FinalProject

Jaegeun

# 교통사고 사망자의 분류 분석

## 

### 데이터 불러오기

csv\_Data<-read.csv('전국\_사망교통사고\_2017.csv')   
head(csv\_Data)

## 발생년 발생년월일시 발생분 주야 요일 사망자수 사상자수 중상자수 경상자수  
## 1 2017 2017010101 15 야간 일 1 2 1 0  
## 2 2017 2017010102 28 야간 일 1 1 0 0  
## 3 2017 2017010102 43 야간 일 1 2 0 1  
## 4 2017 2017010102 22 야간 일 1 1 0 0  
## 5 2017 2017010104 37 야간 일 1 1 0 0  
## 6 2017 2017010104 0 야간 일 1 2 1 0  
## 부상신고자수 발생지시도 발생지시군구 사고유형\_대분류 사고유형\_중분류  
## 1 0 전남 장성군 차대차 기타  
## 2 0 서울 송파구 차대사람 횡단중  
## 3 0 충남 홍성군 차대차 추돌  
## 4 0 충남 아산시 차량단독 공작물충돌  
## 5 0 서울 금천구 차대사람 기타  
## 6 0 강원 춘천시 차대사람 횡단중  
## 사고유형 법규위반\_대분류 법규위반 도로형태\_대분류  
## 1 기타 운전자법규위반 안전운전 의무 불이행 단일로  
## 2 횡단중 운전자법규위반 안전운전 의무 불이행 교차로  
## 3 추돌 운전자법규위반 안전운전 의무 불이행 단일로  
## 4 공작물충돌 운전자법규위반 안전운전 의무 불이행 단일로  
## 5 기타 운전자법규위반 안전운전 의무 불이행 단일로  
## 6 횡단중 운전자법규위반 안전운전 의무 불이행 단일로  
## 도로형태 당사자종별\_1당\_대분류 당사자종별\_1당 당사자종별\_2당\_대분류  
## 1 기타단일로 승용차 승용차 승용차  
## 2 교차로부근 승용차 승용차 보행자  
## 3 기타단일로 승용차 승용차 승용차  
## 4 기타단일로 승용차 승용차 없음  
## 5 기타단일로 승용차 승용차 보행자  
## 6 기타단일로 승용차 승용차 보행자  
## 당사자종별\_2당 발생위치X\_UTMK 발생위치Y\_UTMK 경도 위도  
## 1 승용차 933501 1700129 126.7686 35.29446  
## 2 보행자 967570 1944453 127.1331 37.49874  
## 3 승용차 916497 1842880 126.5666 36.58007  
## 4 없음 961608 1864573 127.0698 36.77851  
## 5 보행자 946778 1941695 126.8981 37.47295  
## 6 보행자 1021195 1985798 127.7410 37.87175

### 불필요한 데이터 삭제

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Data<-csv\_Data%>%  
 select(-발생년,-발생년월일시,-발생분,-당사자종별\_1당\_대분류,-당사자종별\_2당\_대분류,-사고유형)

### 데이터 확인

str(Data)

## 'data.frame': 4064 obs. of 21 variables:  
## $ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 1 2 1 1 ...  
## $ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ 사망자수 : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ 사상자수 : int 2 1 2 1 1 2 1 5 3 7 ...  
## $ 중상자수 : int 1 0 0 0 0 1 0 3 1 4 ...  
## $ 경상자수 : int 0 0 1 0 0 0 0 1 1 2 ...  
## $ 부상신고자수 : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 9 16 16 9 1 2 5 15 16 ...  
## $ 발생지시군구 : Factor w/ 206 levels "가평군","강남구",..: 158 100 201 108 39 182 203 81 89 21 ...  
## $ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2 1 2 3 1 1 1 3 2 2 ...  
## $ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 2 14 12 1 2 14 14 4 13 2 ...  
## $ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 9 9 ...  
## $ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 1 3 3 3 3 3 3 3 3 ...  
## $ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 4 7 7 7 7 7 2 7 7 ...  
## $ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 6 6 12 6 6 ...  
## $ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 7 4 7 9 4 4 4 9 14 7 ...  
## $ 발생위치X\_UTMK : int 933501 967570 916497 961608 946778 1021195 959140 942432 922514 962530 ...  
## $ 발생위치Y\_UTMK : int 1700129 1944453 1842880 1864573 1941695 1985798 1912805 1692541 1476665 1823895 ...  
## $ 경도 : num 127 127 127 127 127 ...  
## $ 위도 : num 35.3 37.5 36.6 36.8 37.5 ...

