

R프로그래밍및실습 연습_3

Jinseog Kim
Dongguk University
jskim1986@gmail.com

2018-05-31

1 사망사고와 관련된 연도별 예제 데이터

1. 2015_사망사고.csv
2. 2016년_사망교통사고.csv

```
acc_2015 <- read.csv("2015_사망사고.csv", fileEncoding = "CP949")
```

2 2015년 자료에서 사망사고가 일어나는 요인을 (발생지시도, 주야, 도로형태, 법규위반, 요일)라고 할 때, 각 요인들의 조합에 따른 사망사고의 건수를 구하라.

```
x <- aggregate(사망자수~발생지시도+주야+도로형태+법규위반+요일, acc_2015, sum)
head(x, 20)
```

	발생지시도	주야	도로형태	법규위반	요일	사망자수
1	대전	야간	교차로내	과속	금	1
2	부산	야간	교차로내	과속	금	1
3	경기	주간	교차로내	과속	금	1
4	경남	주간	교차로내	과속	금	1
5	인천	야간	교차로부근	과속	금	1
6	대구	주간	교차로부근	과속	금	1
7	경기	야간	기타단일로	과속	금	3
8	경북	야간	기타단일로	과속	금	1
9	대구	야간	기타단일로	과속	금	1
10	대전	야간	기타단일로	과속	금	1
11	울산	야간	기타단일로	과속	금	2
12	전남	야간	기타단일로	과속	금	1
13	전북	야간	기타단일로	과속	금	1
14	전남	주간	기타단일로	과속	금	1
15	제주	주간	기타단일로	과속	금	1
16	광주	야간	횡단보도상	과속	금	1

17	경기 야간	교차로내	교차로	통행방법 위반	금	2
18	경북 야간	교차로내	교차로	통행방법 위반	금	1
19	충북 야간	교차로내	교차로	통행방법 위반	금	1
20	강원 주간	교차로내	교차로	통행방법 위반	금	1

3 위의 요인들 (발생지시도, 주야, 도로형태, 법규위반, 요일)중에서 가장 사망사고가 많이 발생하는 조합을 상위 10개만 찾으시오.

```
sidx1 <- order(x$사망자수, decreasing = T)
x[sidx1[1:10],]
```

	발생지시도	주야	도로형태	법규위반	요일	사망자수
1615	경기	야간	기타단일로	안전운전 의무 불이행	화	41
172	경기	야간	기타단일로	안전운전 의무 불이행	금	37
1140	경기	야간	기타단일로	안전운전 의무 불이행	일	37
412	경기	야간	기타단일로	안전운전 의무 불이행	목	35
1374	경기	야간	기타단일로	안전운전 의무 불이행	토	32
910	경기	야간	기타단일로	안전운전 의무 불이행	월	29
1617	경북	야간	기타단일로	안전운전 의무 불이행	화	28
666	경기	야간	기타단일로	안전운전 의무 불이행	수	27
1389	경기	주간	기타단일로	안전운전 의무 불이행	토	27
912	경북	야간	기타단일로	안전운전 의무 불이행	월	25

4 2015년 자료의 처리와 동일하게 2016년 자료를 처리하시오.

```
acc_2016 <- read.csv("2016년_사망교통사고.csv", fileEncoding = "CP949")
y <- aggregate(사망자수~발생지시도+주야+도로형태+법규위반+요일, acc_2016, sum)
sidx2 <- order(y$사망자수, decreasing = T)
y[sidx2[1:10],]
```

	발생지시도	주야	도로형태	법규위반	요일	사망자수
638	경기	야간	기타단일로	안전운전 의무 불이행	수	32
1340	경기	야간	기타단일로	안전운전 의무 불이행	토	30
1561	경기	야간	기타단일로	안전운전 의무 불이행	화	30
867	경기	야간	기타단일로	안전운전 의무 불이행	월	29
407	경기	야간	기타단일로	안전운전 의무 불이행	목	28
1092	경기	야간	기타단일로	안전운전 의무 불이행	일	28
409	경북	야간	기타단일로	안전운전 의무 불이행	목	27
885	경북	주간	기타단일로	안전운전 의무 불이행	월	27
1094	경북	야간	기타단일로	안전운전 의무 불이행	일	26
170	경기	야간	기타단일로	안전운전 의무 불이행	금	25

5 정리된 두 자료를 행으로 결합하려고 한다.

1. 먼저 두 자료를 식별하기 위하여 각 자료에 발생연도컬럼을 추가하라.

```
x$발생년 <- 2015
y$발생년 <- 2016
```

2. 발생연도컬럼이 추가된 두 자료를 행으로 결합하여라.

```
xy <- rbind(x, y)
sidx <- order(xy$사망자수, decreasing = T)
head(xy[sidx,])
```

	발생지시도	주야	도로형태	법규위반	요일	사망자수	발생년
1615	경기	야간	기타단일로	안전운전 의무 불이행	화	41	2015
172	경기	야간	기타단일로	안전운전 의무 불이행	금	37	2015
1140	경기	야간	기타단일로	안전운전 의무 불이행	일	37	2015
412	경기	야간	기타단일로	안전운전 의무 불이행	목	35	2015
1374	경기	야간	기타단일로	안전운전 의무 불이행	토	32	2015
2338	경기	야간	기타단일로	안전운전 의무 불이행	수	32	2016

6 발생지시도별 연도별 사망자수를 구하라.

```
( o <- aggregate(사망자수~발생지시도+발생년, xy, sum) )
```

	발생지시도	발생년	사망자수
1	강원	2015	239
2	경기	2015	880
3	경남	2015	406
4	경북	2015	507
5	광주	2015	97
6	대구	2015	161
7	대전	2015	97
8	부산	2015	178
9	서울	2015	376
10	세종	2015	19
11	울산	2015	94
12	인천	2015	136
13	전남	2015	410
14	전북	2015	308
15	제주	2015	93
16	충남	2015	382
17	충북	2015	238
18	강원	2016	219

19	경기	2016	780
20	경남	2016	373
21	경북	2016	495
22	광주	2016	85
23	대구	2016	164
24	대전	2016	99
25	부산	2016	159
26	서울	2016	348
27	세종	2016	25
28	울산	2016	81
29	인천	2016	144
30	전남	2016	337
31	전북	2016	284
32	제주	2016	80
33	충남	2016	388
34	충북	2016	231

7 위의 결과 자료를 이용하여 발생지시도별 연도별 사망자수를 아래와 같이 교차표로 나타내어라.

```
out1 <- o[o$발생년==2015,]
out2 <- o[o$발생년==2016,]
names(out1)[3] <- "2015"
names(out2)[3] <- "2016"
out <- merge(out1, out2, by="발생지시도")[, c(1,3,5)]
knitr::kable(out)
```

발생지시도	2015	2016
강원	239	219
경기	880	780
경남	406	373
경북	507	495
광주	97	85
대구	161	164
대전	97	99
부산	178	159
서울	376	348
세종	19	25
울산	94	81
인천	136	144
전남	410	337
전북	308	284

발생지시도	2015	2016
제주	93	80
충남	382	388
충북	238	231

```
barplot(as.matrix(t(out[, c(2:3)])), names.arg=out[,1],beside=T, col=2:3, las=2)
```

