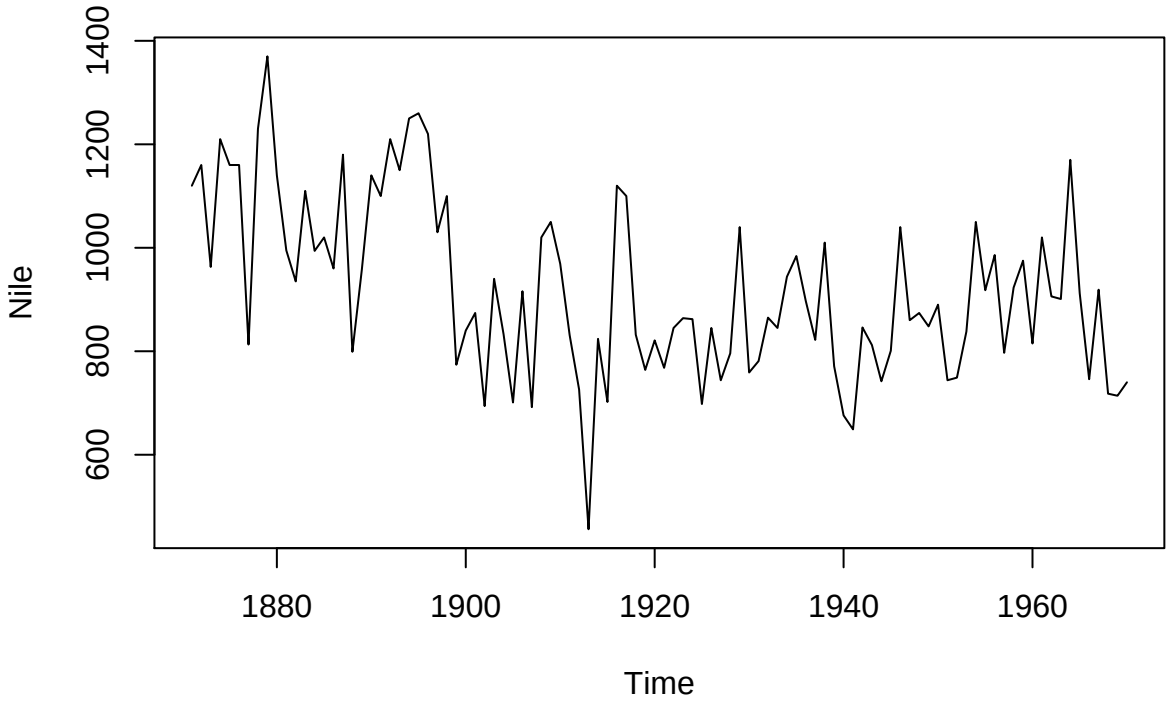
```
str(Nile)
```

```
##  Time-Series [1:100] from 1871 to 1970: 1120 1160 963 1210 1160 1160 813 1230 1370 1140 ...
```

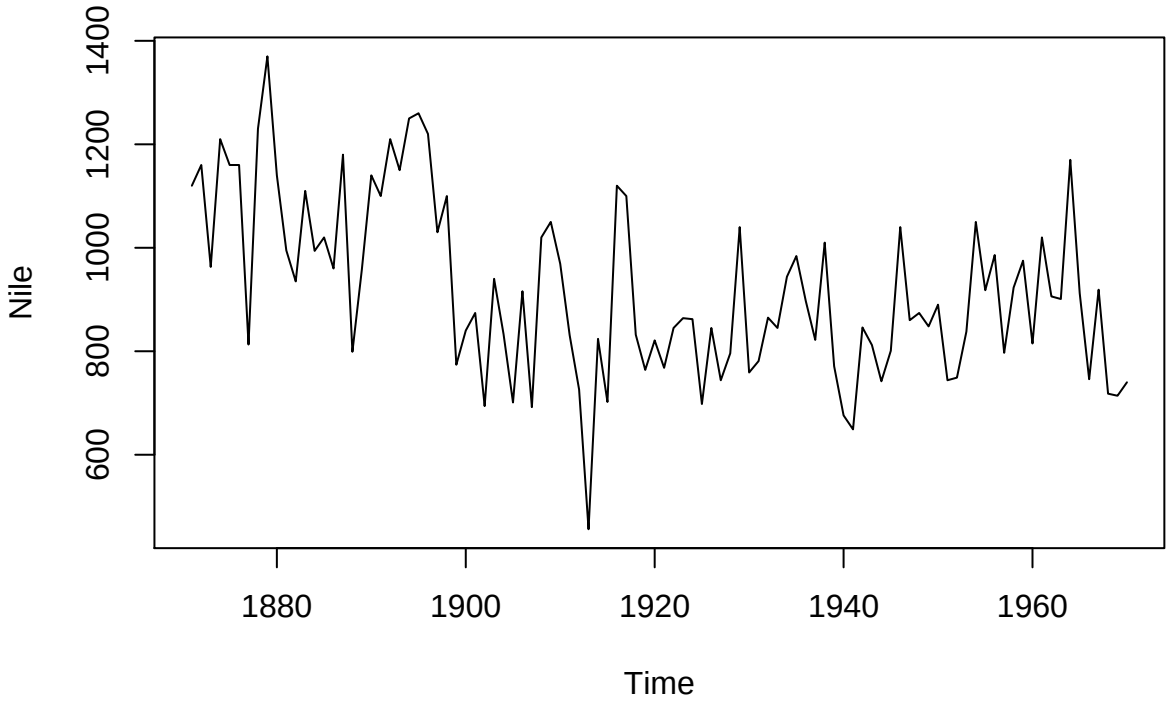
```
summary(Nile)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   456.0   798.5   893.5   919.4  1032.0  1370.0
```