### 데이터 타입 형식 변환

Data$발생지시군구<-as.character(Data$발생지시군구)

### 교차검증 준비

###### (createDataPartition()사용)하기 위해 caret 라이브러리 사용

library(caret)

## Warning: package 'caret' was built under R version 3.4.4

## Loading required package: lattice

## Loading required package: ggplot2

set.seed(137) #항상 같은데이터로 분리하기 위해서  
test\_idx<-createDataPartition(Data$주야,p=0.3)$Resample1   
  
test\_Data<-Data[test\_idx,]  
train\_Data<-Data[-test\_idx,]  
NROW(test\_Data)

## [1] 1220

prop.table(table(test\_Data$주야))

##   
## 야간 주간   
## 0.504918 0.495082

NROW(train\_Data)

## [1] 2844

prop.table(table(train\_Data$주야))

##   
## 야간 주간   
## 0.504571 0.495429

### 교차검증 함수 생성

create\_ten\_fold\_cv<-function(){  
 set.seed(137)  
 lapply(createFolds(train\_Data,k=10),function(idx){  
 return(list(train=train\_Data[-idx,],  
 validation=train\_Data[idx,]))  
 })  
}  
x<-create\_ten\_fold\_cv()  
head(str(x))

## List of 8  
## $ Fold1:List of 2  
## ..$ train :'data.frame': 2840 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 1 1 1 1 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 4 4 ...  
## .. ..$ 사망자수 : int [1:2840] 1 1 1 1 1 1 1 2 1 1 ...  
## .. ..$ 사상자수 : int [1:2840] 1 2 1 2 1 5 3 2 2 1 ...  
## .. ..$ 중상자수 : int [1:2840] 0 0 0 1 0 3 1 0 0 0 ...  
## .. ..$ 경상자수 : int [1:2840] 0 1 0 0 0 1 1 0 1 0 ...  
## .. ..$ 부상신고자수 : int [1:2840] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 1 2 5 15 13 11 9 ...  
## .. ..$ 발생지시군구 : chr [1:2840] "송파구" "홍성군" "아산시" "춘천시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 1 3 2 3 1 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 14 4 13 7 11 14 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 1 4 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 3 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 2 7 7 7 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 12 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 4 9 14 9 4 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2840] 967570 916497 961608 1021195 959140 942432 922514 905827 1162551 948413 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2840] 1944453 1842880 1864573 1985798 1912805 1692541 1476665 1665350 1720013 1957900 ...  
## .. ..$ 경도 : num [1:2840] 127 127 127 128 127 ...  
## .. ..$ 위도 : num [1:2840] 37.5 36.6 36.8 37.9 37.2 ...  
## ..$ validation:'data.frame': 4 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 2 2 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 4 4 4  
## .. ..$ 사망자수 : int [1:4] 1 1 1 1  
## .. ..$ 사상자수 : int [1:4] 7 1 1 5  
## .. ..$ 중상자수 : int [1:4] 4 0 0 0  
## .. ..$ 경상자수 : int [1:4] 2 0 0 4  
## .. ..$ 부상신고자수 : int [1:4] 0 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 16 9 4 9  
## .. ..$ 발생지시군구 : chr [1:4] "공주시" "영등포구" "울진군" "금천구"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2 3 1 2  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 2 7 14 13  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 7  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 1 3 1  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 3 7 3  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 9 7 6  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 7 9 4 8  
## .. ..$ 발생위치X\_UTMK : int [1:4] 962530 946757 1166337 947235  
## .. ..$ 발생위치Y\_UTMK : int [1:4] 1823895 1943309 1901018 1938476  
## .. ..$ 경도 : num [1:4] 127 127 129 127  
## .. ..$ 위도 : num [1:4] 36.4 37.5 37.1 37.4  
## $ Fold2:List of 2  
## ..$ train :'data.frame': 2843 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 1 1 1 1 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 5 4 ...  
## .. ..$ 사망자수 : int [1:2843] 1 1 1 1 1 1 1 1 2 1 ...  
## .. ..$ 사상자수 : int [1:2843] 1 2 1 2 1 5 3 7 2 2 ...  
## .. ..$ 중상자수 : int [1:2843] 0 0 0 1 0 3 1 4 0 0 ...  
## .. ..$ 경상자수 : int [1:2843] 0 1 0 0 0 1 1 2 0 1 ...  
## .. ..$ 부상신고자수 : int [1:2843] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 1 2 5 15 16 13 11 ...  
## .. ..$ 발생지시군구 : chr [1:2843] "송파구" "홍성군" "아산시" "춘천시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 1 3 2 2 3 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 14 4 13 2 7 11 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 9 1 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 3 3 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 2 7 7 7 7 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 12 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 4 9 14 7 9 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2843] 967570 916497 961608 1021195 959140 942432 922514 962530 905827 1162551 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2843] 1944453 1842880 1864573 1985798 1912805 1692541 1476665 1823895 1665350 1720013 ...  
## .. ..$ 경도 : num [1:2843] 127 127 127 128 127 ...  
## .. ..$ 위도 : num [1:2843] 37.5 36.6 36.8 37.9 37.2 ...  
## ..