FinalProject

Jaegeun

# 교통사고 사망자의 분류 분석

## 

### 데이터 불러오기

csv\_Data<-read.csv('data//교통사고정보//2017\_교통사고정보.csv')   
csv\_Data2<-read.csv('data//교통사고정보//2016\_교통사고정보.csv')   
csv\_Data3<-read.csv('data//교통사고정보//2015\_교통사고정보.csv')   
csv\_Data4<-read.csv('data//교통사고정보//2012\_2014\_교통사고정보.csv')

### 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

bind\_Data<-bind\_rows(csv\_Data,csv\_Data2,csv\_Data3,csv\_Data4)

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

## Warning in bind\_rows\_(x, .id): Unequal factor levels: coercing to character

## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector  
  
## Warning in bind\_rows\_(x, .id): binding character and factor vector,  
## coercing into character vector

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

### 데이터 확인

str(Data)

## 'data.frame': 27254 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 : chr "장성군" "송파구" "홍성군" "아산시" ...  
## $ 사고유형\_대분류: chr "차대차" "차대사람" "차대차" "차량단독" ...  
## $ 사고유형\_중분류: chr "기타" "횡단중" "추돌" "공작물충돌" ...  
## $ 법규위반\_대분류: chr "운전자법규위반" "운전자법규위반" "운전자법규위반" "운전자법규위반" ...  
## $ 법규위반 : chr "안전운전 의무 불이행" "안전운전 의무 불이행" "안전운전 의무 불이행" "안전운전 의무 불이행" ...  
## $ 도로형태\_대분류: chr "단일로" "교차로" "단일로" "단일로" ...  
## $ 도로형태 : chr "기타단일로" "교차로부근" "기타단일로" "기타단일로" ...  
## $ 당사자종별\_1당 : chr "승용차" "승용차" "승용차" "승용차" ...  
## $ 당사자종별\_2당 : chr "승용차" "보행자" "승용차" "없음" ...  
## $ 발생위치X\_UTMK : num 933501 967570 916497 961608 946778 ...  
## $ 발생위치Y\_UTMK : num 1700129 1944453 1842880 1864573 1941695 ...  
## $ 경도 : num 127 127 127 127 127 ...  
## $ 위도 : num 35.3 37.5 36.6 36.8 37.5 ...

### 데이터 타입 형식 변환

Data$발생지시도<-as.factor(Data$발생지시도)  
Data$사고유형\_대분류<-as.factor(Data$사고유형\_대분류)  
Data$사고유형\_중분류<-as.factor(Data$사고유형\_중분류)  
Data$법규위반\_대분류<-as.factor(Data$법규위반\_대분류)  
Data$법규위반<-as.factor(Data$법규위반)  
Data$도로형태\_대분류<-as.factor(Data$도로형태\_대분류)  
Data$도로형태<-as.factor(Data$도로형태)  
Data$당사자종별\_1당<-as.factor(Data$당사자종별\_1당)  
Data$당사자종별\_2당<-as.factor(Data$당사자종별\_2당)

### 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] 8177

prop.table(table(test\_Data$주야))

##   
## 야간 주간   
## 0.5151033 0.4848967

NROW(train\_Data)

## [1] 19077

prop.table(table(train\_Data$주야))

