

3. 빈도분석 실습작업



교통사건사고 데이터를 이용하여 빈도분석작업 실습

I. 지역별 교통사고 빈도분석

- 작업 데이터 - 2012~2014년 교통사건사고파일 accdata2.csv
- 분석 목표 - 지역별 교통사고 사상자 합계를 요약하여 시각화하는 프로그램구현

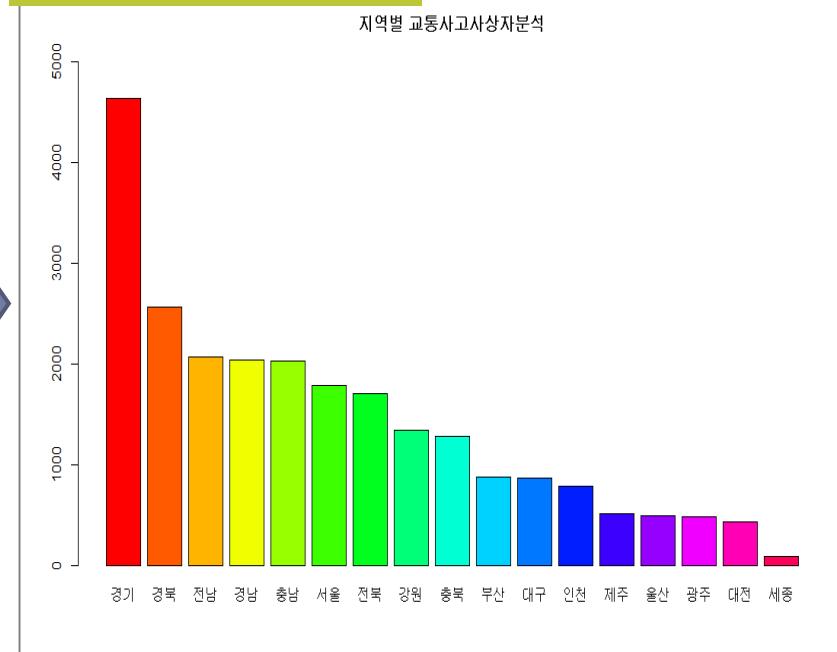
accdata2.csv 데이터셋

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	발생년	주야	요일	사망자수	사상자수	발생지시도발생지시군	사고유형_	사고유형_	법규위반_	법규위반_	경도	위도	
2	2012	야간	수	1	1	경남	거창군	차대차	추돌	진행중	운전자법규안전운전	127.8532	35.58072
3	2012	주간	금	1	1	경기	용인시	차대차	추돌	주정차중	운전자법규안전운전	127.0529	37.31761
4	2012	야간	금	1	2	경북	김천시	차대차	즉면직각충	즉면직각충	운전자법규안전운전	128.0032	36.19092
5	2012	주간	목	1	1	경기	이천시	차대차	추돌	주정차중	운전자법규안전운전	127.4204	37.23896
6	2012	야간	월	1	1	전남	광양시	차대차	추돌	진행중	운전자법규안전운전	127.6484	34.96437
7	2012	주간	일	1	5	전남	광양시	차량단독	전도전복	전도전복	운전자법규안전운전	127.7275	34.98728
8	2012	야간	목	1	2	경기	하남시	차대차	추돌	진행중	운전자법규안전운전	127.1774	37.57841
9	2012	주간	월	1	1	충남	논산시	차량단독	공작물충	공작물충	운전자법규안전운전	127.1355	36.12286
10	2012	야간	월	1	5	경남	밀양시	차대차	즉면직각충	즉면직각충	운전자법규부당한회	128.774	35.44976
11	2012	주간	목	1	1	전북	진안군	차량단독	전도전복	전도전복	운전자법규안전운전	127.4248	35.77755
12	2012	주간	목	1	5	대구	북구	차대차	추돌	진행중	운전자법규안전운전	128.5375	35.89471
13	2012	야간	일	1	1	대구	달성군	차대차	추돌	진행중	운전자법규안전운전	128.4843	35.80141
14	2012	주간	월	2	2	전북	무주군	차대차	기타	기타	운전자법규안전운전	127.6584	35.96462
15	2012	야간	화	1	1	경기	화성시	차대차	추돌	진행중	운전자법규안전운전	126.9016	37.28902
16	2012	야간	금	1	2	경기	광명시	차대차	추돌	진행중	운전자법규안전운전	126.852	37.43926
17	2012	주간	일	1	1	경남	진주시	차대사람	기타	기타	운전자법규안전운전	128.2759	35.24882
18	2012	야간	화	1	1	경남	진주시	차량단독	공작물충	공작물충	운전자법규안전운전	128.2716	35.23129
19	2012	야간	토	1	1	경북	구미시	차대차	추돌	진행중	운전자법규안전운전	128.2576	36.26773
20	2012	야간	목	1	3	경북	구미시	차대차	추돌	주정차중	운전자법규안전운전	128.2489	36.2767
21	2012	야간	일	1	6	전북	정읍시	차대차	정면충돌	정면충돌	운전자법규중앙선침	126.9094	35.63408
22	2012	야간	일	1	1	경기	화성시	차대사람	차도통행	차도통행	운전자법규안전운전	126.8107	37.20664
23	2012	야간	일	1	2	충남	서산시	차량단독	공작물충	공작물충	운전자법규안전운전	126.5045	36.75058
24	2012	야간	일	1	1	광주	서구	차대사람	횡단중	횡단중	운전자법규안전운전	126.8339	35.15331
25	2012	주간	수	1	1	경남	김해시	차대차	즉면직각충	즉면직각충	운전자법규안전거리	128.8591	35.22071
26	2012	야간	수	1	1	경기	안성시	차대사람	기타	기타	운전자법규안전운전	127.2018	37.06876