$ validation:'data.frame': 1 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 4  
## .. ..$ 사망자수 : int 1  
## .. ..$ 사상자수 : int 1  
## .. ..$ 중상자수 : int 0  
## .. ..$ 경상자수 : int 0  
## .. ..$ 부상신고자수 : int 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 14  
## .. ..$ 발생지시군구 : chr "전주시"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 4  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 3  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4  
## .. ..$ 발생위치X\_UTMK : int 966955  
## .. ..$ 발생위치Y\_UTMK : int 1754577  
## .. ..$ 경도 : num 127  
## .. ..$ 위도 : num 35.8  
## $ Fold3:List of 2  
## ..$ train :'data.frame': 2842 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 1 1 1 1 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 4 4 ...  
## .. ..$ 사망자수 : int [1:2842] 1 1 1 1 1 1 1 2 1 1 ...  
## .. ..$ 사상자수 : int [1:2842] 1 2 1 2 1 5 7 2 2 1 ...  
## .. ..$ 중상자수 : int [1:2842] 0 0 0 1 0 3 4 0 0 0 ...  
## .. ..$ 경상자수 : int [1:2842] 0 1 0 0 0 1 2 0 1 0 ...  
## .. ..$ 부상신고자수 : int [1:2842] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 1 2 5 16 13 11 9 ...  
## .. ..$ 발생지시군구 : chr [1:2842] "송파구" "홍성군" "아산시" "춘천시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 1 3 2 3 1 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 14 4 2 7 11 14 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 1 4 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 3 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 2 7 7 7 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 12 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 4 9 7 9 4 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2842] 967570 916497 961608 1021195 959140 942432 962530 905827 1162551 948413 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2842] 1944453 1842880 1864573 1985798 1912805 1692541 1823895 1665350 1720013 1957900 ...  
## .. ..$ 경도 : num [1:2842] 127 127 127 128 127 ...  
## .. ..$ 위도 : num [1:2842] 37.5 36.6 36.8 37.9 37.2 ...  
## ..$ validation:'data.frame': 2 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 4  
## .. ..$ 사망자수 : int [1:2] 1 1  
## .. ..$ 사상자수 : int [1:2] 3 1  
## .. ..$ 중상자수 : int [1:2] 1 0  
## .. ..$ 경상자수 : int [1:2] 1 0  
## .. ..$ 부상신고자수 : int [1:2] 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 15 9  
## .. ..$ 발생지시군구 : chr [1:2] "서귀포시" "광진구"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2 3  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 13 1  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 3  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 7  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 14 9  
## .. ..$ 발생위치X\_UTMK : int [1:2] 922514 964223  
## .. ..$ 발생위치Y\_UTMK : int [1:2] 1476665 1950661  
## .. ..$ 경도 : num [1:2] 127 127  
## .. ..$ 위도 : num [1:2] 33.3 37.6  
## $ Fold4:List of 2  
## ..$ train :'data.frame': 2843 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 1 1 1 1 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 5 4 ...  
## .. ..$ 사망자수 : int [1:2843] 1 1 1 1 1 1 1 1 2 1 ...  
## .. ..$ 사상자수 : int [1:2843] 1 2 1 2 1 5 3 7 2 2 ...  
## .. ..$ 중상자수 : int [1:2843] 0 0 0 1 0 3 1 4 0 0 ...  
## .. ..$ 경상자수 : int [1:2843] 0 1 0 0 0 1 1 2 0 1 ...  
## .. ..$ 부상신고자수 : int [1:2843] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 1 2 5 15 16 13 11 ...  
## .. ..$ 발생지시군구 : chr [1:2843] "송파구" "홍성군" "아산시" "춘천시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 1 3 2 2 3 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 14 4 13 2 7 11 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 9 1 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 3 3 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 2 7 7 7 7 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 12 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 4 9 14 7 9 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2843] 967570 916497 961608 1021195 959140 942432 922514 962530 905827 1162551 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2843] 1944453 1842880 1864573 1985798 1912805 1692541 1476665 1823895 1665350 1720013 ...  
## .. ..$ 경도 : num [1:2843] 127 127 127 128 127 ...  
## .. ..$ 위도 : num [1:2843] 37.5 36.6 36.8 37.9 37.2 ...  
## ..$ validation:'data.frame': 1 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 4  
## .. ..$ 사망자수 : int 1  
## .. ..$ 사상자수 : int 2  
## .. ..$ 중상자수 : int 1  
## .. ..$ 경상자수 : int 0  
## .. ..$ 부상신고자수 : int 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 4  
## .. ..$ 발생지시군구 : chr "청송군"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 13  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 13  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 7  
## .. ..$ 발생위치X\_UTMK : int 1134705  
## .. ..$ 발생위치Y\_UTMK : int 1810932  
## .. ..$ 경도 : num 129  
## .. ..$ 위도 : num 36.3  
## $ Fold5:List of 2  
## ..$ train :'data.frame': 2841 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 1 1 1 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 4 4 ...  