##   
## 야간 주간   
## 0.5150705 0.4849295

### 교차검증 함수 생성

create\_ten\_fold\_cv<-function(){  
 set.seed(117)  
 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 9  
## $ Fold1:List of 2  
## ..$ train :'data.frame': 19073 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19073] 1 1 1 1 2 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19073] 2 2 2 7 2 2 1 1 1 2 ...  
## .. ..$ 중상자수 : int [1:19073] 1 0 1 4 0 0 0 0 0 1 ...  
## .. ..$ 경상자수 : int [1:19073] 0 1 0 2 0 1 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19073] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 13 11 14 15 9 4 ...  
## .. ..$ 발생지시군구 : chr [1:19073] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 4 2 2 2 4 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 8 14 18 18 8 17 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 11 2 5 5 11 15 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 6 3 6 3 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 9 4 9 4 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 18 14 11 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 16 9 9 9 16 14 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19073] 933501 916497 1021195 962530 905827 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19073] 1700129 1842880 1985798 1823895 1665350 ...  
## .. ..$ 경도 : num [1:19073] 127 127 128 127 126 ...  
## .. ..$ 위도 : num [1:19073] 35.3 36.6 37.9 36.4 35 ...  
## ..$ validation:'data.frame': 4 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 2 2 1  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 4 4 4 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 4 2 13  
## .. ..$ 발생지시군구 : chr [1:4] "은평구" "울진군" "안성시" "순천시"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 2 2 2 3  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 18 18 3 17  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 5 11 11 9  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 3 6 6 3  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 4 9 9 4  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 12 11 13  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 9 9 9 23  
## .. ..$ 발생위치X\_UTMK : num [1:4] 948413 1166337 978913 998903  
## .. ..$ 발생위치Y\_UTMK : num [1:4] 1957900 1901018 1900967 1664061  
## .. ..$ 경도 : num [1:4] 127 129 127 127  
## .. ..$ 위도 : num [1:4] 37.6 37.1 37.1 35  
## $ Fold2:List of 2  
## ..$ train :'data.frame': 19076 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 4 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19076] 1 1 1 2 1 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19076] 2 2 7 2 2 1 1 1 1 2 ...  
## .. ..$ 중상자수 : int [1:19076] 1 0 4 0 0 0 0 0 0 1 ...  
## .. ..$ 경상자수 : int [1:19076] 0 1 2 0 1 0 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19076] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 16 13 11 9 14 15 9 4 ...  
## .. ..$ 발생지시군구 : chr [1:19076] "장성군" "홍성군" "공주시" "무안군" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 3 4 2 2 2 2 4 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 3 8 14 18 18 18 8 17 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 2 5 5 5 11 15 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 3 3 6 3 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 4 4 9 4 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 18 14 11 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 14 16 9 9 9 9 16 14 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19076] 933501 916497 962530 905827 1162551 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19076] 1700129 1842880 1823895 1665350 1720013 ...  
## .. ..$ 경도 : num [1:19076] 127 127 127 126 129 ...  
## .. ..$ 위도 : num [1:19076] 35.3 36.6 36.4 35 35.5 ...  
## ..$ validation:'data.frame': 1 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5  
## .. ..$ 사망자수 : int 1  
## .. ..$ 사상자수 : int 2  
## .. ..$ 중상자수 : int 1  
## .. ..$ 경상자수 : int 0  
## .. ..$ 부상신고자수 : int 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 1  
## .. ..$ 발생지시군구 : chr "춘천시"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 2  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 18  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 9  
## .. ..$ 발생위치X\_UTMK : num 1021195  
## .. ..$ 발생위치Y\_UTMK : num 1985798  
## .. ..$ 경도 : num 128  
## .. ..$ 위도 : num 37.9  
## $ Fold3:List of 2  
## ..$ train :'data.frame': 19076 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 4 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19076] 1 1 1 2 1 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19076] 2 2 7 2 2 1 1 1 1 2 ...  
## .. ..$ 중상자수 : int [1:19076] 1 1 4 0 0 0 0 0 0 1 ...  
## .. ..$ 경상자수 : int [1:19076] 0 0 2 0 1 0 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19076] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 1 16 13 11 9 14 15 9 4 ...  
## .. ..