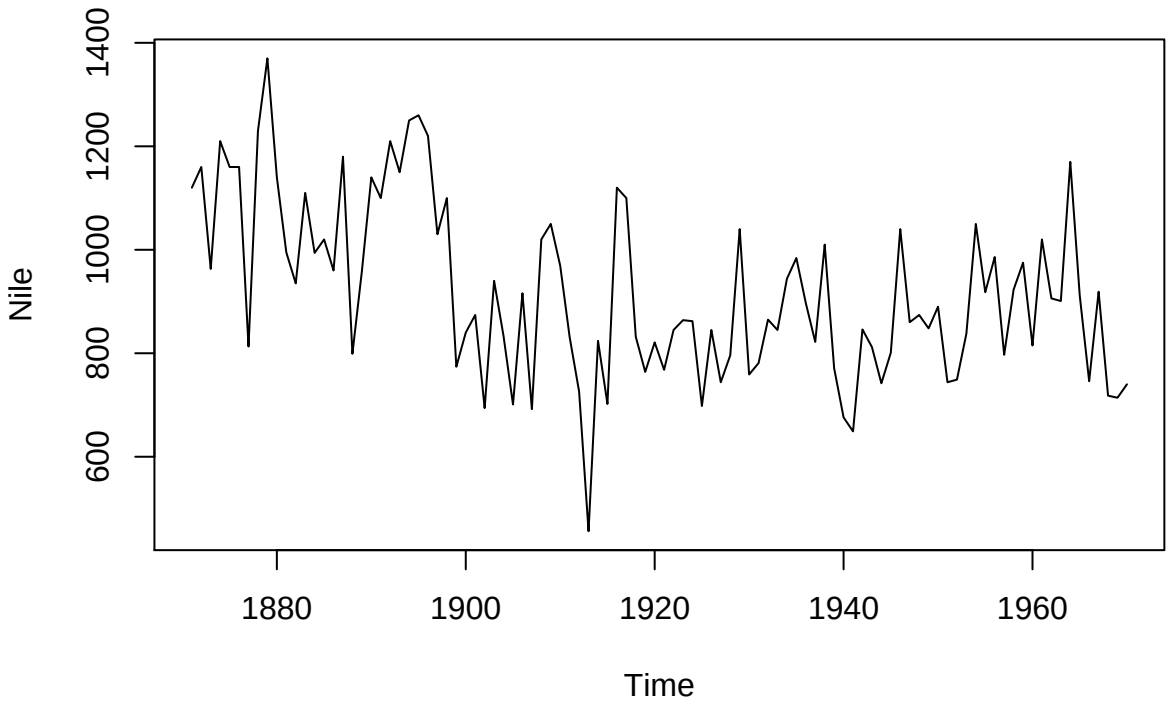
```
#그냥 보는 그래프
plot(Nile)
```



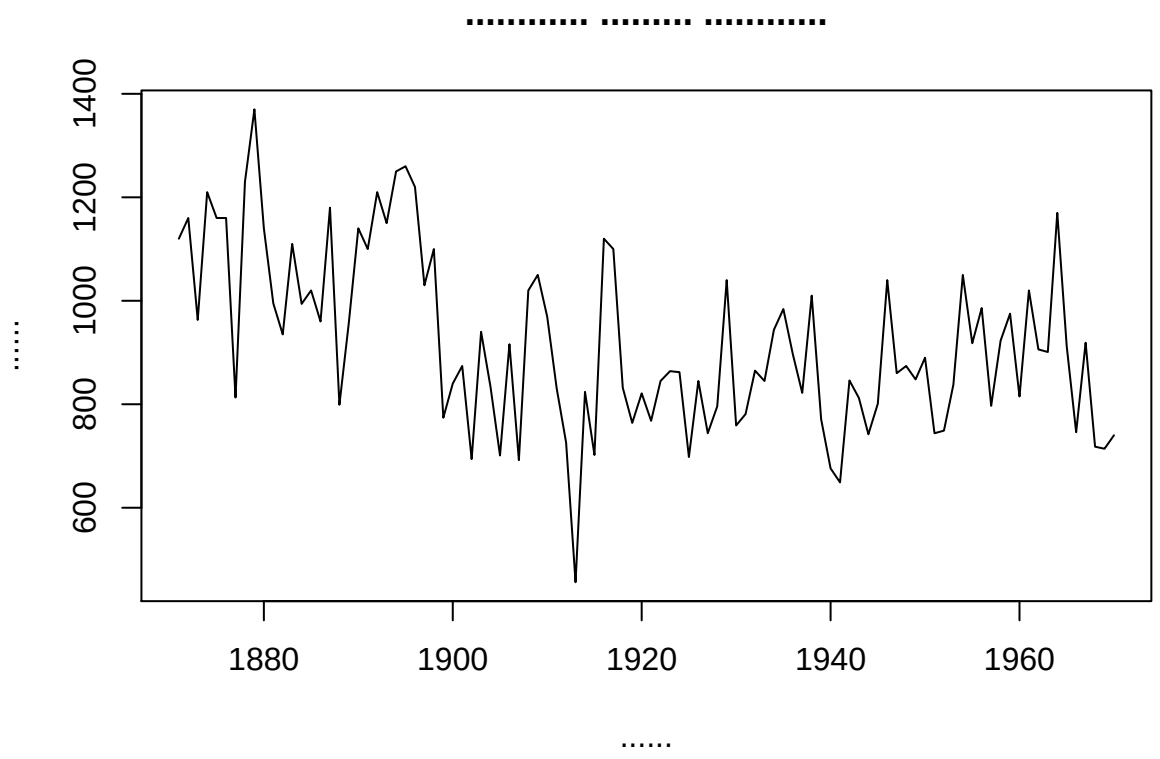

```
plot((1871:1970),Nile, xlab="Time",type="l")
```



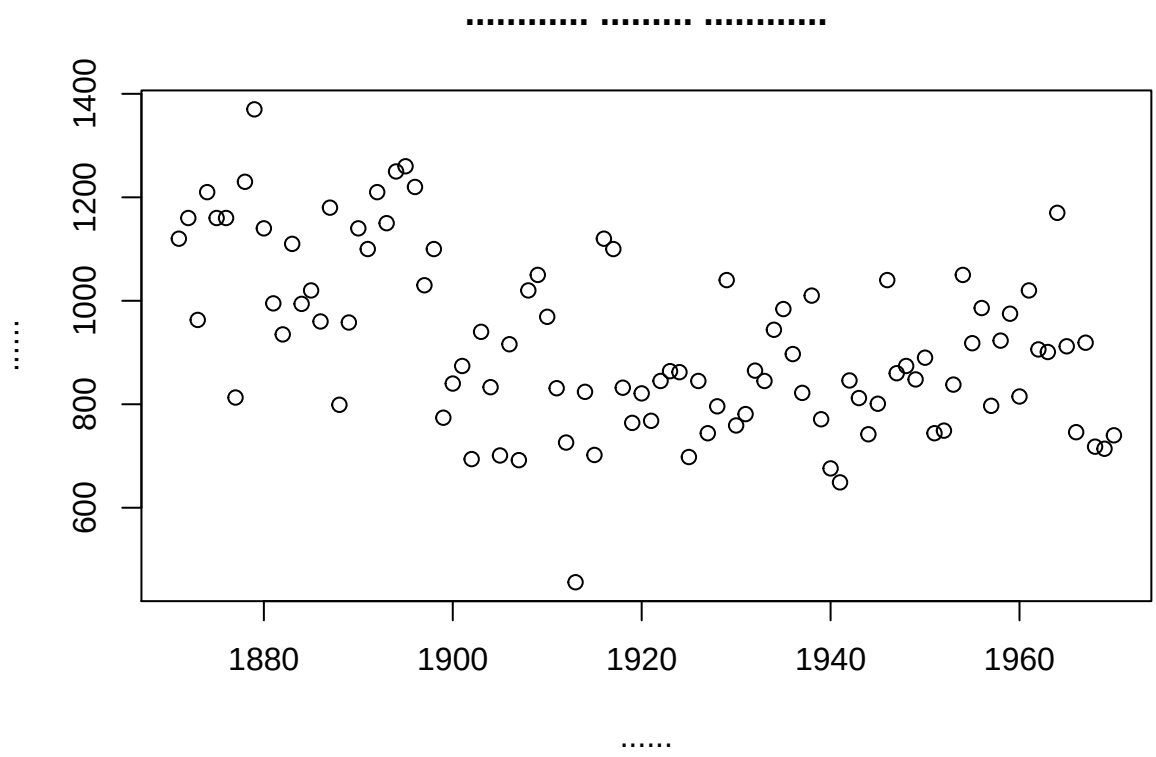

```
plot((start(Nile)[1]:end(Nile)[1]),Nile, xlab="Time",type="l")
```




```
#남에게 보여주는 그래프
plot(Nile, main="나일강의 연도별 유량변화", xlab="연도",ylab="유량")#선그래프
```




```
plot(Nile, type="p",main="나일강의 연도별 유량변화", xlab="연도",ylab="유량")#점그래프
```



- 막대그래프와 히스토그램: barplot(), hist()

```
load("data.rda")
str(data)
```

```
## 'data.frame':    468284 obs. of  5 variables:
##  $ V1: Factor w/ 2 levels "남자","여자": 1 1 1 1 1 1 1 1 1 1 ...
##  $ V2: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ V3: Factor w/ 14 levels "가구주","가구주의 배우자",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ V4: Factor w/ 8 levels "안받음","초등학교",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ V5: int  NA NA NA NA NA NA NA NA NA NA ...
```

```
summary(data)
```

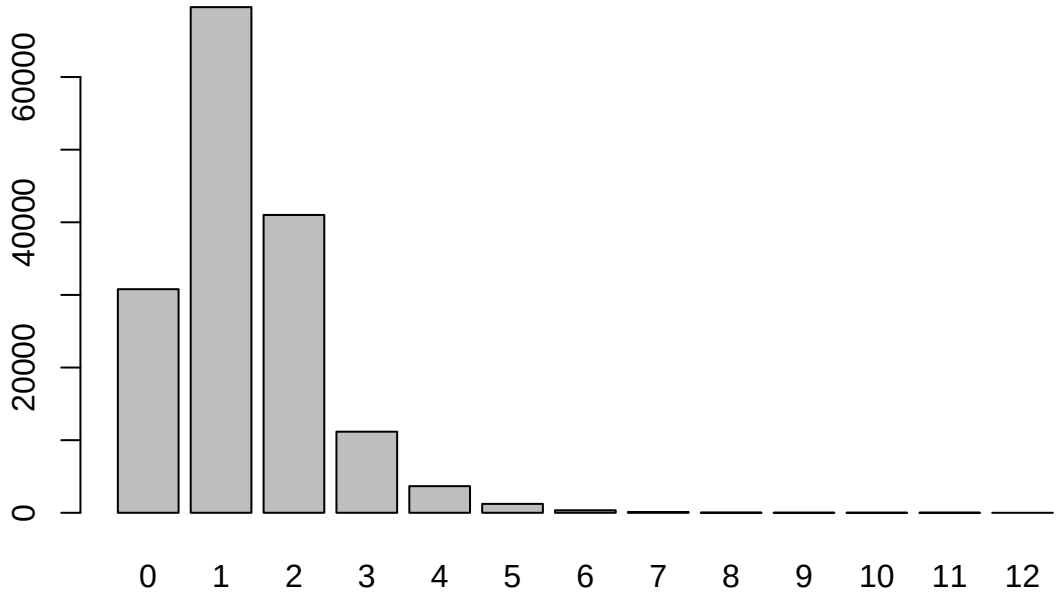
```
##     V1               V2                        V3
##  남자:226965   Min.   : 0.00   가구주           :179293
##  여자:241319   1st Qu.:22.00   자녀             :145533
##                Median :40.00   가구주의 배우자 :106311
##                Mean   :39.34   가구주의 부모    : 12069
##                3rd Qu.:55.00   손자녀및배우자   :  7832
##                Max.   :85.00   형제자매및배우자:  5332
```

```
##                                (Other)         : 11914
##              V4                V5
##  고등학교      :134246   Min.   : 0.00
##  대학4년제이상: 81110   1st Qu.: 1.00
##  초등학교      : 80710   Median : 1.00
##  중학교        : 55704   Mean   : 1.32
##  안받음        : 51085   3rd Qu.: 2.00
##  대학4년제미만: 50753   Max.   :12.00
##  (Other)       : 14676   NA's   :310308
```