## .. ..$ 사망자수 : int [1:2841] 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:2841] 1 2 1 2 1 5 3 7 1 1 ...  
## .. ..$ 중상자수 : int [1:2841] 0 0 0 1 0 3 1 4 0 0 ...  
## .. ..$ 경상자수 : int [1:2841] 0 1 0 0 0 1 1 2 0 0 ...  
## .. ..$ 부상신고자수 : int [1:2841] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 1 2 5 15 16 9 14 ...  
## .. ..$ 발생지시군구 : chr [1:2841] "송파구" "홍성군" "아산시" "춘천시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 1 3 2 2 1 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 14 4 13 2 14 14 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 4 4 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 1 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 2 7 7 3 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 12 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 4 9 14 7 4 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2841] 967570 916497 961608 1021195 959140 942432 922514 962530 948413 966955 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2841] 1944453 1842880 1864573 1985798 1912805 1692541 1476665 1823895 1957900 1754577 ...  
## .. ..$ 경도 : num [1:2841] 127 127 127 128 127 ...  
## .. ..$ 위도 : num [1:2841] 37.5 36.6 36.8 37.9 37.2 ...  
## ..$ validation:'data.frame': 3 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 4 4  
## .. ..$ 사망자수 : int [1:3] 2 1 1  
## .. ..$ 사상자수 : int [1:3] 2 2 1  
## .. ..$ 중상자수 : int [1:3] 0 0 0  
## .. ..$ 경상자수 : int [1:3] 0 1 0  
## .. ..$ 부상신고자수 : int [1:3] 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 11 2  
## .. ..$ 발생지시군구 : chr [1:3] "무안군" "울주군" "안성시"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 3 1 1  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 7 11 2  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 1 9  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 3 3  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 7 7  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 9 4 4  
## .. ..$ 발생위치X\_UTMK : int [1:3] 905827 1162551 978913  
## .. ..$ 발생위치Y\_UTMK : int [1:3] 1665350 1720013 1900967  
## .. ..$ 경도 : num [1:3] 126 129 127  
## .. ..$ 위도 : num [1:3] 35 35.5 37.1  
## $ Fold6:List of 2  
## ..$ train :'data.frame': 2841 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 1 1 1 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 4 4 4 ...  
## .. ..$ 사망자수 : int [1:2841] 1 1 1 1 1 1 2 1 1 1 ...  
## .. ..$ 사상자수 : int [1:2841] 1 2 1 1 3 7 2 2 1 1 ...  
## .. ..$ 중상자수 : int [1:2841] 0 0 0 0 1 4 0 0 0 0 ...  
## .. ..$ 경상자수 : int [1:2841] 0 1 0 0 1 2 0 1 0 0 ...  
## .. ..$ 부상신고자수 : int [1:2841] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 16 2 15 16 13 11 9 14 ...  
## .. ..$ 발생지시군구 : chr [1:2841] "송파구" "홍성군" "아산시" "화성시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 2 3 1 2 2 3 1 1 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 12 1 14 13 2 7 11 14 14 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 1 4 4 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 1 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 7 7 7 7 3 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 6 6 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 7 9 4 14 7 9 4 4 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2841] 967570 916497 961608 959140 922514 962530 905827 1162551 948413 966955 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2841] 1944453 1842880 1864573 1912805 1476665 1823895 1665350 1720013 1957900 1754577 ...  
## .. ..$ 경도 : num [1:2841] 127 127 127 127 127 ...  
## .. ..$ 위도 : num [1:2841] 37.5 36.6 36.8 37.2 33.3 ...  
## ..$ validation:'data.frame': 3 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 2 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 4  
## .. ..$ 사망자수 : int [1:3] 1 1 1  
## .. ..$ 사상자수 : int [1:3] 2 5 1  
## .. ..$ 중상자수 : int [1:3] 1 3 0  
## .. ..$ 경상자수 : int [1:3] 0 1 0  
## .. ..$ 부상신고자수 : int [1:3] 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 1 5 3  
## .. ..$ 발생지시군구 : chr [1:3] "춘천시" "북구" "남해군"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 3 1  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 4 14  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 3 1  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 2 4  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 12 12  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 9 4  
## .. ..$ 발생위치X\_UTMK : int [1:3] 1021195 942432 1048067  
## .. ..$ 발생위치Y\_UTMK : int [1:3] 1985798 1692541 1657068  
## .. ..$ 경도 : num [1:3] 128 127 128  
## .. ..$ 위도 : num [1:3] 37.9 35.2 34.9  
## $ Fold7:List of 2  
## ..$ train :'data.frame': 2841 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 2 1 1 1 1 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 5 5 4 4 ...  
## .. ..$ 사망자수 : int [1:2841] 1 1 1 1 1 1 1 2 1 1 ...  