$ 발생지시군구 : chr [1:19076] "장성군" "춘천시" "공주시" "무안군" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 2 3 4 2 2 2 2 4 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 18 3 8 14 18 18 18 8 17 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 2 5 5 5 11 15 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 3 3 6 3 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 4 4 9 4 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 18 14 11 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 9 14 16 9 9 9 9 16 14 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19076] 933501 1021195 962530 905827 1162551 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19076] 1700129 1985798 1823895 1665350 1720013 ...  
## .. ..$ 경도 : num [1:19076] 127 128 127 126 129 ...  
## .. ..$ 위도 : num [1:19076] 35.3 37.9 36.4 35 35.5 ...  
## ..$ validation:'data.frame': 1 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5  
## .. ..$ 사망자수 : int 1  
## .. ..$ 사상자수 : int 2  
## .. ..$ 중상자수 : int 0  
## .. ..$ 경상자수 : int 1  
## .. ..$ 부상신고자수 : int 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 16  
## .. ..$ 발생지시군구 : chr "홍성군"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 15  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14  
## .. ..$ 발생위치X\_UTMK : num 916497  
## .. ..$ 발생위치Y\_UTMK : num 1842880  
## .. ..$ 경도 : num 127  
## .. ..$ 위도 : num 36.6  
## $ Fold4:List of 2  
## ..$ train :'data.frame': 19076 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 1 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19076] 1 1 1 1 2 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19076] 2 2 2 7 2 2 1 1 1 1 ...  
## .. ..$ 중상자수 : int [1:19076] 1 0 1 4 0 0 0 0 0 0 ...  
## .. ..$ 경상자수 : int [1:19076] 0 1 0 2 0 1 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19076] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 13 11 9 14 15 9 ...  
## .. ..$ 발생지시군구 : chr [1:19076] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 4 2 2 2 2 4 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 8 14 18 18 18 8 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 11 2 5 5 5 11 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 6 3 3 6 3 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 9 4 4 9 4 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 11 18 14 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 16 9 9 9 9 16 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19076] 933501 916497 1021195 962530 905827 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19076] 1700129 1842880 1985798 1823895 1665350 ...  
## .. ..$ 경도 : num [1:19076] 127 127 128 127 126 ...  
## .. ..$ 위도 : num [1:19076] 35.3 36.6 37.9 36.4 35 ...  
## ..$ validation:'data.frame': 1 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 7  
## .. ..$ 사망자수 : int 1  
## .. ..$ 사상자수 : int 1  
## .. ..$ 중상자수 : int 0  
## .. ..$ 경상자수 : int 0  
## .. ..$ 부상신고자수 : int 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 1  
## .. ..$ 발생지시군구 : chr "원주시"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 2  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 18  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 5  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 3  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 5  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 9  
## .. ..$ 발생위치X\_UTMK : num 1040515  
## .. ..$ 발생위치Y\_UTMK : num 1927745  
## .. ..$ 경도 : num 128  
## .. ..$ 위도 : num 37.3  
## $ Fold5:List of 2  
## ..$ train :'data.frame': 19073 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 4 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19073] 1 1 1 1 1 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19073] 2 2 2 7 2 1 1 1 1 1 ...  
## .. ..$ 중상자수 : int [1:19073] 1 0 1 4 0 0 0 0 0 0 ...  
## .. ..$ 경상자수 : int [1:19073] 0 1 0 2 1 0 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19073] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 11 9 15 9 3 4 ...  
## .. ..$ 발생지시군구 : chr [1:19073] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 2 2 2 4 3 2 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 14 18 18 8 15 18 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 2 5 5 11 11 11 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 3 6 3 6 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 4 9 4 9 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 18 14 13 12 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 9 9 9 16 23 9 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19073] 933501 916497 1021195 962530 1162551 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19073] 1700129 1842880 1985798 1823895 1720013 ...  
## .. ..$ 경도 : num [1:19073] 127 127 128 127 129 ...  
## .. ..$ 위도 : num [1:19073] 35.3 36.6 37.9 36.4 35.5 ...  
## ..$ 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 7  
## .. ..$ 사망자수 : int [1:4] 2 1 1 2  
## .. ..$ 사상자수 : int [1:4] 2 1 2 4  
## .. ..$ 중상자수 : int [1:4] 0 0 1 0  
## .. ..$ 경상자수 : int [1:4] 0 0 0 0  
## .. ..$ 부상신고자수 : int [1:4] 0 0 0 2  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 14 4 8  
## .. ..