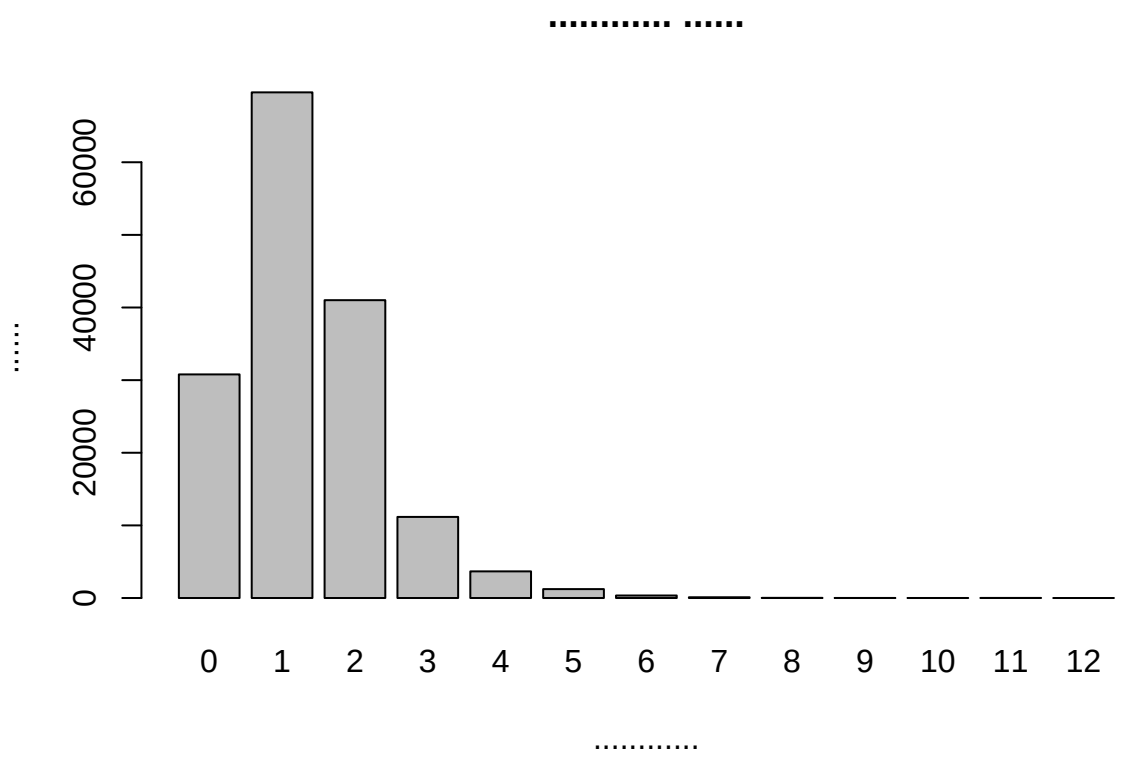
```
table(data$V5)
```

```
##
##     0     1     2     3     4     5     6     7     8     9    10    11
## 30788 69624 41010 11165  3667  1228   346   104    21     8     4    10
##    12
##     1
```

```
table.V5=table(data$V5)# 자녀 수의 분포
barplot(table.V5)
```



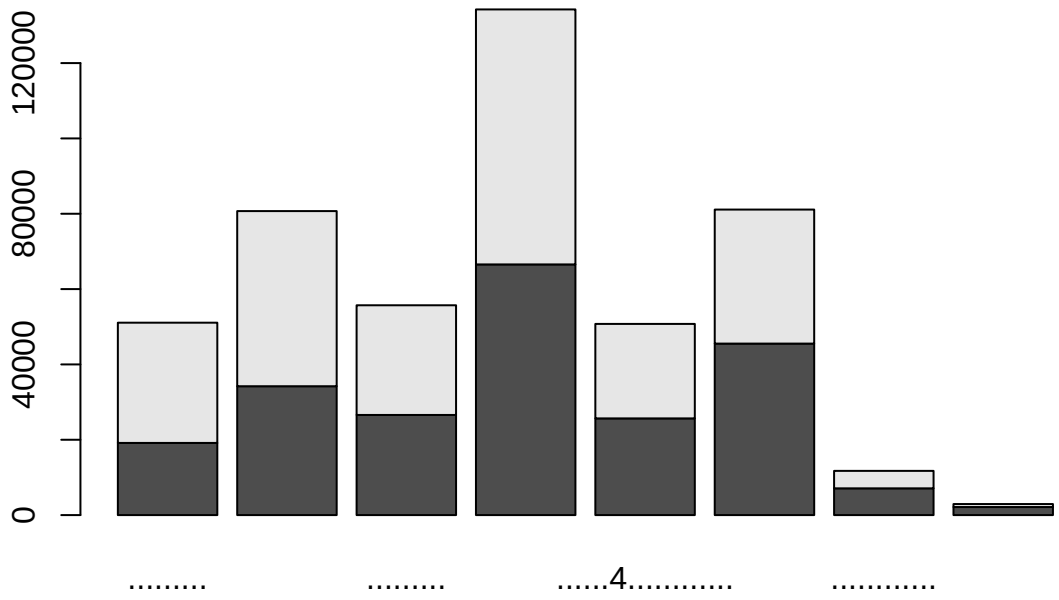
```r
barplot(table.V5, main="출생아별 빈도", xlab="출생아수",ylab="빈도")
```



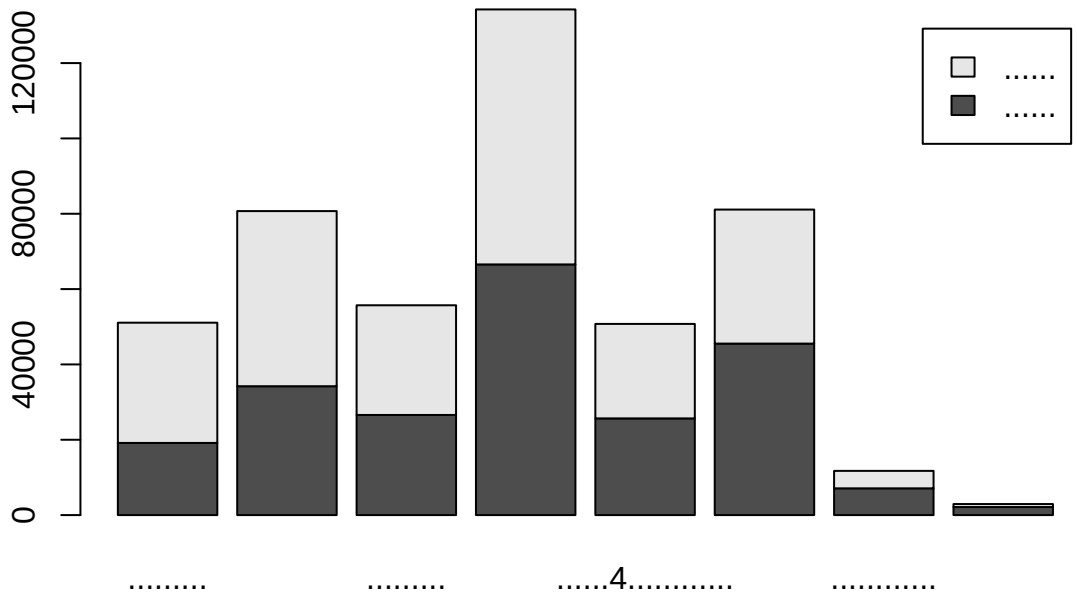

```r
table(data$V1,data$V4)#남녀-교육수준 분포
```

```
##
##          안받음 초등학교 중학교 고등학교 대학4년제미만 대학4년제이상
##    남자  19161     34214   26588     66548          25673          45530
##    여자  31924     46496   29116     67698          25080          35580
##
##          석사과정 박사과정
##    남자     7107     2144
##    여자     4634      791
```