## .. ..$ 사상자수 : int [1:2841] 1 1 2 1 5 3 7 2 2 1 ...  
## .. ..$ 중상자수 : int [1:2841] 0 0 1 0 3 1 4 0 0 0 ...  
## .. ..$ 경상자수 : int [1:2841] 0 0 0 0 1 1 2 0 1 0 ...  
## .. ..$ 부상신고자수 : int [1:2841] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 1 2 5 15 16 13 11 14 ...  
## .. ..$ 발생지시군구 : chr [1:2841] "송파구" "아산시" "춘천시" "화성시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 3 1 1 3 2 2 3 1 1 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 1 14 14 4 13 2 7 11 14 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 9 9 1 4 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3 3 3 3 3 3 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7 2 7 7 7 7 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 6 12 6 6 6 6 6 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 9 4 4 9 14 7 9 4 4 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2841] 967570 961608 1021195 959140 942432 922514 962530 905827 1162551 966955 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2841] 1944453 1864573 1985798 1912805 1692541 1476665 1823895 1665350 1720013 1754577 ...  
## .. ..$ 경도 : num [1:2841] 127 127 128 127 127 ...  
## .. ..$ 위도 : num [1:2841] 37.5 36.8 37.9 37.2 35.2 ...  
## ..$ validation:'data.frame': 3 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 4 4  
## .. ..$ 사망자수 : int [1:3] 1 1 1  
## .. ..$ 사상자수 : int [1:3] 2 1 1  
## .. ..$ 중상자수 : int [1:3] 0 0 0  
## .. ..$ 경상자수 : int [1:3] 1 0 0  
## .. ..$ 부상신고자수 : int [1:3] 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 16 9 4  
## .. ..$ 발생지시군구 : chr [1:3] "홍성군" "은평구" "의성군"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2 1 3  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 12 14 4  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 4 9  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 1 3  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 3 2  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 9  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 7 4 9  
## .. ..$ 발생위치X\_UTMK : int [1:3] 916497 948413 1080719  
## .. ..$ 발생위치Y\_UTMK : int [1:3] 1842880 1957900 1818230  
## .. ..$ 경도 : num [1:3] 127 127 128  
## .. ..$ 위도 : num [1:3] 36.6 37.6 36.4  
## $ Fold8:List of 2  
## ..$ train :'data.frame': 2840 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 2 1 1 1 1 1 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 5 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:2840] 1 1 1 1 1 2 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:2840] 2 2 5 3 7 2 2 1 1 1 ...  
## .. ..$ 중상자수 : int [1:2840] 0 1 3 1 4 0 0 0 0 0 ...  
## .. ..$ 경상자수 : int [1:2840] 1 0 1 1 2 0 1 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:2840] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 16 1 5 15 16 13 11 9 14 9 ...  
## .. ..$ 발생지시군구 : chr [1:2840] "홍성군" "춘천시" "북구" "서귀포시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 2 1 3 2 2 3 1 1 1 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 12 14 4 13 2 7 11 14 14 7 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 9 9 9 1 4 4 9 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 3 3 3 3 3 3 3 1 1 1 ...  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 7 7 2 7 7 7 7 3 3 3 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 12 6 6 6 6 6 6 9 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 7 4 9 14 7 9 4 4 4 9 ...  
## .. ..$ 발생위치X\_UTMK : int [1:2840] 916497 1021195 942432 922514 962530 905827 1162551 948413 966955 946757 ...  
## .. ..$ 발생위치Y\_UTMK : int [1:2840] 1842880 1985798 1692541 1476665 1823895 1665350 1720013 1957900 1754577 1943309 ...  
## .. ..$ 경도 : num [1:2840] 127 128 127 127 127 ...  
## .. ..$ 위도 : num [1:2840] 36.6 37.9 35.2 33.3 36.4 ...  
## ..$ validation:'data.frame': 4 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 4  
## .. ..$ 사망자수 : int [1:4] 1 1 1 1  
## .. ..$ 사상자수 : int [1:4] 1 1 1 1  
## .. ..$ 중상자수 : int [1:4] 0 0 0 0  
## .. ..$ 경상자수 : int [1:4] 0 0 0 0  
## .. ..$ 부상신고자수 : int [1:4] 0 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 16 2 15  
## .. ..$ 발생지시군구 : chr [1:4] "송파구" "아산시" "화성시" "제주시"  
## .. ..$ 사고유형\_대분류: Factor w/ 3 levels "차대사람","차대차",..: 1 3 1 1  
## .. ..$ 사고유형\_중분류: Factor w/ 15 levels "공작물충돌","기타",..: 14 1 14 14  
## .. ..$ 법규위반\_대분류: Factor w/ 2 levels "운전자법규위반",..: 1 1 1 1  
## .. ..$ 법규위반 : Factor w/ 15 levels "과속","교차로 통행방법 위반",..: 9 9 9 4  
## .. ..$ 도로형태\_대분류: Factor w/ 5 levels "교차로","기타",..: 1 3 3 3  
## .. ..$ 도로형태 : Factor w/ 11 levels "고가도로위","교량위",..: 4 7 7 7  
## .. ..$ 당사자종별\_1당 : Factor w/ 12 levels "개인형이동수단(PM)",..: 6 6 6 12  
## .. ..$ 당사자종별\_2당 : Factor w/ 14 levels "건설기계","기타",..: 4 9 4 4  
## .. ..$ 발생위치X\_UTMK : int [1:4] 967570 961608 959140 910531  
## .. ..$ 발생위치Y\_UTMK : int [1:4] 1944453 1864573 1912805 1502842  
## .. ..$ 경도 : num [1:4] 127 127 127 127  
## .. ..$ 위도 : num [1:4] 37.5 36.8 37.2 33.5