$ 발생지시군구 : chr [1:4] "무안군" "전주시" "청송군" "강서구"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 4 2 3 3  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 8 18 17 17  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 5 15 10  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 3 6 6  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 4 9 9  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 18  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 16 9 14 14  
## .. ..$ 발생위치X\_UTMK : num [1:4] 905827 966955 1134705 1126079  
## .. ..$ 발생위치Y\_UTMK : num [1:4] 1665350 1754577 1810932 1685284  
## .. ..$ 경도 : num [1:4] 126 127 129 129  
## .. ..$ 위도 : num [1:4] 35 35.8 36.3 35.2  
## $ Fold6:List of 2  
## ..$ train :'data.frame': 19075 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19075] 1 1 1 1 2 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19075] 2 2 2 7 2 1 1 1 1 2 ...  
## .. ..$ 중상자수 : int [1:19075] 1 0 1 4 0 0 0 0 0 1 ...  
## .. ..$ 경상자수 : int [1:19075] 0 1 0 2 0 0 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19075] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 13 9 14 15 9 4 ...  
## .. ..$ 발생지시군구 : chr [1:19075] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 4 2 2 2 4 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 8 18 18 18 8 17 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 11 5 5 5 11 15 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 3 3 6 3 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 4 4 9 4 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 18 14 11 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 16 9 9 9 16 14 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19075] 933501 916497 1021195 962530 905827 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19075] 1700129 1842880 1985798 1823895 1665350 ...  
## .. ..$ 경도 : num [1:19075] 127 127 128 127 126 ...  
## .. ..$ 위도 : num [1:19075] 35.3 36.6 37.9 36.4 35 ...  
## ..$ 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",..: 4 4  
## .. ..$ 사망자수 : int [1:2] 1 1  
## .. ..$ 사상자수 : int [1:2] 2 1  
## .. ..$ 중상자수 : int [1:2] 0 0  
## .. ..$ 경상자수 : int [1:2] 1 0  
## .. ..$ 부상신고자수 : int [1:2] 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 11 3  
## .. ..$ 발생지시군구 : chr [1:2] "울주군" "양산시"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 2 3  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 14 15  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 2 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 13  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 9 23  
## .. ..$ 발생위치X\_UTMK : num [1:2] 1162551 1138995  
## .. ..$ 발생위치Y\_UTMK : num [1:2] 1720013 1709400  
## .. ..$ 경도 : num [1:2] 129 129  
## .. ..$ 위도 : num [1:2] 35.5 35.4  
## $ Fold7:List of 2  
## ..$ train :'data.frame': 19076 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 1 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19076] 1 1 1 1 2 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19076] 2 2 2 7 2 2 1 1 1 2 ...  
## .. ..$ 중상자수 : int [1:19076] 1 0 1 4 0 0 0 0 0 1 ...  
## .. ..$ 경상자수 : int [1:19076] 0 1 0 2 0 1 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19076] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 13 11 9 14 15 4 ...  
## .. ..$ 발생지시군구 : chr [1:19076] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 4 2 2 2 2 3 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 8 14 18 18 18 17 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 11 2 5 5 5 15 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 6 3 3 6 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 9 4 4 9 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 11 18 11 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 16 9 9 9 9 14 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19076] 933501 916497 1021195 962530 905827 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19076] 1700129 1842880 1985798 1823895 1665350 ...  
## .. ..$ 경도 : num [1:19076] 127 127 128 127 126 ...  
## .. ..$ 위도 : num [1:19076] 35.3 36.6 37.9 36.4 35 ...  
## ..$ 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",..: 9  
## .. ..$ 발생지시군구 : chr "영등포구"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 4  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 8  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 3  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 4  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 14  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 16  
## .. ..$ 발생위치X\_UTMK : num 946757  
## .. ..$ 발생위치Y\_UTMK : num 1943309  
## .. ..$ 경도 : num 127  
## .. ..$ 위도 : num 37.5  
## $ Fold8:List of 2  
## ..$ train :'data.frame': 19072 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 2 2 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 4 4 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19072] 1 1 2 1 1 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19072] 2 2 2 2 1 1 1 2 1 1 ...  