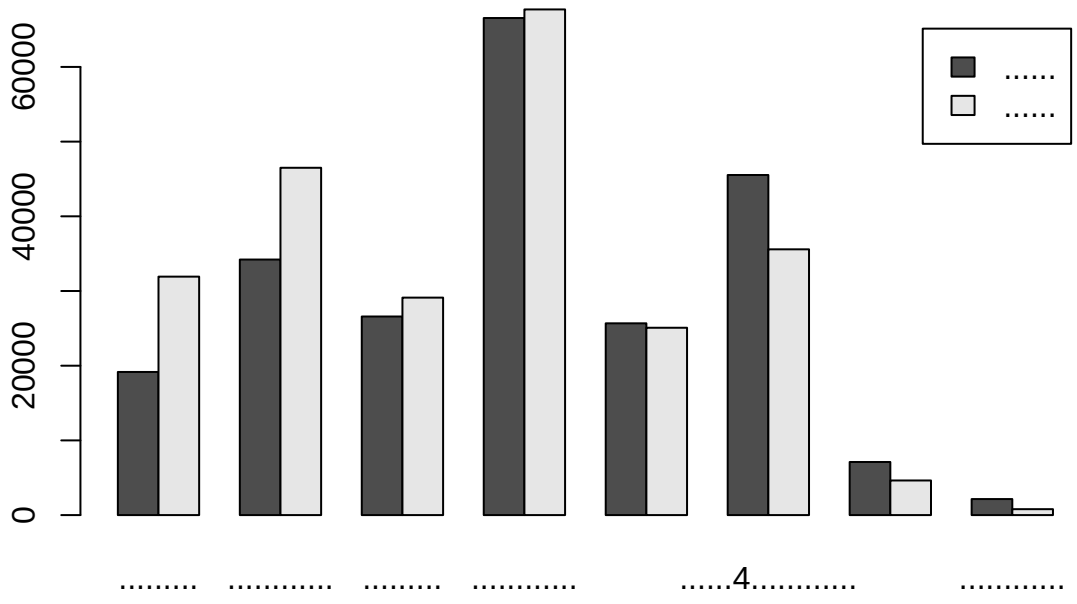
```r
tableV1.V4=table(data$V1,data$V4)
barplot(tableV1.V4)# 내가 보기용. 어? 뭔가 부족하네?
```



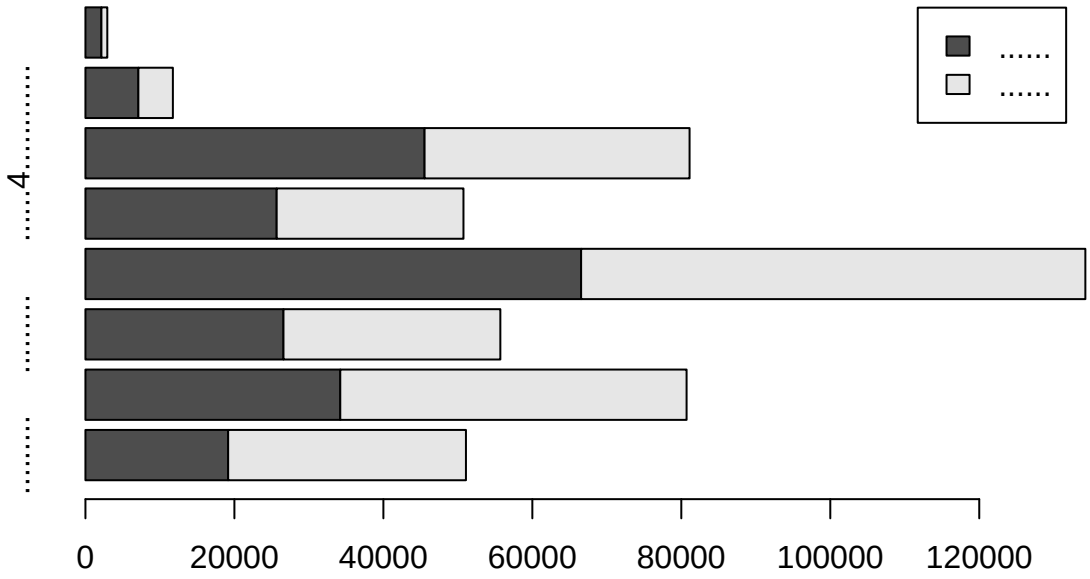

```
barplot(tableV1.V4, legend.text=T)#내가 보기용
```



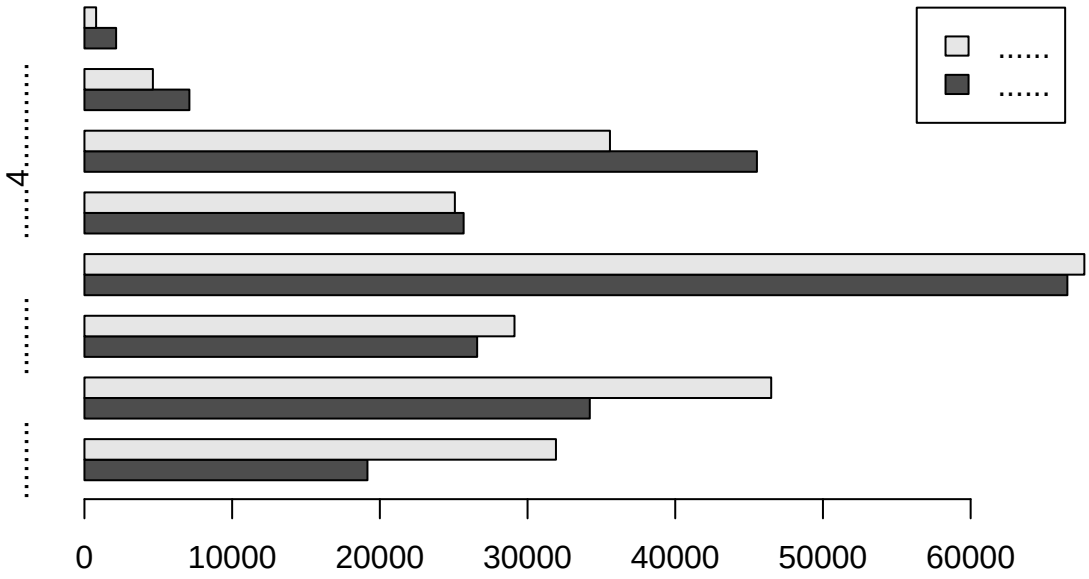

```
barplot(tableV1.V4, legend.text=T,beside=T)#나란히 보기
```



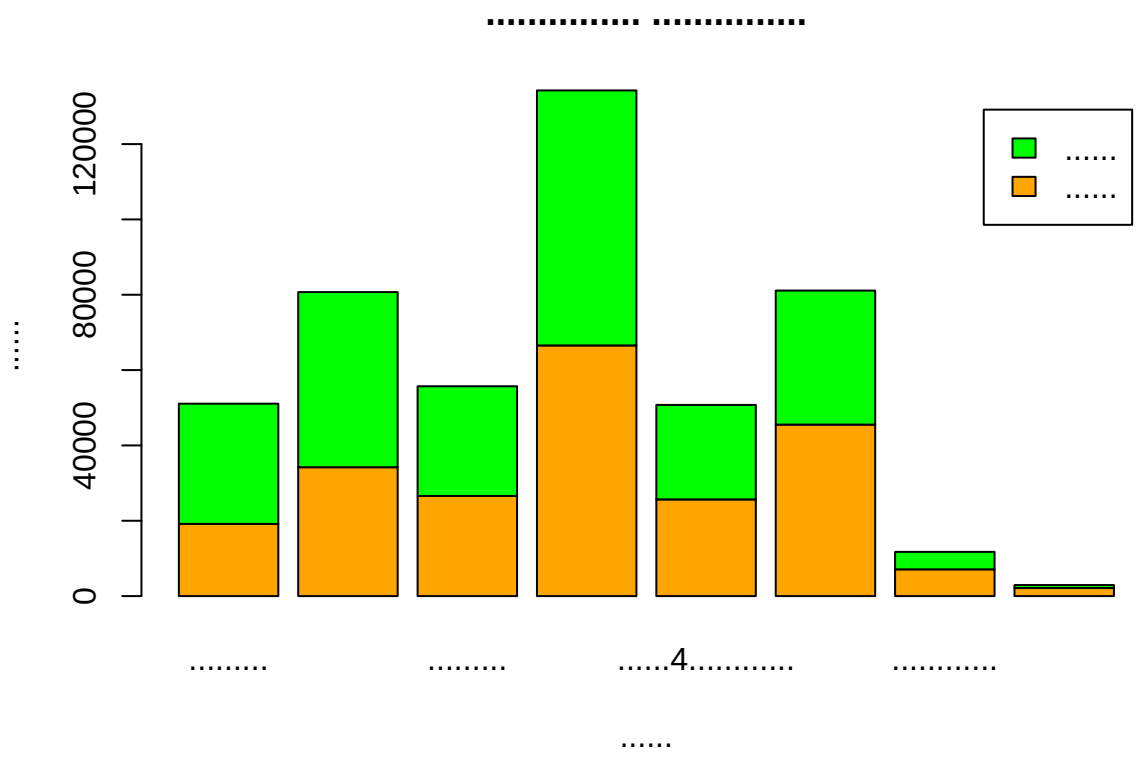

```
barplot(tableV1.V4, legend.text=T,horiz=T)#옆으로 뻗는 막대그래프
```