## NULL

head(x$Fold1$train)

## 주야 요일 사망자수 사상자수 중상자수 경상자수 부상신고자수 발생지시도  
## 2 야간 일 1 1 0 0 0 서울  
## 3 야간 일 1 2 0 1 0 충남  
## 4 야간 일 1 1 0 0 0 충남  
## 6 야간 일 1 2 1 0 0 강원  
## 7 야간 일 1 1 0 0 0 경기  
## 8 주간 일 1 5 3 1 0 광주  
## 발생지시군구 사고유형\_대분류 사고유형\_중분류 법규위반\_대분류  
## 2 송파구 차대사람 횡단중 운전자법규위반  
## 3 홍성군 차대차 추돌 운전자법규위반  
## 4 아산시 차량단독 공작물충돌 운전자법규위반  
## 6 춘천시 차대사람 횡단중 운전자법규위반  
## 7 화성시 차대사람 횡단중 운전자법규위반  
## 8 북구 차량단독 도로이탈 운전자법규위반  
## 법규위반 도로형태\_대분류 도로형태 당사자종별\_1당  
## 2 안전운전 의무 불이행 교차로 교차로부근 승용차  
## 3 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 4 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 6 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 7 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 8 안전운전 의무 불이행 단일로 교량위 화물차  
## 당사자종별\_2당 발생위치X\_UTMK 발생위치Y\_UTMK 경도 위도  
## 2 보행자 967570 1944453 127.1331 37.49874  
## 3 승용차 916497 1842880 126.5666 36.58007  
## 4 없음 961608 1864573 127.0698 36.77851  
## 6 보행자 1021195 1985798 127.7410 37.87175  
## 7 보행자 959140 1912805 127.0395 37.21317  
## 8 없음 942432 1692541 126.8674 35.22660

### 데이터 탐색

##### Hmisc패키지를 이용하여 summary에 formula지정 가능하게 한다.

library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.4.4

## Loading required package: survival

##   
## Attaching package: 'survival'

## The following object is masked from 'package:caret':  
##   
## cluster

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.4.4

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

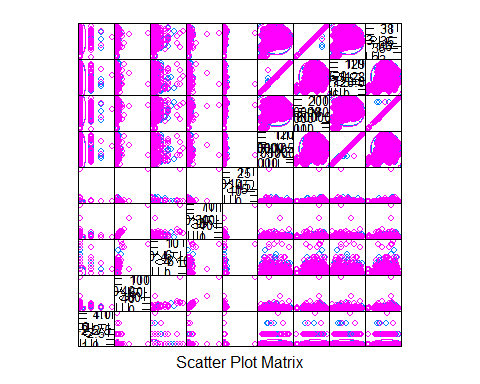
## The following objects are masked from 'package:base':  
##   
## format.pval, units

summary(주야~요일+발생지시도+사고유형\_대분류+사고유형\_중분류+도로형태,data=x$Fold1$train,method='reverse')