## .. ..$ 중상자수 : int [1:19072] 0 1 0 0 0 0 0 1 0 0 ...  
## .. ..$ 경상자수 : int [1:19072] 1 0 0 1 0 0 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19072] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 16 1 13 11 9 14 9 4 3 4 ...  
## .. ..$ 발생지시군구 : chr [1:19072] "홍성군" "춘천시" "무안군" "울주군" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 2 4 2 2 2 4 3 3 2 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 15 18 8 14 18 18 8 17 15 18 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 2 5 5 11 15 11 11 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 3 3 3 6 6 6 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 4 4 4 9 9 9 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 14 11 13 12 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 9 16 9 9 9 16 14 23 9 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19072] 916497 1021195 905827 1162551 948413 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19072] 1842880 1985798 1665350 1720013 1957900 ...  
## .. ..$ 경도 : num [1:19072] 127 128 126 129 127 ...  
## .. ..$ 위도 : num [1:19072] 36.6 37.9 35 35.5 37.6 ...  
## ..$ validation:'data.frame': 5 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 2 1 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 4 7 7  
## .. ..$ 사망자수 : int [1:5] 1 1 1 1 1  
## .. ..$ 사상자수 : int [1:5] 2 7 1 1 1  
## .. ..$ 중상자수 : int [1:5] 1 4 0 0 0  
## .. ..$ 경상자수 : int [1:5] 0 2 0 0 0  
## .. ..$ 부상신고자수 : int [1:5] 0 0 0 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 15 2 16  
## .. ..$ 발생지시군구 : chr [1:5] "장성군" "공주시" "제주시" "안성시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 4 3  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 3 18 3 17  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 5 11 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 3  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 4  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 18 14 13  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 16 14  
## .. ..$ 발생위치X\_UTMK : num [1:5] 933501 962530 910531 975120 972938  
## .. ..$ 발생위치Y\_UTMK : num [1:5] 1700129 1823895 1502842 1889560 1875546  
## .. ..$ 경도 : num [1:5] 127 127 127 127 127  
## .. ..$ 위도 : num [1:5] 35.3 36.4 33.5 37 36.9  
## $ Fold9:List of 2  
## ..$ train :'data.frame': 19075 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 1 1 1 1 1 1 1 2 2 2 ...  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 5 5 5 5 5 4 4 4 4 4 ...  
## .. ..$ 사망자수 : int [1:19075] 1 1 1 1 2 1 1 1 1 1 ...  
## .. ..$ 사상자수 : int [1:19075] 2 2 2 7 2 2 1 1 1 1 ...  
## .. ..$ 중상자수 : int [1:19075] 1 0 1 4 0 0 0 0 0 0 ...  
## .. ..$ 경상자수 : int [1:19075] 0 1 0 2 0 1 0 0 0 0 ...  
## .. ..$ 부상신고자수 : int [1:19075] 0 0 0 0 0 0 0 0 0 0 ...  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 13 16 1 16 13 11 9 14 15 9 ...  
## .. ..$ 발생지시군구 : chr [1:19075] "장성군" "홍성군" "춘천시" "공주시" ...  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 3 2 3 4 2 2 2 2 4 ...  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 3 15 18 3 8 14 18 18 18 8 ...  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2 2 2 2 2 2 2 2 2 ...  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 11 11 11 11 11 2 5 5 5 11 ...  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 6 6 6 6 6 6 3 3 6 3 ...  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 9 9 9 9 9 9 4 4 9 4 ...  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11 11 11 11 11 11 11 18 14 ...  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 14 14 9 14 16 9 9 9 9 16 ...  
## .. ..$ 발생위치X\_UTMK : num [1:19075] 933501 916497 1021195 962530 905827 ...  
## .. ..$ 발생위치Y\_UTMK : num [1:19075] 1700129 1842880 1985798 1823895 1665350 ...  
## .. ..$ 경도 : num [1:19075] 127 127 128 127 126 ...  
## .. ..$ 위도 : num [1:19075] 35.3 36.6 37.9 36.4 35 ...  
## ..$ validation:'data.frame': 2 obs. of 21 variables:  
## .. ..$ 주야 : Factor w/ 2 levels "야간","주간": 2 2  
## .. ..$ 요일 : Factor w/ 7 levels "금","목","수",..: 4 4  
## .. ..$ 사망자수 : int [1:2] 1 1  
## .. ..$ 사상자수 : int [1:2] 5 1  
## .. ..$ 중상자수 : int [1:2] 0 0  
## .. ..$ 경상자수 : int [1:2] 4 0  
## .. ..$ 부상신고자수 : int [1:2] 0 0  
## .. ..$ 발생지시도 : Factor w/ 17 levels "강원","경기",..: 9 9  
## .. ..$ 발생지시군구 : chr [1:2] "금천구" "광진구"  
## .. ..$ 사고유형\_대분류: Factor w/ 4 levels "건널목","차대사람",..: 3 4  
## .. ..$ 사고유형\_중분류: Factor w/ 19 levels "경보기무시","공작물충돌",..: 17 2  
## .. ..$ 법규위반\_대분류: Factor w/ 3 levels "보행자과실","운전자법규위반",..: 2 2  
## .. ..$ 법규위반 : Factor w/ 20 levels "과로","과속",..: 9 11  
## .. ..$ 도로형태\_대분류: Factor w/ 9 levels "건널목","고가도로위",..: 3 6  
## .. ..$ 도로형태 : Factor w/ 16 levels "건널목","고가도로위",..: 4 9  
## .. ..$ 당사자종별\_1당 : Factor w/ 18 levels "개인형이동수단(PM)",..: 11 11  
## .. ..$ 당사자종별\_2당 : Factor w/ 23 levels "","0","00","건설기계",..: 15 16  
## .. ..$ 발생위치X\_UTMK : num [1:2] 947235 964223  
## .. ..$ 발생위치Y\_UTMK : num [1:2] 1938476 1950661  
## .. ..$ 경도 : num [1:2] 127 127  
## .. ..$ 위도 : num [1:2] 37.4 37.6