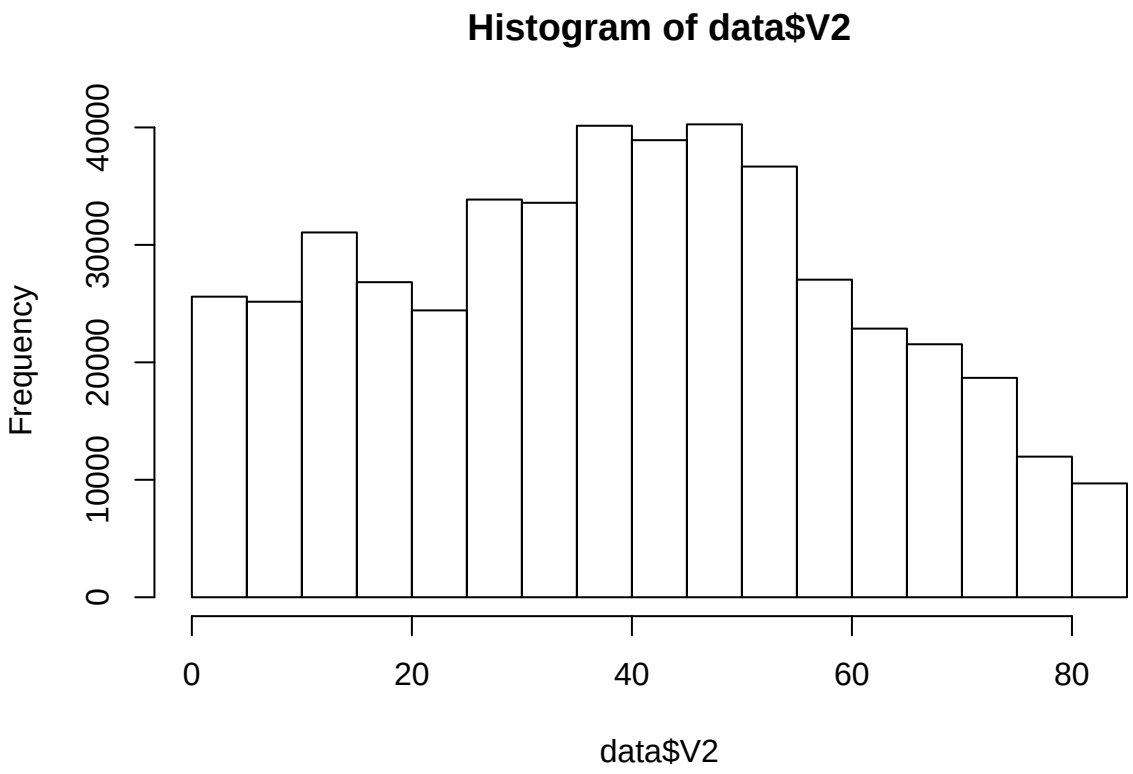

```
barplot(tableV1.V4, legend.text=T,horiz=T,beside=T)#옆으로 뻗는 막대그래프
```



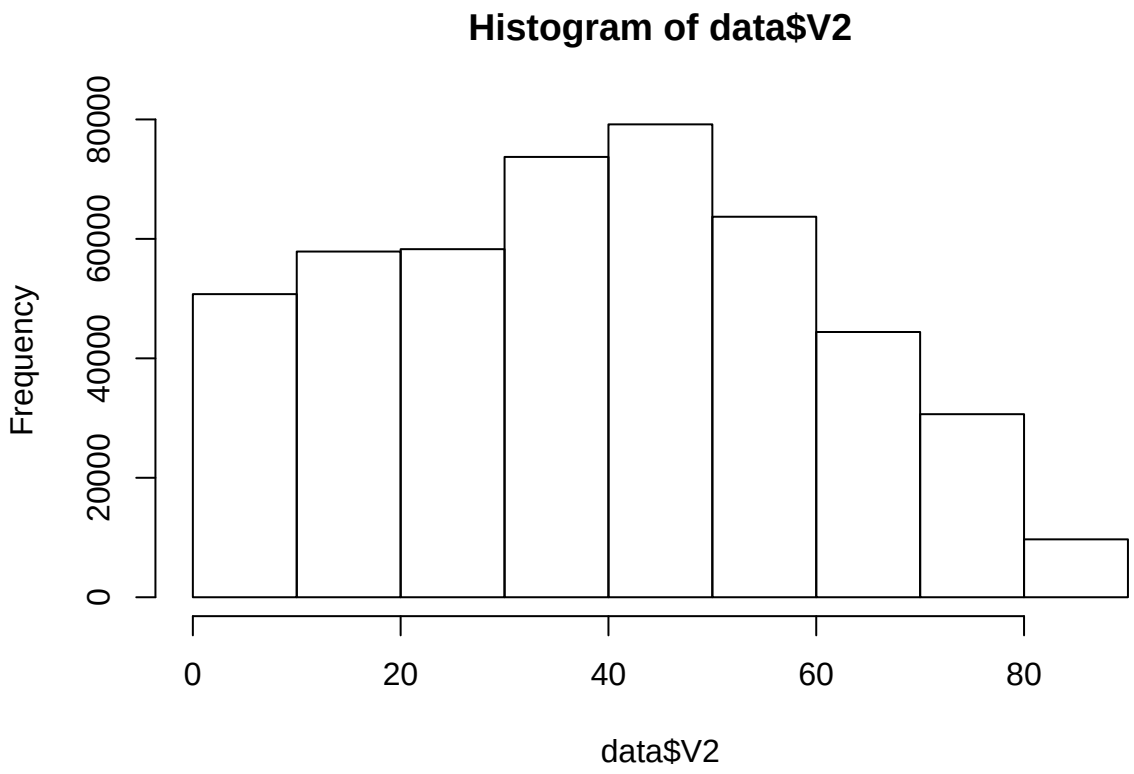

```
#남에게 보여주기용
barplot(tableV1.V4, legend.text=T, col=c("orange","green"), main="학력에따른 성별인원수",xlab="학력",ylab=
```



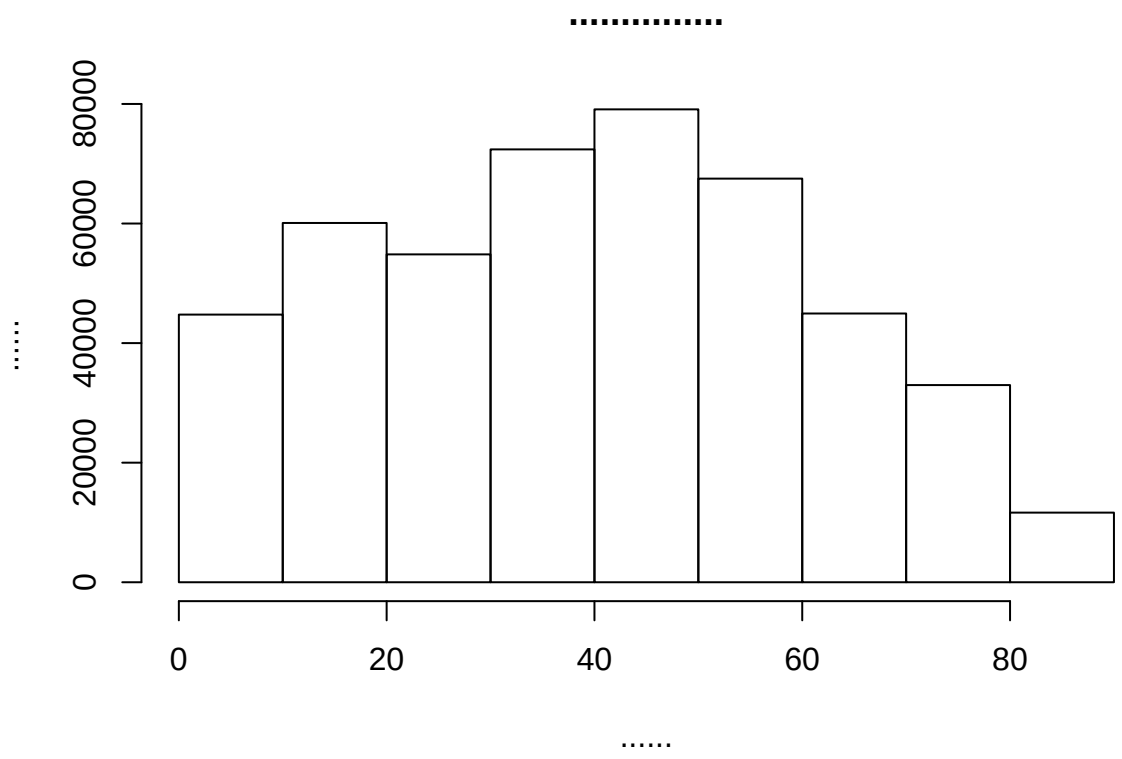

```
hist(data$V2)#나이의 분포
```




```
hist(data$V2, breaks=c(seq(0,90,10)))#나이의 분포
```




```r
#보고용
hist(data$V2, breaks=c(seq(0,90,10)), right=F,main="연령별분포",xlab="연령", ylab="빈도")
```

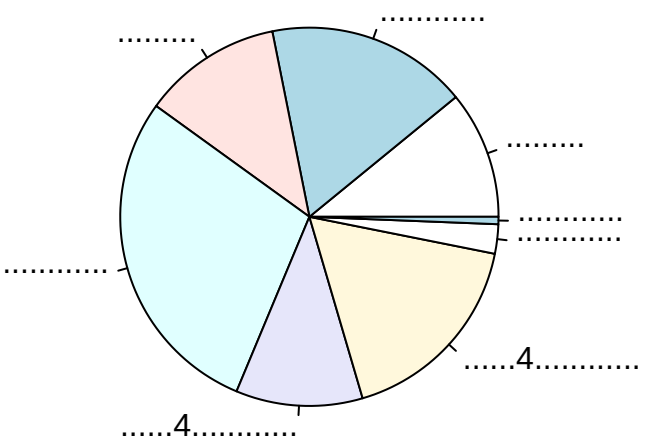


- 파이그래프: pie()

