
통계로 보는 역사학

| 목차

- 데이터 탐색하기

- 기존 데이터를 활용하여 데이터 생성하기

데이터 탐색하기

데이터 탐색하기

- 메모리 및 데이터 변수 초기화하기

```
> gc()
```

```
      used (Mb) gc trigger  (Mb) limit (Mb) max used  (Mb)
Ncells 1443759 77.2    2540470 135.7      NA  2540470 135.7
Vcells 3142129 24.0    8388608  64.0    32768 6331978  48.4
```

```
> rm(list=ls())
```

- 한글 출력을 위해 인코딩 설정하기

```
> #Encoding_mac
```

```
> Sys.setlocale(category = "LC_CTYPE", locale = "ko_KR.UTF-8")
```

```
[1] "ko_KR.UTF-8"
```

```
> #Encoding_window
```

```
> options(encoding = 'UTF-8') #euc-kr
```

데이터 탐색하기

- 경로 지정하기

```
> getwd()
[1] "/Users/seongminmun/Desktop/Class/Genealogy_Ann"
> setwd("/Users/seongminmun/Desktop/Class/Genealogy_Ann")
> dir()
[1] "Ann_revised_300.csv" "original"
```

- CSV데이터 불러오기

```
> Ann <- read.csv("Ann_revised_300.csv",header = TRUE)
> str(Ann)
'data.frame':  300 obs. of  8 variables:
 $ X          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ number     : int  1 2 3 4 5 6 7 8 9 10 ...
 $ last_name  : Factor w/ 39 levels "강","고","곽",...: 4 4 4 4 4 4 4 4 4 4 ...
 $ first_name : Factor w/ 249 levels "가","가균","가봉",...: 221 160 199 36 165 88 53 152 185 208 ...
 $ string     : Factor w/ 300 levels "a 1","a 1 1",...: 1 2 3 4 5 6 7 8 9 288 ...
 $ sex       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ grade     : int  10 6 1 1 1 1 1 1 1 1 ...
 $ final_period: int  1 1 2 2 3 3 4 5 6 6 ...
```

데이터 탐색하기

- 불 필요한 데이터 제거하기

```
> Ann_2<-Ann[,-1]
```

```
> head(Ann_2)
```

	number	last_name	first_name	string	sex	grade	final_period
1	1	권	행	a 1	0	10	1
2	2	권	인행	a 1 1	0	6	1
3	3	권	책	a 1 1 1	0	1	2
4	4	권	균한	a 1 1 1 1	0	1	2
5	5	권	자팽	a 1 1 1 1 1	0	1	3
6	6	권	선개	a 1 1 1 1 1 1	0	1	3

```
> str(Ann_2)
```

```
'data.frame':  300 obs. of  7 variables:
 $ number      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ last_name   : Factor w/ 39 levels "강","고","곽",...: 4 4 4 4 4 4 4 4 4 4 ...
 $ first_name  : Factor w/ 249 levels "가","가균","가봉",...: 221 160 199 36 165 88 53 152 185 208 ...
 $ string      : Factor w/ 300 levels "a 1","a 1 1",...: 1 2 3 4 5 6 7 8 9 288 ...
 $ sex         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ grade       : int  10 6 1 1 1 1 1 1 1 1 ...
 $ final_period: int  1 1 2 2 3 3 4 5 6 6 ...
```

데이터 탐색하기

- 데이터에서 성이 '권'씨인 사람들만 추출하여 데이터 재생성하기

```
kwon <- NULL
for(i in 1:length(Ann_2$last_name)){
  if(Ann_2[i,2]=="권"){
    kwon <- rbind(kwon,Ann_2[i,])
  }
}
head(kwon)
str(kwon)
summary(kwon)
```

- 데이터에서 성이 '권'씨가 아닌 사람들만 추출하여 데이터 재생성하기

```
kwon_not <- NULL
for(i in 1:length(Ann_2$last_name)){
  if(Ann_2[i,2]!="권"){
    kwon_not <- rbind(kwon_not,Ann_2[i,])
  }
}
head(kwon_not)
str(kwon_not)
summary(kwon_not)
```

데이터 탐색하기

- 성이 '권'씨인 사람들의 계급 평균 구하기

```
total_grade <- 0
people <- length(kwon$last_name)
for(i in 1:length(kwon$last_name)){
  total_grade <- total_grade+kwon[i,6]
}
mean_grade <- total_grade/people ; mean_grade
```

- 계급 데이터 변환하기

(계급이 1-5인 경우 1, 6-9인 경우 2, 10-14인 경우 3으로 표기하기)

```
Grade_change <- NULL
for(i in 1:length(Ann_2$grade)){
  if(1<=Ann_2[i,6]&&Ann_2[i,6]<=5){
    Grade_change <- c(Grade_change,1)
  } else if(6<=Ann_2[i,6]&&Ann_2[i,6]<=9){
    Grade_change <- c(Grade_change,2)
  } else if(10<=Ann_2[i,6]&&Ann_2[i,6]<=14){
    Grade_change <- c(Grade_change,3)
  }
}
head(Grade_change)
```


데이터 탐색하기

- 성이 '권'씨인 사람들의 계급 평균 구하기

```
total_grade <- 0
people <- length(kwon$last_name)
for(i in 1:length(kwon$last_name)){
  total_grade <- total_grade+kwon[i,6]
}
mean_grade <- total_grade/people ; mean_grade
```

데이터 탐색하기

- 적절한 함수를 사용하여 성이 '권'씨인 사람 추출하기

```
> ft_kwon <- subset(Ann_2,last_name=='권')
> head(ft_kwon)
  number last_name first_name      string sex grade final_period
1      1        권        행        a 1  0    10            1
2      2        권       인행      a 1 1  0     6            1
3      3        권        책      a 1 1 1  0     1            2
4      4        권       균한      a 1 1 1 1  0     1            2
5      5        권       자평      a 1 1 1 1 1  0     1            3
6      6        권       선개      a 1 1 1 1 1 1  0     1            3
> str(ft_kwon)
'data.frame':   108 obs. of  7 variables:
 $ number      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ last_name   : Factor w/ 39 levels "강","고","곽",...: 4 4 4 4 4 4 4 4 4 4 ...
 $ first_name  : Factor w/ 249 levels "가","가균","가봉",...: 221 160 199 36 165 88 53 152 185 208 ...
 $ string      : Factor w/ 300 levels "a 1","a 1 1",...: 1 2 3 4 5 6 7 8 9 288 ...
 $ sex         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ grade       : int  10 6 1 1 1 1 1 1 1 1 ...
 $ final_period: int  1 1 2 2 3 3 4 5 6 6 ...
```

- 적절한 함수를 사용하여 성이 '권'씨인 사람들의 계급 평균 구하기

```
> mean(ft_kwon$grade)
[1] 4.592593
```

기존 데이터를 활용하여 데이터 생성하기

기존 데이터를 활용하여 데이터 생성하기

- 직계 데이터 생성하기

(직계이면 0이고 직계가 아닌 경우(사위), 1로 표기하기)

- 계급 비교 데이터 생성하기

(자식의 계급이 아버지보다 크면 1, 작거나 같으면 0)