## NULL

head(x$Fold1$train)

## 주야 요일 사망자수 사상자수 중상자수 경상자수 부상신고자수 발생지시도  
## 1 야간 일 1 2 1 0 0 전남  
## 3 야간 일 1 2 0 1 0 충남  
## 6 야간 일 1 2 1 0 0 강원  
## 10 야간 일 1 7 4 2 0 충남  
## 11 야간 일 2 2 0 0 0 전남  
## 12 야간 월 1 2 0 1 0 울산  
## 발생지시군구 사고유형\_대분류 사고유형\_중분류 법규위반\_대분류  
## 1 장성군 차대차 기타 운전자법규위반  
## 3 홍성군 차대차 추돌 운전자법규위반  
## 6 춘천시 차대사람 횡단중 운전자법규위반  
## 10 공주시 차대차 기타 운전자법규위반  
## 11 무안군 차량단독 전도전복 운전자법규위반  
## 12 울주군 차대사람 차도통행중 운전자법규위반  
## 법규위반 도로형태\_대분류 도로형태 당사자종별\_1당  
## 1 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 3 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 6 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 10 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 11 안전운전 의무 불이행 단일로 기타단일로 승용차  
## 12 과속 단일로 기타단일로 승용차  
## 당사자종별\_2당 발생위치X\_UTMK 발생위치Y\_UTMK 경도 위도  
## 1 승용차 933501 1700129 126.7686 35.29446  
## 3 승용차 916497 1842880 126.5666 36.58007  
## 6 보행자 1021195 1985798 127.7410 37.87175  
## 10 승용차 962530 1823895 127.0821 36.41184  
## 11 없음 905827 1665350 126.4683 34.97880  
## 12 보행자 1162551 1720013 129.2914 35.46267

### 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

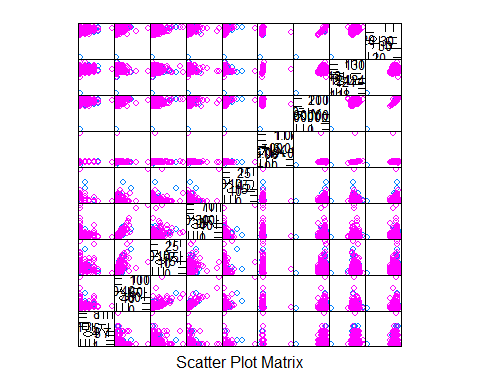
## The following objects are masked from 'package:base':  
##   
## format.pval, units