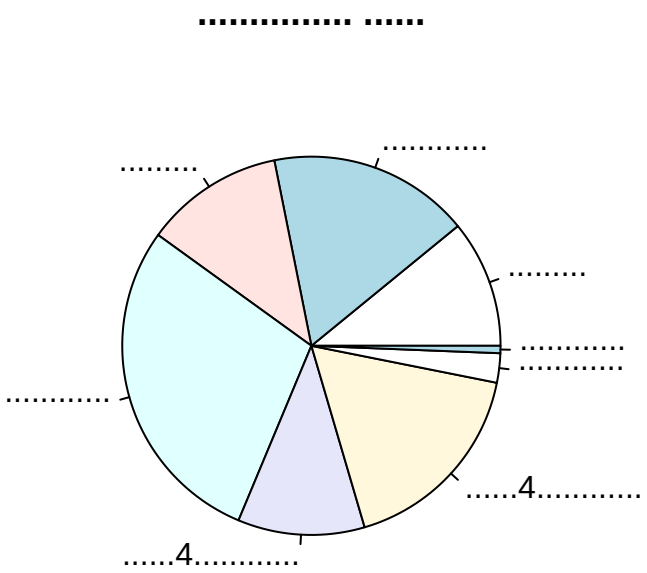
```
table(data$V4)
```

```
##
##          안받음        초등학교          중학교       고등학교  대학4년제미만
##           51085           80710           55704          134246           50753
## 대학4년제이상        석사과정        박사과정
##           81110           11741            2935
```

```
table.V4=table(data$V4)
pie(table.V4)#내가 보가
```




```
pie(table.V4, main="학력수준별 비중")
```




```
jpeg(filename="pie.jpg")
pie(table.V4, main="학력수준별 비중")
dev.off()
```

```
## pdf
##   2
```

2.모수와 통계량(parameter, statistics) - rara의 Cafe 자료

```
#ranicafe=read.csv("C:/work_2016/Rstat/Lesson 2/data/cafedata.csv",header=T,na.strings=c("na"))
ranicafe=read.csv("../data/cafedata.csv",header=T,na.strings=c("na"))
str(ranicafe)
```

```
## 'data.frame':    48 obs. of  22 variables:
##  $ t                         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Date                      : Factor w/ 48 levels "2010-01-19","2010-01-20",..: 1 2 3 4 5 6 7 8 9 10
##  $ Day.Code                  : int  2 3 4 5 1 2 3 4 5 1 ...
##  $ Day.of.Week               : Factor w/ 5 levels "Fri","Mon","Thu",..: 4 5 3 1 2 4 5 3 1 2 ...
##  $ Bread.Sand.Sold           : int  5 6 8 4 3 7 6 0 3 2 ...
##  $ Bread.Sand.Waste          : int  3 8 2 2 0 1 6 0 4 6 ...
##  $ Wraps.Sold                : int  25 7 14 5 10 5 19 7 4 13 ...
```

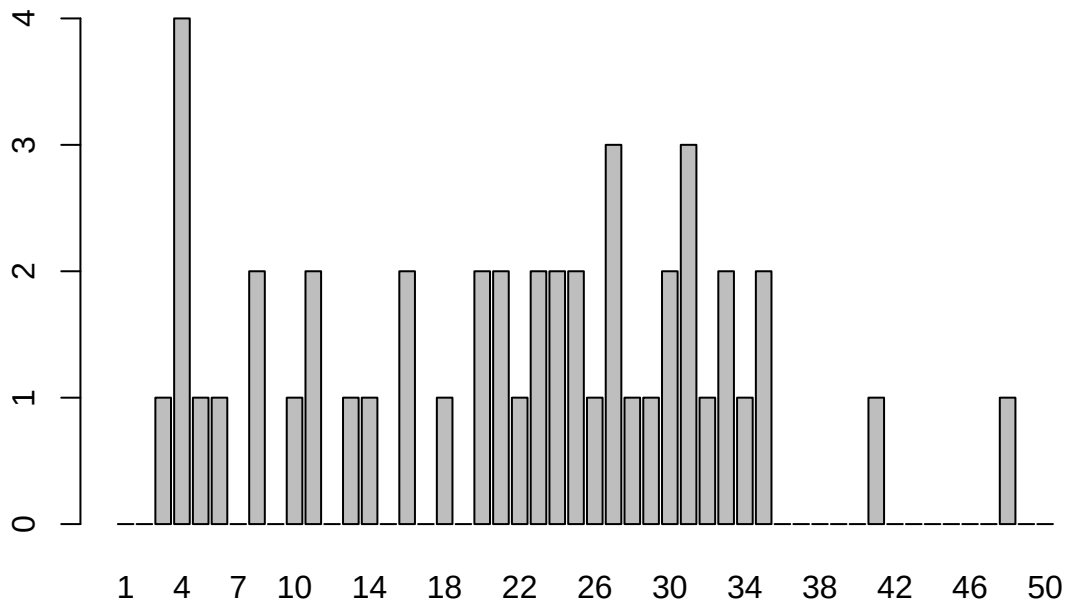
```
##  $ Wraps.Waste              : int  5 17 0 7 0 3 3 0 9 3 ...
##  $ Muffins.Sold             : int  5 3 4 5 8 1 6 6 0 3 ...
##  $ Muffins.Waste            : int  1 5 0 0 0 0 0 1 4 0 ...
##  $ Cookies.Sold             : int  5 1 1 3 3 5 10 0 3 6 ...
##  $ Cookies.Waste            : int  3 6 0 1 0 0 0 0 2 0 ...
##  $ Fruit.Cup.Sold           : int  1 0 0 3 2 2 2 0 1 2 ...
##  $ Fruit.Cup.Waste          : int  4 3 3 0 0 0 0 0 1 0 ...
##  $ Chips                    : int  12 0 0 20 0 4 2 20 3 16 ...
##  $ Juices                   : int  8 0 13 0 5 4 5 6 4 7 ...
##  $ Sodas                    : int  20 13 23 13 13 33 15 27 12 19 ...
##  $ Coffees                  : int  41 33 34 27 20 23 32 31 30 27 ...
##  $ Total.Soda.and.Coffee    : int  61 46 57 40 33 56 47 58 42 46 ...
##  $ Sales                    : num  200 196 103 163 102 ...
##  $ Max.Daily.Temperature..F.: int  36 34 39 40 36 26 34 33 20 37 ...
##  $ Total.Items.Wasted       : int  16 39 5 10 0 4 9 1 20 9 ...
```

```
summary(ranicafe)
```

```
##        t                 Date       Day.Code Day.of.Week Bread.Sand.Sold
##  Min.   : 1.00   2010-01-19: 1   Min.   :1   Fri: 9      Min.   :0.000
##  1st Qu.:12.75   2010-01-20: 1   1st Qu.:2   Mon: 9      1st Qu.:3.000
##  Median :24.50   2010-01-21: 1   Median :3   Thu:10      Median :4.000
##  Mean   :24.50   2010-01-22: 1   Mean   :3   Tue:10      Mean   :4.702
##  3rd Qu.:36.25   2010-01-25: 1   3rd Qu.:4   Wed:10      3rd Qu.:6.000
##  Max.   :48.00   2010-01-26: 1   Max.   :5               Max.   :9.000
##                  (Other)   :42                           NA's   :1
##  Bread.Sand.Waste   Wraps.Sold     Wraps.Waste     Muffins.Sold
##  Min.   :0.000    Min.   : 4.00   Min.   : 0.00   Min.   : 0.000
##  1st Qu.:0.000    1st Qu.: 9.00   1st Qu.: 0.00   1st Qu.: 3.000
##  Median :0.000    Median :13.00   Median : 0.00   Median : 5.000
##  Mean   :1.574    Mean   :13.15   Mean   : 1.66   Mean   : 5.851
##  3rd Qu.:3.000    3rd Qu.:16.50   3rd Qu.: 2.00   3rd Qu.: 8.000
##  Max.   :8.000    Max.   :25.00   Max.   :17.00   Max.   :28.000
##  NA's   :1        NA's   :1       NA's   :1       NA's   :1
##  Muffins.Waste    Cookies.Sold    Cookies.Waste   Fruit.Cup.Sold
##  Min.   :0.000   Min.   : 0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:0.000   1st Qu.: 3.000   1st Qu.:0.000   1st Qu.:1.000
##  Median :0.000   Median : 5.000   Median :0.000   Median :2.000
##  Mean   :0.617   Mean   : 5.787   Mean   :1.043   Mean   :1.702
##  3rd Qu.:1.000   3rd Qu.: 8.000   3rd Qu.:1.500   3rd Qu.:2.000
```

```
##  Max.   :5.000   Max.   :13.000   Max.   :6.000   Max.   :4.000
##  NA's   :1       NA's   :1        NA's   :1       NA's   :1
##  Fruit.Cup.Waste      Chips            Juices           Sodas
##  Min.   :0.0000   Min.   : 0.000   Min.   : 0.000   Min.   :11.00
##  1st Qu.:0.0000   1st Qu.: 6.500   1st Qu.: 3.000   1st Qu.:21.00
##  Median :0.0000   Median : 9.000   Median : 4.000   Median :29.00
##  Mean   :0.3617   Mean   : 9.149   Mean   : 4.936   Mean   :29.57
##  3rd Qu.:0.0000   3rd Qu.:11.000   3rd Qu.: 6.000   3rd Qu.:36.00
##  Max.   :4.0000   Max.   :25.000   Max.   :21.000   Max.   :55.00
##  NA's   :1        NA's   :1        NA's   :1        NA's   :1
##     Coffees      Total.Soda.and.Coffee     Sales
##  Min.   : 3.00   Min.   :27.00         Min.   : 61.94
##  1st Qu.:12.00   1st Qu.:41.00         1st Qu.:119.88
##  Median :23.00   Median :52.00         Median :150.51
##  Mean   :21.51   Mean   :51.09         Mean   :148.22
##  3rd Qu.:30.00   3rd Qu.:60.00         3rd Qu.:179.02
##  Max.   :48.00   Max.   :74.00         Max.   :240.87
##  NA's   :1       NA's   :1             NA's   :1
##  Max.Daily.Temperature..F. Total.Items.Wasted
##  Min.   :20.00             Min.   : 0.000
##  1st Qu.:33.00             1st Qu.: 1.000
##  Median :37.50             Median : 4.000
##  Mean   :41.92             Mean   : 5.255
##  3rd Qu.:48.25             3rd Qu.: 7.000
##  Max.   :80.00             Max.   :39.000
##                            NA's   :1
```

```r
table.coffee=table(ranicafe$Coffees)
coffe.ext=rep(0,50)
coffe.ext[as.numeric(names(table.coffee))]=table.coffee
names(coffe.ext)=(1:50)
barplot(coffe.ext)
```