##   
##   
## Descriptive Statistics by 주야  
##   
## +----------------------------+--------+--------+  
## | |야간 |주간 |  
## | |(N=1434) |(N=1406) |  
## +----------------------------+--------+--------+  
## |요일 : 금 |15% (217)|15% (205)|  
## +----------------------------+--------+--------+  
## | 목 |15% (219)|14% (199)|  
## +----------------------------+--------+--------+  
## | 수 |13% (190)|14% (193)|  
## +----------------------------+--------+--------+  
## | 월 |13% (186)|17% (233)|  
## +----------------------------+--------+--------+  
## | 일 |14% (195)|10% (145)|  
## +----------------------------+--------+--------+  
## | 토 |15% (217)|13% (189)|  
## +----------------------------+--------+--------+  
## | 화 |15% (210)|17% (242)|  
## +----------------------------+--------+--------+  
## |발생지시도 : 강원 | 5% ( 71)| 6% ( 82)|  
## +----------------------------+--------+--------+  
## | 경기 |21% (308)|17% (239)|  
## +----------------------------+--------+--------+  
## | 경남 | 7% (102)|10% (136)|  
## +----------------------------+--------+--------+  
## | 경북 | 9% (127)|11% (156)|  
## +----------------------------+--------+--------+  
## | 광주 | 3% ( 48)| 2% ( 35)|  
## +----------------------------+--------+--------+  
## | 대구 | 4% ( 54)| 4% ( 52)|  
## +----------------------------+--------+--------+  
## | 대전 | 2% ( 33)| 2% ( 25)|  
## +----------------------------+--------+--------+  
## | 부산 | 3% ( 49)| 4% ( 57)|  
## +----------------------------+--------+--------+  
## | 서울 | 9% (136)| 7% (100)|  
## +----------------------------+--------+--------+  
## | 세종 | 0% ( 6)| 0% ( 6)|  
## +----------------------------+--------+--------+  
## | 울산 | 2% ( 29)| 1% ( 15)|  
## +----------------------------+--------+--------+  
## | 인천 | 3% ( 40)| 2% ( 31)|  
## +----------------------------+--------+--------+  
## | 전남 | 8% (116)|10% (143)|  
## +----------------------------+--------+--------+  
## | 전북 | 6% ( 92)| 7% (105)|  
## +----------------------------+--------+--------+  
## | 제주 | 2% ( 27)| 2% ( 28)|  
## +----------------------------+--------+--------+  
## | 충남 | 8% (115)| 9% (129)|  
## +----------------------------+--------+--------+  
## | 충북 | 6% ( 81)| 5% ( 67)|  
## +----------------------------+--------+--------+  
## |사고유형\_대분류 : 차대사람 |46% (666)|33% (464)|  
## +----------------------------+--------+--------+  
## | 차대차 |34% (483)|46% (640)|  
## +----------------------------+--------+--------+  
## | 차량단독 |20% (285)|21% (302)|  
## +----------------------------+--------+--------+  
## |사고유형\_중분류 : 공작물충돌| 8% (111)| 5% ( 74)|  
## +----------------------------+--------+--------+  
## | 기타 |18% (264)|16% (223)|  
## +----------------------------+--------+--------+  
## | 길가장자리구역통행중 | 1% ( 20)| 2% ( 22)|  
## +----------------------------+--------+--------+  
## | 도로이탈 | 1% ( 21)| 2% ( 33)|  
## +----------------------------+--------+--------+  
## | 보도통행중 | 1% ( 9)| 1% ( 16)|  
## +----------------------------+--------+--------+  
## | 전도 | 3% ( 43)| 5% ( 69)|  
## +----------------------------+--------+--------+  
## | 전도전복 | 5% ( 72)| 6% ( 83)|  
## +----------------------------+--------+--------+  
## | 전복 | 0% ( 3)| 1% ( 8)|  
## +----------------------------+--------+--------+  
## | 정면충돌 | 5% ( 78)| 9% (122)|  
## +----------------------------+--------+--------+  
## | 주/정차차량 충돌 | 0% ( 2)| 0% ( 0)|  
## +----------------------------+--------+--------+  
## | 차도통행중 | 6% ( 82)| 3% ( 46)|  
## +----------------------------+--------+--------+  
## | 추돌 |11% (151)| 9% (132)|  
## +----------------------------+--------+--------+  
## | 측면충돌 |11% (162)|21% (292)|  
## +----------------------------+--------+--------+  
## | 횡단중 |29% (415)|20% (284)|  
## +----------------------------+--------+--------+  
## | 후진중충돌 | 0% ( 1)| 0% ( 2)|  
## +----------------------------+--------+--------+  
## |도로형태 : 고가도로위 | 1% ( 12)| 0% ( 2)|  
## +----------------------------+--------+--------+  
## | 교량위 | 2% ( 23)| 1% ( 19)|  
## +----------------------------+--------+--------+  
## | 교차로내 |16% (231)|22% (306)|  
## +----------------------------+--------+--------+  
## | 교차로부근 |11% (163)|10% (138)|  
## +----------------------------+--------+--------+  
## | 교차로횡단보도내 | 5% ( 70)| 4% ( 58)|  
## +----------------------------+--------+--------+  
## | 기타 | 2% ( 31)| 3% ( 41)|  
## +----------------------------+--------+--------+  
## | 기타단일로 |62% (886)|59% (827)|  
## +----------------------------+--------+--------+  
## | 불명 | 0% ( 2)| 0% ( 3)|  
## +----------------------------+--------+--------+  
## | 주차장 | 0% ( 4)| 0% ( 1)|  
## +----------------------------+--------+--------+  
## | 지하차도(도로)내 | 0% ( 4)| 0% ( 2)|  
## +----------------------------+--------+--------+  
## | 터널안 | 1% ( 8)| 1% ( 9)|  
## +----------------------------+--------+--------+

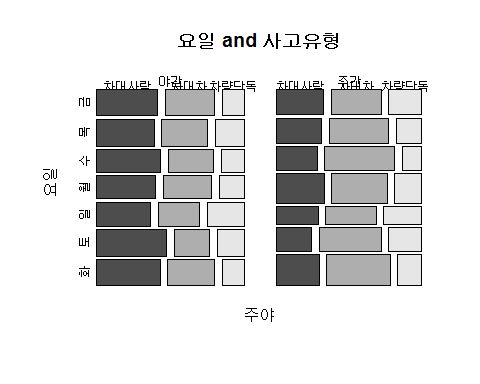
### featureplot를 이용한 데이터 시각화

Data\_complete<-x$Fold1$train[complete.cases(x$Fold1$train),]  
featurePlot(  
 Data\_complete[,  
 sapply(names(Data\_complete),  
 function(n){is.numeric(Data\_complete[,n])})],  
 Data\_complete[,c('주야')],  
 "ellipse")



### mosaicplot를 이용한 데이터 시각화

mosaicplot(주야~요일+사고유형\_대분류,data=x$Fold1$train,color=TRUE,main='요일 and 사고유형')



### xtabs를 이용한 사망자수

xtabs(~요일+사고유형\_대분류,data=x$Fold1$train)

## 사고유형\_대분류  
## 요일 차대사람 차대차 차량단독  
## 금 174 160 88  
## 목 164 167 87  
## 수 152 169 62  
## 월 168 168 83  
## 일 126 118 96  
## 토 166 148 92  
## 화 180 193 79

xtabs(주야=="야간"~요일+사고유형\_대분류,data=x$Fold1$train)/xtabs(~요일+사고유형\_대분류,data=x$Fold1$train)