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

##   
##   
## Descriptive Statistics by 주야  
##   
## +----------------------------+---------+---------+  
## | |야간 |주간 |  
## | |(N=9824) |(N=9249) |  
## +----------------------------+---------+---------+  
## |요일 : 금 |15% (1446)|15% (1376)|  
## +----------------------------+---------+---------+  
## | 목 |14% (1372)|14% (1323)|  
## +----------------------------+---------+---------+  
## | 수 |14% (1333)|14% (1317)|  
## +----------------------------+---------+---------+  
## | 월 |14% (1346)|16% (1481)|  
## +----------------------------+---------+---------+  
## | 일 |14% (1391)|11% (1039)|  
## +----------------------------+---------+---------+  
## | 토 |15% (1522)|14% (1307)|  
## +----------------------------+---------+---------+  
## | 화 |14% (1414)|15% (1406)|  
## +----------------------------+---------+---------+  
## |발생지시도 : 강원 | 4% ( 415)| 5% ( 508)|  
## +----------------------------+---------+---------+  
## | 경기 |20% (1992)|17% (1587)|  
## +----------------------------+---------+---------+  
## | 경남 | 8% ( 796)| 9% ( 841)|  
## +----------------------------+---------+---------+  
## | 경북 |10% ( 954)|11% (1061)|  
## +----------------------------+---------+---------+  
## | 광주 | 3% ( 258)| 2% ( 175)|  
## +----------------------------+---------+---------+  
## | 대구 | 4% ( 395)| 3% ( 282)|  
## +----------------------------+---------+---------+  
## | 대전 | 2% ( 228)| 2% ( 168)|  
## +----------------------------+---------+---------+  
## | 부산 | 4% ( 416)| 4% ( 325)|  
## +----------------------------+---------+---------+  
## | 서울 | 9% ( 906)| 7% ( 641)|  
## +----------------------------+---------+---------+  
## | 세종 | 0% ( 35)| 0% ( 46)|  
## +----------------------------+---------+---------+  
## | 울산 | 2% ( 223)| 2% ( 157)|  
## +----------------------------+---------+---------+  
## | 인천 | 3% ( 324)| 3% ( 265)|  
## +----------------------------+---------+---------+  
## | 전남 | 8% ( 737)|10% ( 959)|  
## +----------------------------+---------+---------+  
## | 전북 | 6% ( 631)| 8% ( 710)|  
## +----------------------------+---------+---------+  
## | 제주 | 2% ( 192)| 2% ( 187)|  
## +----------------------------+---------+---------+  
## | 충남 | 8% ( 808)| 9% ( 808)|  
## +----------------------------+---------+---------+  
## | 충북 | 5% ( 514)| 6% ( 529)|  
## +----------------------------+---------+---------+  
## |사고유형\_대분류 : 건널목 | 0% ( 3)| 0% ( 0)|  
## +----------------------------+---------+---------+  
## | 차대사람 |47% (4583)|31% (2829)|  
## +----------------------------+---------+---------+  
## | 차대차 |33% (3243)|47% (4340)|  
## +----------------------------+---------+---------+  
## | 차량단독 |20% (1995)|22% (2080)|  
## +----------------------------+---------+---------+  
## |사고유형\_중분류 : 경보기무시| 0% ( 0)| 0% ( 0)|  
## +----------------------------+---------+---------+  
## | 공작물충돌 |10% (1004)| 8% ( 755)|  
## +----------------------------+---------+---------+  
## | 기타 |23% (2269)|23% (2136)|  
## +----------------------------+---------+---------+  
## | 길가장자리구역통행중 | 2% ( 202)| 2% ( 173)|  
## +----------------------------+---------+---------+  
## | 도로이탈 | 2% ( 206)| 3% ( 290)|  
## +----------------------------+---------+---------+  
## | 보도통행중 | 1% ( 106)| 1% ( 103)|  
## +----------------------------+---------+---------+  
## | 전도 | 0% ( 40)| 1% ( 57)|  
## +----------------------------+---------+---------+  
## | 전도전복 | 5% ( 495)| 7% ( 665)|  
## +----------------------------+---------+---------+  
## | 전복 | 0% ( 4)| 0% ( 7)|  
## +----------------------------+---------+---------+  
## | 정면충돌 | 5% ( 531)| 7% ( 686)|  
## +----------------------------+---------+---------+  
## | 주/정차차량 충돌 | 0% ( 4)| 0% ( 5)|  
## +----------------------------+---------+---------+  
## | 직전진행 | 0% ( 0)| 0% ( 0)|  
## +----------------------------+---------+---------+  
## | 차단기돌파 | 0% ( 2)| 0% ( 0)|  
## +----------------------------+---------+---------+  
## | 차도통행중 | 5% ( 454)| 2% ( 229)|  
## +----------------------------+---------+---------+  
## | 추돌 |11% (1070)|10% ( 938)|  
## +----------------------------+---------+---------+  
## | 측면직각충돌 | 8% ( 737)|15% (1418)|  
## +----------------------------+---------+---------+  
## | 측면충돌 | 2% ( 167)| 3% ( 309)|  
## +----------------------------+---------+---------+  
## | 횡단중 |26% (2532)|16% (1474)|  
## +----------------------------+---------+---------+  
## | 후진중충돌 | 0% ( 1)| 0% ( 4)|  
## +----------------------------+---------+---------+  
## |도로형태 : 건널목 | 0% ( 3)| 0% ( 0)|  
## +----------------------------+---------+---------+  
## | 고가도로위 | 0% ( 49)| 0% ( 23)|  
## +----------------------------+---------+---------+  
## | 교량위 | 1% ( 146)| 1% ( 103)|  
## +----------------------------+---------+---------+  
## | 교차로내 |16% (1580)|22% (1996)|  
## +----------------------------+---------+---------+  
## | 교차로부근 |11% (1115)|11% ( 975)|  
## +----------------------------+---------+---------+  
## | 교차로횡단보도내 | 1% ( 78)| 1% ( 55)|  
## +----------------------------+---------+---------+  
## | 기타 | 0% ( 33)| 0% ( 39)|  
## +----------------------------+---------+---------+  
## | 기타/불명 | 2% ( 201)| 2% ( 218)|  
## +----------------------------+---------+---------+  
## | 기타단일로 |62% (6050)|58% (5379)|  
## +----------------------------+---------+---------+  
## | 불명 | 0% ( 2)| 0% ( 4)|  
## +----------------------------+---------+---------+  
## | 주차장 | 0% ( 4)| 0% ( 2)|  
## +----------------------------+---------+---------+  
## | 지하도로내 | 0% ( 19)| 0% ( 20)|  
## +----------------------------+---------+---------+  
## | 지하차도(도로)내 | 0% ( 4)| 0% ( 4)|  
## +----------------------------+---------+---------+  
## | 터널안 | 0% ( 42)| 1% ( 73)|  
## +----------------------------+---------+---------+  
## | 횡단보도부근 | 1% ( 115)| 1% ( 83)|  
## +----------------------------+---------+---------+  
## | 횡단보도상 | 4% ( 383)| 3% ( 275)|  
## +----------------------------+---------+---------+