27	2012	주간	월	1	1	강원	춘천시	차대사람	기타	기타	운전자법규안전운전	127.6849	37.78842
28	2012	야간	금	1	1	대전	유성구	차대사람	횡단중	횡단중	운전자법규안전운전	127.3253	36.29997
29	2012	주간	목	1	1	충남	태안군	차량단독	기타	기타	운전자법규안전운전	126.2918	36.75975
30	2012	야간	목	1	1	경기	포천시	차대사람	횡단중	횡단중	운전자법규과속	127.2178	38.01433
31	2012	야간	금	1	1	인천	계양구	차대사람	횡단중	횡단중	운전자법규안전운전	126.721	37.54354
32	2012	주간	화	1	1	서울	강서구	차량단독	기타	기타	운전자법규안전운전	126.822	37.54416
33	2012	야간	목	1	1	전남	순천시	차대사람	기타	기타	운전자법규안전운전	127.5214	34.95367
34	2012	야간	금	1	1	경기	양평군	차대차	추돌	진행중	운전자법규안전운전	127.4149	37.51546
35	2012	주간	화	1	1	이천	가하구	차대차	저며추돌	저며추돌	운전자법규안전운전	126.4716	37.64689

지역별 사상자요약

지역	사상자합계
강원	1339
경기	4642
경남	2035
경북	2560
광주	480
대구	861
대전	426
부산	876
서울	1781
세종	83
울산	493
인천	781
전남	2074
전북	1707
제주	513
충남	2024
충북	1280

지역별 사상자요약 시각화



how to

1. 작업디렉토리와 파일을 읽어온 다음 파일내용의 일부분을 확인하는 작업을 그림처럼 수행해보아요.

```
> setwd("d:\\machine_r")
> source1<-read.csv("accdata2.csv",header=TRUE)
> head(source1,5)
  발생년 주야 요일 사망자수 사상자수 발생지시도 발생지시군구 사고유형_대분류 사고유형_중분류      사고유형_법규위반 대분류      법규위반 경도   위도
1  2012 야간 수      1       1     경남     거창군      차대차      추돌      진행중      추돌 운전자법규위반 안전운전 의무 불이행 127.8532 35.58072
2  2012 주간 금      1       1     경기     용인시      차대차      추돌      주정차중      추돌 운전자법규위반 안전운전 의무 불이행 127.0529 37.31761
3  2012 야간 금      1       2     경북     김천시      차대차  측면직각충돌  측면직각충돌 운전자법규위반 안전운전 의무 불이행 128.0032 36.19092
4  2012 주간 목      1       1     경기     이천시      차대차      추돌      주정차중      추돌 운전자법규위반 안전운전 의무 불이행 127.4204 37.23896
5  2012 야간 월      1       1     전남     광양시      차대차      추돌      진행중      추돌 운전자법규위반 안전운전 의무 불이행 127.6484 34.96437
>
```

2. 발생지시도별로 사상자수 합계를 요약하는 작업을 수행해야겠죠? tapply()함수를 이용하여 그림처럼 작업해요.

```
> totinfo<-tapply(source1$사상자수,source1$발생지시도,sum)
> totinfo
강원 경기 경북 광주 대구 대전 부산 서울 세종 울산 인천 전남 전북 제주 총남 총북
1339 4642 2035 2560 480 861 426 876 1781 83 493 781 2074 1707 513 2024 1280
>
>
```

tapply(data,index,function) – data를 index에서 지정한 팩터값으로 분류하여 지정된 function을 적용

이렇게 사상자수의 합계가 요약돼요.