- 표본통계량1: 자료의 범위(max min)

```
ranicafe$Coffee
```

```
##  [1] 41 33 34 27 20 23 32 31 30 27 30 27 26 24 18 22 21 28 23 31 29 48 25
## [24] 31 25 35 33 35 16 24 20 11 21 NA  8  8  4  4  3  5  6  4 13  4 16 14
## [47] 10 11
```

```
#max
sort(ranicafe$Coffee)
```

```
##  [1]  3  4  4  4  4  5  6  8  8 10 11 11 13 14 16 16 18 20 20 21 21 22 23
## [24] 23 24 24 25 25 26 27 27 27 28 29 30 30 31 31 31 32 33 33 34 35 35 41
## [47] 48
```

```
sort(ranicafe$Coffee)[1]
```

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

```
min(ranicafe$Coffee,na.rm=T)
```

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

```r
#min
sort(ranicafe$Coffee,decreasing=T)
```

```
##  [1] 48 41 35 35 34 33 33 32 31 31 31 30 30 29 28 27 27 27 26 25 25 24 24
## [24] 23 23 22 21 21 20 20 18 16 16 14 13 11 11 10  8  8  6  5  4  4  4  4
## [47]  3
```

```r
sort(ranicafe$Coffee,decreasing=T)[1]
```

```
## [1] 48
```

```r
max(ranicafe$Coffee,na.rm=T)
```

```
## [1] 48
```

```r
range(ranicafe$Coffee,na.rm=T)
```

```
## [1]  3 48
```

- 표본통계량2: 대표값(최빈값,평균,중앙값)

```r
#최빈값(mode)
rc=ranicafe$Coffees
stem(rc)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   0 | 34444
##   0 | 5688
##   1 | 01134
##   1 | 668
##   2 | 001123344
##   2 | 55677789
##   3 | 001112334
##   3 | 55
##   4 | 1
##   4 | 8
```

```r
sort(table(ranicafe$Coffees),decreasing=T)[1]
```

```
## 4
## 4
```

```r
names(sort(table(ranicafe$Coffees),decreasing=T)[1])
```

```
## [1] "4"
```

```
as.numeric(names(sort(table(ranicafe$Coffees),decreasing=T)[1]))
```

```
## [1] 4
```

```
#평균
## 평균 가중치
weight=1/length(rc)
sum(rc*weight,na.rm=T)
```

```
## [1] 21.0625
```

```
weight2=1/length(rc[!is.na(rc)]) # Size without mission obs
sum(rc*weight2,na.rm=T)
```

```
## [1] 21.51064
```

```
## 평균
mean(rc)
```

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

```
mean(rc,na.rm=T)
```

```
## [1] 21.51064
```

```
rc=c(rc,NA)
tail(rc,n=5)
```

```
## [1] 16 14 10 11 NA
```

```
mean(rc)
```

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

```
mean(rc,na.rm=T)
```

```
## [1] 21.51064
```

```
## 평균은 자주 보이지 않을 수 있다.
rc[(rc==21)|(rc==22)] # cannot determine, then we include
```

```
## [1] 22 21 21 NA NA
```

```
rc[which(rc==21|rc==22)] # we include wha we can determine. Things cannot be determined our out.
```

```
## [1] 22 21 21
```

```
ranicafe[(rc==21)|(rc==22),]
```

```
##       t       Date Day.Code Day.of.Week Bread.Sand.Sold Bread.Sand.Waste
## 16   16 2010-02-09        2         Tue               8                0
## 17   17 2010-02-10        3         Wed               7                0
## 33   33 2010-03-04        4         Thu               4                0
## NA   NA       <NA>       NA        <NA>              NA               NA
## NA.1 NA       <NA>       NA        <NA>              NA               NA
##      Wraps.Sold Wraps.Waste Muffins.Sold Muffins.Waste Cookies.Sold
## 16           16           0           11             0            9
## 17           12           0            5             0            7
## 33           14           0            6             0            8
## NA           NA          NA           NA            NA           NA
## NA.1         NA          NA           NA            NA           NA
##      Cookies.Waste Fruit.Cup.Sold Fruit.Cup.Waste Chips Juices Sodas
## 16               1              2               0    11      8    31
## 17               3              1               0    14      3    24
## 33               0              2               0     8      5    43
## NA              NA             NA              NA    NA     NA    NA
## NA.1            NA             NA              NA    NA     NA    NA
##      Coffees Total.Soda.and.Coffee  Sales Max.Daily.Temperature..F.
## 16        22                   53 181.43                       29
## 17        21                   45 125.57                       26
## 33        21                   64 168.08                       45
## NA        NA                   NA     NA                       NA
## NA.1      NA                   NA     NA                       NA
##      Total.Items.Wasted
## 16                    1
## 17                    3
## 33                    0
## NA                   NA
## NA.1                 NA
```

```
ranicafe[which((rc==21)|(rc==22)),]
```

```
##     t       Date Day.Code Day.of.Week Bread.Sand.Sold Bread.Sand.Waste
## 16 16 2010-02-09        2         Tue               8                0
## 17 17 2010-02-10        3         Wed               7                0
## 33 33 2010-03-04        4         Thu               4                0
##    Wraps.Sold Wraps.Waste Muffins.Sold Muffins.Waste Cookies.Sold
```

```
## 16         16           0           11             0           9
## 17         12           0            5             0           7
## 33         14           0            6             0           8
##    Cookies.Waste Fruit.Cup.Sold Fruit.Cup.Waste Chips Juices Sodas Coffees
## 16             1              2               0    11      8    31      22
## 17             3              1               0    14      3    24      21
## 33             0              2               0     8      5    43      21
##    Total.Soda.and.Coffee  Sales Max.Daily.Temperature..F.
## 16                    53 181.43                        29
## 17                    45 125.57                        26
## 33                    64 168.08                        45
##    Total.Items.Wasted
## 16                  1
## 17                  3
## 33                  0
```

```
## 평균은 outlier에 민감하다.
rc[which.max(rc)]=480
mean(rc,na.rm=T)
```

```
## [1] 30.70213
```

```
## 중위수
rc_m=rc[!is.na(rc)]
median.idx=(length(rc_m)+1)/2
rc.s=sort(rc_m)
rc.s[median.idx]
```

```
## [1] 23
```

```
median(rc,na.rm=T)
```

```
## [1] 23
```

```
## 중위수는 outlier에 덜 민감하다.
rc[which.max(rc)]=48
median(rc,na.rm=T)
```

```
## [1] 23
```