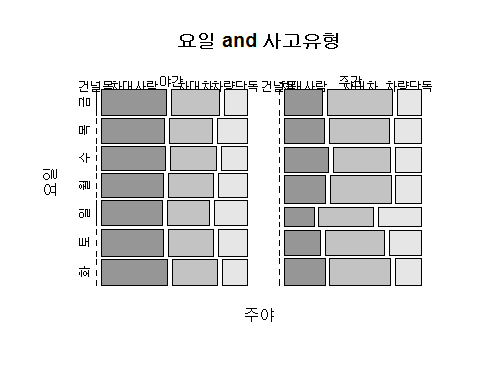
### featureplot를 이용한 데이터 시각화

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 건널목 차대사람 차대차 차량단독  
## 금 2 1107 1206 507  
## 목 0 1053 1063 579  
## 수 0 1086 1035 529  
## 월 0 1091 1154 582  
## 일 1 866 875 688  
## 토 0 1074 1105 650  
## 화 0 1135 1145 540

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

## 사고유형\_대분류  
## 요일 건널목 차대사람 차대차 차량단독  
## 금 1.0000000 0.6278229 0.4170813 0.4852071  
## 목 0.6020893 0.4167451 0.5094991  
## 수 0.5801105 0.4318841 0.4839319  
## 월 0.5627864 0.3899480 0.4845361  
## 일 1.0000000 0.7205543 0.4880000 0.4927326  
## 토 0.6517691 0.4515837 0.4969231  
## 화 0.6044053 0.4139738 0.4703704

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

## 사고유형\_대분류  
## 요일 건널목 차대사람 차대차 차량단독  
## 금 0.0000000 0.3721771 0.5829187 0.5147929  
## 목 0.3979107 0.5832549 0.4905009  
## 수 0.4198895 0.5681159 0.5160681  
## 월 0.4372136 0.6100520 0.5154639  
## 일 0.0000000 0.2794457 0.5120000 0.5072674  
## 토 0.3482309 0.5484163 0.5030769  
## 화 0.3955947 0.5860262 0.5296296

### 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)

## 1 3 6 10 11 12   
## 주간 야간 야간 주간 주간 야간   
## 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]]$actual  
## [1] 야간 주간 주간 야간  
## Levels: 야간 주간  
##   
## [[1]]$predicted  
## 13 20 24 26   
## 야간 야간 야간 주간   
## Levels: 야간 주간  
##   
##   
## [[2]]  
## [[2]]$actual  
## [1] 야간  
## Levels: 야간 주간  
##   
## [[2]]$predicted  
## 6   
## 야간   
## Levels: 야간 주간  
##   
##   
## [[3]]  
## [[3]]$actual  
## [1] 야간  
## Levels: 야간 주간  
##   
## [[3]]$predicted  
## 3   
## 야간   
## Levels: 야간 주간  
##   
##   
## [[4]]  
## [[4]]$actual  
## [1] 야간  
## Levels: 야간 주간  
##   
## [[4]]$predicted  
## 28   
## 야간   
## Levels: 야간 주간  
##   
##   
## [[5]]  
## [[5]]$actual  
## [1] 야간 주간 주간 주간  
## Levels: 야간 주간  
##   
## [[5]]$predicted  
## 11 14 18 29   
## 주간 야간 주간 주간   
## Levels: 야간 주간  
##   
##   
## [[6]]  
## [[6]]$actual  
## [1] 야간 주간  
## Levels: 야간 주간  
##   
## [[6]]$predicted  
## 12 19   
## 야간 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: 0.661 +/- 0.339"

### 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: 0.467 +/- 0.426"

## [1] 0.25 1.00 1.00 1.00 0.25 0.50 0.00 0.20 0.00

### 정확도 분포

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