3. 계산된 결과를 데이터프레임구조로 변환시켜주는 작업을 그림처럼 코딩해요.

```
> df <- data.frame(지역=names(totinfo), 사상자총계=totinfo)
> df
  지역 사상자총계
  강원 강원      1339
  경기 경기      4642
  경남 경남      2035
  경북 경북      2560
  광주 광주      480
  대구 대구      861
  대전 대전      426
  부산 부산      876
  서울 서울      1781
  세종 세종      83
  울산 울산      493
  인천 인천      781
  전남 전남      2074
  전북 전북      1707
  제주 제주      513
  충남 충남      2024
  충북 충북      1280
>
```

4. df를 실행한 결과 사상자총계값이 큰 순서대로 표시된다면 더 직관적으로 지역별 사상자수 합계를 빨리 파악할 수 있겠죠? 그래서, 사상자총계가 높은 순서로 정렬작업을 코딩해요.

```
> df<-df[order(df$사상자총계, decreasing = TRUE), ]
```

```
> df
  지역 사상자총계
  경기 경기      4642
  경북 경북      2560
  전남 전남      2074
  경남 경남      2035
  충남 충남      2024
  서울 서울      1781
  전북 전북      1707
  강원 강원      1339
  충북 충북      1280
  부산 부산      876
  대구 대구      861
  인천 인천      781
  제주 제주      513
  울산 울산      493
  광주 광주      480
  대전 대전      426
  세종 세종      83
```

decreasing = true로 설정하면 내림차순정렬이 되는거예요.

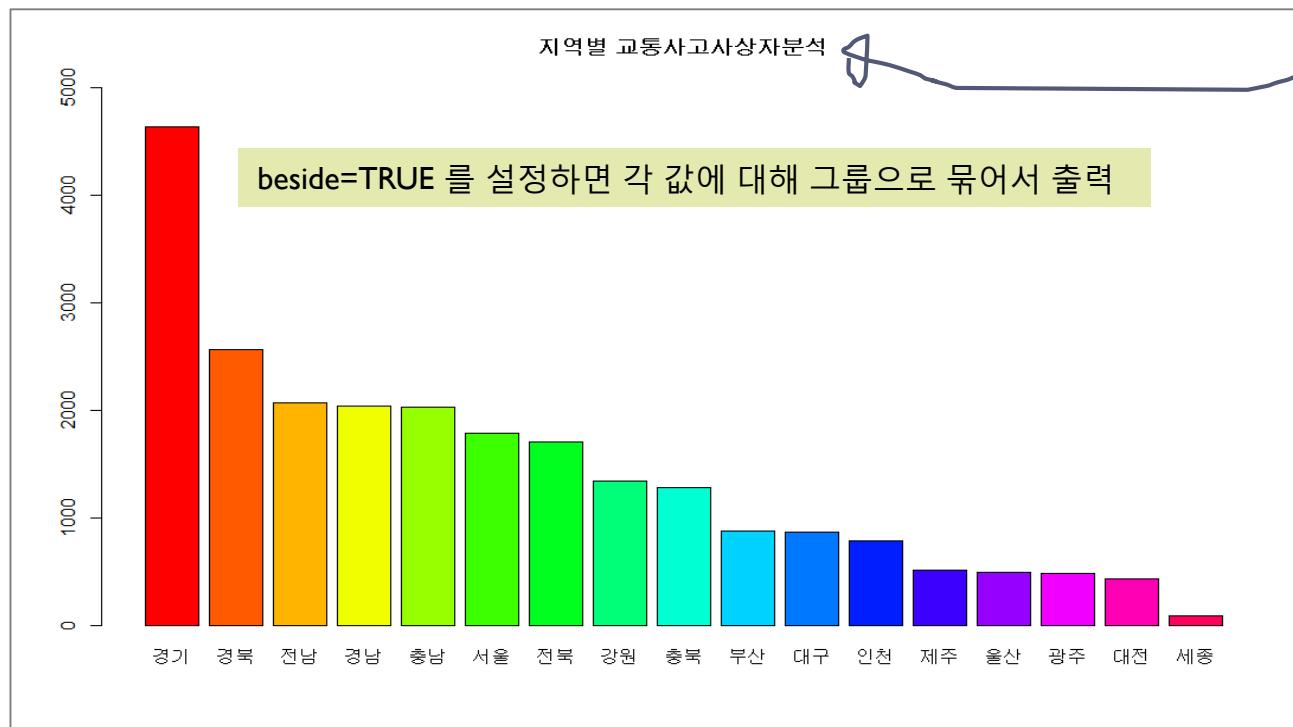
5. 요약결과를 막대도표로 시각화하기 위해서는 이들을 벡터나 행열의 구조로 변경을 해줘야해요. 따라서, 지역사상자총계를 벡터로 각각 변환해요.

```
> vcol1<-as.vector(df[,1])
> vcol1
[1] "경기" "경북" "전남" "경남" "충남" "서울" "전북" "강원" "충북" "부산" "대구" "인천" "제주" "울산" "광주" "대전" "세종"
> vcol2<-as.vector(df[,2])
> vcol2
[1] 4642 2560 2074 2035 2024 1781 1707 1339 1280 876 861 781 513 493 480 426 83
>
```