- 표본통계량 3: 펴저있는 정도(표준편차 사분위수 범위)
  a. 표준편차 (Standard deviation)

```
## 편차
height=c(164, 166, 168, 170,172,174,176)
height.m=mean(height)
height.dev=height-mean(height)
sum(height.dev)
```

```
## [1] 0
```

```
# (더하면 0이더라)

## 분산
var(height.dev)
```

```
## [1] 18.66667
```

```
#? var
mean(height.dev^2) # 확률 1/6로 키가 가지는 값이 6개 밖에 없는 경우
```

```
## [1] 16
```

```
var(height.dev)*(length(height)-1)/length(height)
```

```
## [1] 16
```

```
##표준편차
sd(height)
```

```
## [1] 4.320494
```

```
##?sd
sqrt(mean(height.dev^2)) # 확률 1/6로 키가 가지는 값이 6개 밖에 없는 경우
```

```
## [1] 4
```

```
sd(height)*sqrt((length(height)-1)/length(height))
```

```
## [1] 4
```

```
## 커피 판매 평균, 표준편차
rc.m=mean(rc,na.rm=T)
rc.sd=sd(rc,na.rm=T)
cat("Coffee sales=", round(rc.m,1),"+/-",round(rc.sd,2))
```

```
## Coffee sales= 21.5 +/- 11.08
```

```
## 변동계수
```

```
rj=ranicafe$Juices
rj.m=mean(rj,na.rm=T)
rj.sd=sd(rj,na.rm=T)
rc.cv=rc.sd/rc.m
rj.cv=rj.sd/rj.m
rc.cv
```

```
## [1] 0.5151163
```

```
rj.cv
```

```
## [1] 0.7502046
```

```
# 쥬스 판매가 커피 판매보다 더 변화가 심하다.
```

b. 4분위수

```
# 최소값, 4분위수, 최대값
qs=quantile(rc,na.rm=T)
qs
```

```
##   0%  25%  50%  75% 100%
##    3   12   23   30   48
```

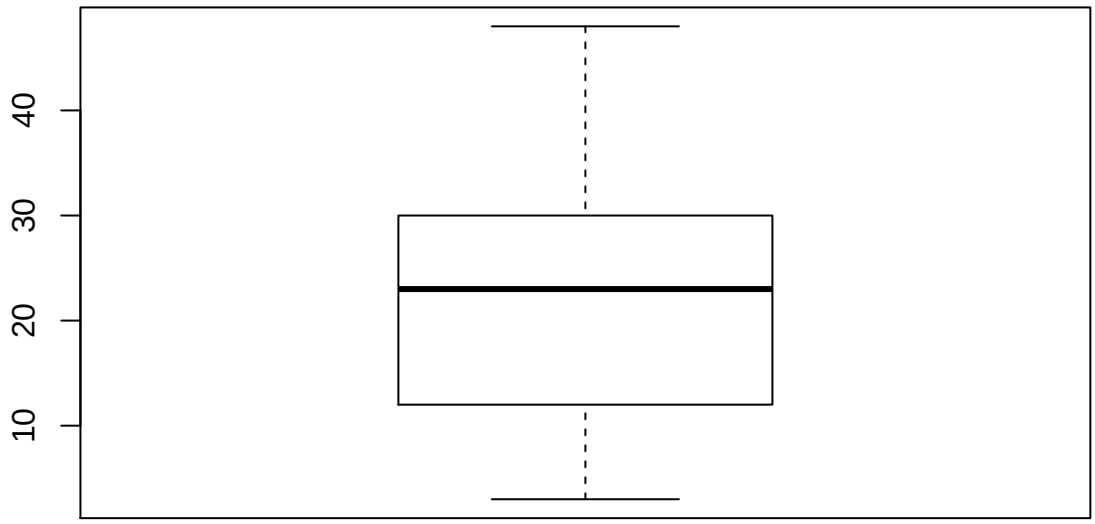
```
# 3분위수 -1분위수: 분위 간 범위
IQR(rc,na.rm=T)
```

```
## [1] 18
```

```
qs[4]-qs[2]
```

```
## 75%
##  18
```

```
bp.rc=boxplot(rc,na.rm=T)
```




```
bp.rc
```

```
## $stats
##      [,1]
## [1,]    3
## [2,]   12
## [3,]   23
## [4,]   30
## [5,]   48
##
## $n
## [1] 47
##
## $conf
##          [,1]
## [1,] 18.8516
## [2,] 27.1484
##
## $out
```

```
## numeric(0)
##
## $group
## numeric(0)
##
## $names
## [1] ""
```

```
## 자동차 제동거리 outlier
is.na(cars$dist)
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [34] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [45] FALSE FALSE FALSE FALSE FALSE FALSE
```

```
qscar=quantile(cars$dist)
qscar
```

```
##   0%  25%  50%  75% 100%
##    2   26   36   56  120
```

```
bp.dist=boxplot(cars$dist)
bp.dist
```

```
## $stats
##      [,1]
## [1,]    2
## [2,]   26
## [3,]   36
## [4,]   56
## [5,]   93
##
## $n
## [1] 50
##
## $conf
##          [,1]
## [1,] 29.29663
## [2,] 42.70337
##
```

```
## $out
## [1] 120
##
## $group
## [1] 1
##
## $names
## [1] "1"
```

```
# outlier 판별 기준: 하한값/상한값

ll.dist=qscar[2]-1.5*IQR(cars$dist,na.rm=T)
ul.dist=qscar[4]+1.5*IQR(cars$dist,na.rm=T)
dist.out=cars$dist[(cars$dist>ul.dist|cars$dist<ll.dist)&!is.na(cars$dist)]
dist.in=cars$dist[(cars$dist<=ul.dist)&(cars$dist>=ll.dist)&!is.na(cars$dist)]
whisker.dist=range(dist.in)
range(bp.dist$stats)
```

```
## [1]  2 93
```

```
dist.out
```

```
## [1] 120
```

```
ll.rc=qs[2]-1.5*IQR(rc,na.rm=T)
ul.rc=qs[4]+1.5*IQR(rc,na.rm=T)
rc.out=rc[(rc>ul.rc|rc<ll.rc)&!is.na(rc)]
rc.in=rc[(rc<=ul.rc)&(rc>=ll.rc)&!is.na(rc)]
whisker.rc=range(rc.in)
range(bp.rc$stats)
```

```
## [1]  3 48
```

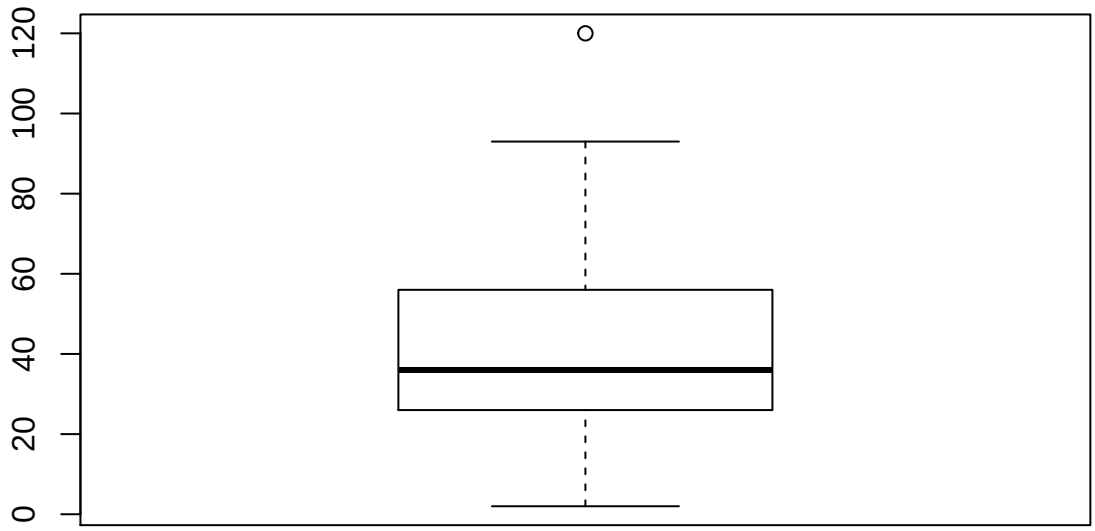
```
rc.out
```

```
## numeric(0)
```

```
## quantile. extended
quantile(rc,seq(0,1,0.1),na.rm=T)
```

```
##   0%  10%  20%  30%  40%  50%  60%  70%  80%  90% 100%
##  3.0  4.6 10.2 15.6 20.4 23.0 25.6 28.2 31.0 33.4 48.0
```

```
#install.packages("statar")
library(statar)
```




```
xtile(rc,n=4)
```

```
##  [1]  4  4  4  3  2  2  4  4  3  3  3  3  3  3  2  2  2  3  2  4  3  4  3
## [24]  4  3  4  4  4  2  3  2  1  2 NA  1  1  1  1  1  1  1  1  2  1  2  2
## [47]  1  1 NA
```

```
xtile(rc,prob=seq(0.25,1,0.25))
```

```
##  [1]  4  4  4  3  2  2  4  4  3  3  3  3  3  3  2  2  2  3  2  4  3  4  3
## [24]  4  3  4  4  4  2  3  2  1  2 NA  1  1  1  1  1  1  1  1  2  1  2  2
## [47]  1  1 NA
```

3. 다음 시간 준비

```
save.image("Lesson2.RData")
#install.packages("prob")
library("prob")
tosscoin(1)
```

```
##   toss1
## 1     H
## 2     T
```