

데이터과학을 위한 **R**프로그래밍

3주차. R 데이터구조(생성, 추출)



이혜선 교수

포항공과대학교 산업경영공학과



목차

3주차. R 데이터구조(생성, 추출)

1차시

R 데이터 생성 (불러들이기)

2차시

R 데이터 활용 I (subset, 내보내기)

3차시

R 데이터 활용 II (dplyr 활용)



3주차

2차시

R 데이터 활용 I

(subset, 내보내기)

데이터 추출

✓ 데이터 추출 - subset(데이터이름, 조건)

▶ 예제 1 : brain 데이터에서 female만 있는 subset 데이터 생성

```
1 # lec3_2.r
2 # Data handling
3 # subset and export
4
5 # set working directory
6 # change working directory
7 setwd("D:/tempstore/moocr")
8
9 brain<-read.csv("brain2210.csv")
10 head(brain)
11
12 attach(brain)
13
14 # subset with female
15 # brainf<-subset(brain, sex=='f') after attach(brain)
16 brainf<-subset(brain, sex=='f')
17 mean(brainf$wt)
```

brain 데이터에서 female(여성)만 추출하여
brainf로 저장

> mean(brainf\$wt)
[1] 1117.169

데이터 추출

☑ 데이터 추출 - subset(데이터이름, 조건)

▶ 예제 2 : brain 데이터에서 male만 있는 subset 데이터 생성

```
# subset with female
# brainf<-subset(brain, sex=='f')
brainf<-subset(brain, sex=='f')
mean(brainf$wt)
sd(brainf$wt)

# subset with male
brainm<-subset(brain, sex=='m')
mean(brainm$wt)
sd(brainm$wt)
```



```
> brainf<-subset(brain, sex=='f')
> mean(brainf$wt)
[1] 1117.169
> sd(brainf$wt)
[1] 98.97094
>
> # subset with male
> brainm<-subset(brain, sex=='m')
> mean(brainm$wt)
[1] 1270.741
> sd(brainm$wt)
[1] 129.22
```

▶ 요약표

| | female | male |
|------|---------|---------|
| mean | 1117.17 | 1270.74 |
| sd | 98.97 | 129.22 |

데이터 추출과 활용

✓ 추출한 데이터의 활용 (그룹별 히스토그램) - 눈여겨보기

```
# histogram with same scale  
hist(brainf$wt, breaks = 12, col = "green", cex=0.7, main="Histogram (Female)")  
hist(brainm$wt, breaks = 12, col = "orange", cex=0.7, main="Histogram (Male)")
```

또 다른
문제는?

남녀간 동일범위로 설정하고
히스토그램을 그리고 비교한다!!



데이터 추출

✓ 데이터 추출 - subset(데이터이름, 조건)

▶ 예제 3 : brain 데이터에서 wt<1300 이하인 데이터 생성

```
# subset with wt<1300  
brain1300<-subset(brain,wt<1300)  
  
# same subset of brain1300  
# brain1300<-subset(brain,!brain$wt>=1300)  
summary(brain1300)  
table(brain1300$sex)
```

```
> summary(brain1300)  
      wt      sex  
Min.   : 915   Length:138  
1st Qu.:1074   Class  :character  
Median :1155   Mode   :character  
Mean    :1145  
3rd Qu.:1230  
Max.    :1289
```

```
> table(brain1300$sex)  
  f  m  
74 64
```

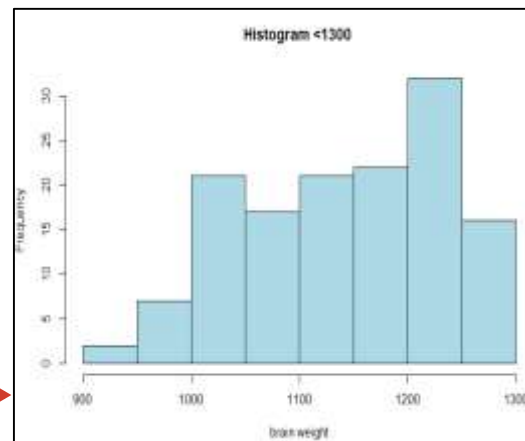
원래 데이터 brain은
여(77), 남(108)명

데이터 추출과 활용

✓ 추출한 데이터의 활용 (히스토그램)

➤ 예제 3 : brain 데이터에서 wt<1300 이하인 데이터 생성

```
# ex3. subset with wt<1300  
brain1300<-subset(brain,wt<1300)  
  
# same subset of brain1300  
# brain1300<-subset(brain,!brain$wt>=1300)  
summary(brain1300)  
table(brain1300$sex)  
  
hist(brain1300$wt, breaks = 12,col = "lightblue", main="Histogram
```

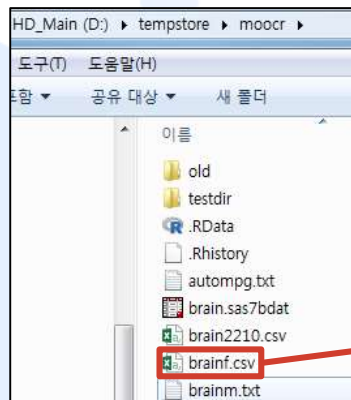


데이터 내보내기

✓ csv로 내보내기 (write.table, write.csv)

```
# export csv file - write out to csv file  
write.table(brainf, file="brainf.csv", row.names = FALSE, sep=",", na=" ")
```

```
write.csv(brainf, file="brainf.csv", row.names = FALSE)
```



A screenshot of an Excel spreadsheet titled 'brainf.csv - Excel'. The spreadsheet displays data with columns 'wt' and 'sex'. The data is as follows:

| | wt | sex |
|----|------|-----|
| 1 | | |
| 2 | 1125 | f |
| 3 | 1027 | f |
| 4 | 1112 | f |
| 5 | 983 | f |
| 6 | 1090 | f |
| 7 | 1247 | f |
| 8 | 1045 | f |
| 9 | 983 | f |
| 10 | 972 | f |
| 11 | 1045 | f |

데이터 내보내기

☑ txt 로 내보내기 (write.table)

```
# export txt file - write out to csv file  
write.table(brainm, file="brainm.txt", row.names = FALSE, na=" ")
```