지역	사상자총계
경기	4642
경북	2560
전남	2074
경남	2035
충남	2024
서울	1781
전북	1707
강원	1339
충북	1280
부산	876
대구	861
인천	781
제주	513
울산	493
광주	480
대전	426
세종	83

6. 벡터변수 vcol1,vcol2를 이용하여 막대차트를 구현해요. barplot()함수를 이용하면 아주 간단하게 그릴 수 있죠?

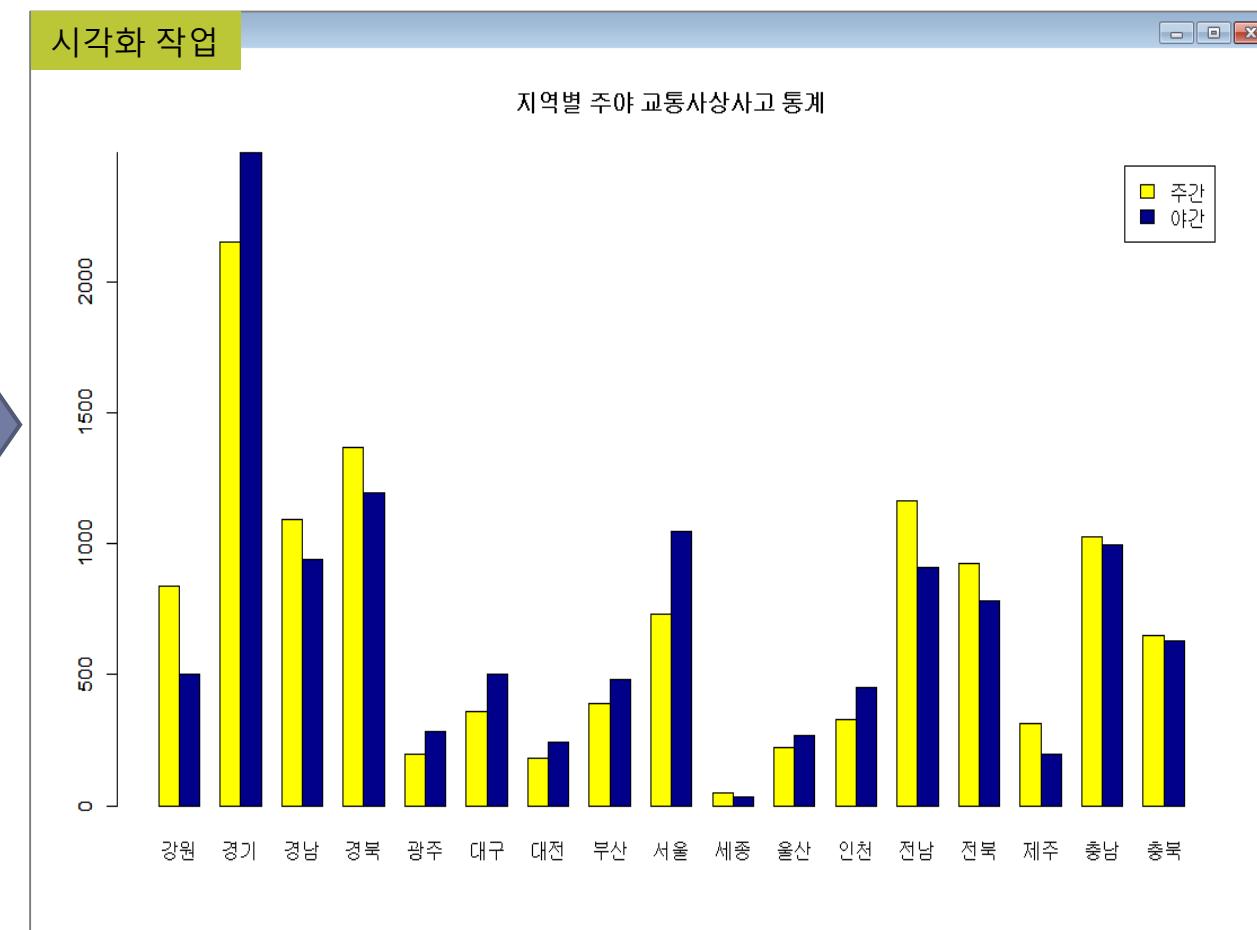
```
> barplot(vcol2,beside=TRUE,names=vcol1,col=rainbow(length(vcol1)),ylim=c(0,5000),main="지역별 교통사고사상자분석")
```