## 사고유형\_대분류  
## 요일 차대사람 차대차 차량단독  
## 금 0.5747126 0.5062500 0.4090909  
## 목 0.5792683 0.4550898 0.5517241  
## 수 0.6052632 0.3846154 0.5322581  
## 월 0.4940476 0.4047619 0.4216867  
## 일 0.6349206 0.5169492 0.5625000  
## 토 0.6927711 0.3918919 0.4782609  
## 화 0.5611111 0.3834197 0.4430380

xtabs(주야=="주간"~요일+사고유형\_대분류,data=x$Fold1$train)/xtabs(~요일+사고유형\_대분류,data=x$Fold1$train)

## 사고유형\_대분류  
## 요일 차대사람 차대차 차량단독  
## 금 0.4252874 0.4937500 0.5909091  
## 목 0.4207317 0.5449102 0.4482759  
## 수 0.3947368 0.6153846 0.4677419  
## 월 0.5059524 0.5952381 0.5783133  
## 일 0.3650794 0.4830508 0.4375000  
## 토 0.3072289 0.6081081 0.5217391  
## 화 0.4388889 0.6165803 0.5569620

### 평가 메트릭

##### rpart 모델

library(rpart)  
m<-rpart(주야~요일+발생지시도+사고유형\_대분류+사고유형\_중분류+도로형태,data=train\_Data)  
p<-predict(m,newdata=train\_Data,type='class')  
head(p)

## 2 3 4 6 7 8   
## 야간 야간 야간 야간 야간 주간   
## Levels: 야간 주간

### foreach 사용하여 분석

library(foreach)

## Warning: package 'foreach' was built under R version 3.4.4

folds<-create\_ten\_fold\_cv()  
rpart\_result<-foreach(f=folds) %do% {  
 model\_rpart<-rpart(  
 주야~요일+발생지시도+사고유형\_대분류+사고유형\_중분류+도로형태, data=f$train)  
 predicted<-predict(model\_rpart, newdata=f$validation,  
 type="class")  
 return(list(actual=f$validation$주야,predicted=predicted))  
}  
head(rpart\_result)

## [[1]]  
## [[1]]$actual  
## [1] 야간 주간 주간 주간  
## Levels: 야간 주간  
##   
## [[1]]$predicted  
## 10 17 20 21   
## 야간 주간 야간 주간   
## Levels: 야간 주간  
##   
##   
## [[2]]  
## [[2]]$actual  
## [1] 주간  
## Levels: 야간 주간  
##   
## [[2]]$predicted  
## 14   
## 야간   
## Levels: 야간 주간  
##   
##   
## [[3]]  
## [[3]]$actual  
## [1] 야간 주간  
## Levels: 야간 주간  
##   
## [[3]]$predicted  
## 9 22   
## 주간 야간   
## Levels: 야간 주간  
##   
##   
## [[4]]  
## [[4]]$actual  
## [1] 주간  
## Levels: 야간 주간  
##   
## [[4]]$predicted  
## 18   
## 주간   
## Levels: 야간 주간  
##   
##   
## [[5]]  
## [[5]]$actual  
## [1] 야간 야간 주간  
## Levels: 야간 주간  
##   
## [[5]]$predicted  
## 11 12 24   
## 주간 야간 야간   
## Levels: 야간 주간  
##   
##   
## [[6]]  
## [[6]]$actual  
## [1] 야간 주간 주간  
## Levels: 야간 주간  
##   
## [[6]]$predicted  
## 6 8 25   
## 야간 주간 야간   
## Levels: 야간 주간

### Accuracy 평가

##### evaluation함수 생성

evaluation<-function(lst){  
 accuracy<-sapply(lst,function(one\_result){ #벡터로 묶음  
 return(sum(one\_result$predicted==one\_result$actual)  
 /NROW(one\_result$actual))  
 })  
 print(sprintf("MEAN +/- SD: %.3f +/- %.3f",  
 mean(accuracy),sd(accuracy))) #평균, 표준편차  
 return(accuracy)  
}  
rpart\_accuracy<-evaluation(rpart\_result) #Accuracy 계산결과 rpart 모델성능 판단

## [1] "MEAN +/- SD: 0.562 +/- 0.405"

### ctree()를 사용한 교차 검증

library(party)

## Warning: package 'party' was built under R version 3.4.4

## Loading required package: grid

## Loading required package: mvtnorm

## Loading required package: modeltools

## Loading required package: stats4

## Loading required package: strucchange

## Warning: package 'strucchange' was built under R version 3.4.4

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.4.4

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: sandwich

## Warning: package 'sandwich' was built under R version 3.4.4

ctree\_result<-foreach(f=folds)%do% {  
 model\_ctree<-ctree(  
 주야~요일+발생지시도+사고유형\_대분류+사고유형\_중분류+도로형태,data=f$train)  
 predicted<-predict(model\_ctree,newdata=f$validation,  
 type='response')  
 return(list(actual=f$validation$주야,predicted=predicted))  
}  
(ctree\_accuracy<-evaluation(ctree\_result))

## [1] "MEAN +/- SD: 0.625 +/- 0.339"

## [1] 0.7500000 0.0000000 0.5000000 1.0000000 0.6666667 0.3333333 1.0000000  
## [8] 0.7500000

### 정확도 분포

plot(density(rpart\_accuracy),main='rpart VS ctree')  
lines(density(ctree\_accuracy),col='red',lty='dashed')