2. 지역별/주간,야간 교통사고사상자 분석

- ▶ 분석 목표 – 지역별 교통사고사상자를 주간/야간으로 분류하여 사상자합계를 요약

	주간	야간
강원	839	500
경기	2150	2492
경남	1094	941
경북	1367	1193
광주	197	283
대구	360	501
대전	181	245
부산	392	484
서울	734	1047
세종	47	36
울산	224	269
인천	331	450
전남	1163	911
전북	927	780
제주	316	197
충남	1028	996
충북	649	631



how to

- 앞서 작업한 source1에 대해 발생지시도, 주야 항목별 사상자수 합계를 요약하기 위한 tapply()함수를 그림처럼 기술해요.

```
> tot2<-tapply(source1$사상자수,list(source1$발생지시도,factor(source1$주야,level=c("주간","야간"))),sum)
```

```
>
```

```
> a<-data.frame(tot2)
```

tot2는 matrix 구조가 되므로, 이를
데이터프레임구조로 변환하여 a에 저장

2개 항목에 대한 그룹화가 이뤄짐
이때 주야항목은 주간,야간 순서로 인덱싱되도록 목록지정

```
>
```

```
> a
```

	주간	야간
강원	839	500
경기	2150	2492
경남	1094	941
경북	1367	1193
광주	197	283
대구	360	501
대전	181	245
부산	392	484
서울	734	1047
세종	47	36
울산	224	269
인천	331	450
전남	1163	911
전북	927	780
제주	316	197
충남	1028	996
충북	649	631

2. 데이터프레임의 행/열 구조를 변경해보죠. 분석목적에 따라 때때로 행/열의 구조를 변경할 필요가 있죠? 여기서는 data.table 패키지에서 지원하는 transpose() 함수를 이용하여 손쉽게 구조를 변경해보죠.

```
> library(data.table)          해당 패키지가 설치되어 있지 않다면 설치작업부터 진행해주세요.
> t_1<-transpose(a)
> colnames(t_1)<-rownames(a)  기존 데이터프레임 a의 행이름→t_1의 열이름, a의 열 이름을 t_1의 행이름으로 대입해요.
> rownames(t_1)<-colnames(a)
> t_1
      강원 경기 경남 경북 광주 대구 대전 부산 서울 세종 울산 인천 전남 전북 제주 충남 충북
주간  839 2150 1094 1367 197  360 181  392  734   47  224  331 1163  927  316 1028  649
야간  500 2492  941 1193 283  501 245  484 1047   36  269  450  911  780  197  996  631
>
```

3. barplot()함수를 이용하여 간단하게 결과를 시각화해보죠.

```
> barplot(as.matrix(t_1[,]),beside=T,col=c('yellow','darkblue'),main="지역별 주야 교통사상사고 통계",legend=c("주간","야간"))
> axis(2,seq(0,3000,500))
>
```

