사망교통사고 정보 분석 및 시각화

프로젝트 결과 리포트

김강현, 조수빈, 최홍찬

11/25/2022

## 팀소개

* 김강현: 경북소프트웨어고등학교, 2학년, 소프트웨어개발
* 조수빈: 경북소프트웨어고등학교, 2학년, 소프트웨어개발
* 최홍찬: 경북소프트웨어고등학교, 2학년, 소프트웨어개발

## 1. 서론

1. 이 프로젝트는 전국에서 오는 저희 학교 학생들이 여러 교통수단을 이용하며 다양한 교통사고를 접할 수 있다고 생각했고 이에 따라 사망교통사고 데이터의 교통사고의 시간, 유형 등을 시각화 및 분석을 통해 학생들이 교통사고에 대한 경각심을 키우는 것을 목적으로 한다.
2. 이 레포트는 서론, 프로젝트 배경, 데이터 설명, 데이터 분석 설명, 중간 결과, 프로젝트 플랜의 순서로 전국 사망교통사고과 관련된 프로젝트를 설명한다.

### 1.1 프로젝트 배경/목적

이 프로젝트는 사망교통사고에 관련된 다양한 데이터를 비교 분석하고 시각화 하는 프로젝트이다. 저희 경북소프트웨어고등학교는 전국에서 학생들이 오기 때문에 그만큼 학생들이 귀가를 하거나, 학교에 올 때 다양한 교통수단을 이용하고 있다. 때문에 교통사고가 언제(시간), 어떻게(유형) 일어나는지 알면 도움이 될 것이라 생각했고 교통사고에 대한 데이터를 다뤄 보기로 결정했다. 그래서 2012~2020년 동안의 전국 사망교통사고의 사고년도, 위치, 사고유형, 원인 그리고 시간등을 가진 자료를 찾아 시각화를 했다. 데이터를 찾아본 후 사고유형, 원인, 날짜, 시간, 위치 데이터를 비교 분석하여 그 관련성에 대해 중점적으로 알아보고자 했다. 이 프로젝트는 학생들이 교통사고에 대한 경각심을 키우는 것이 목적이다.

### 1.2 리포트 개요

서론에서는 사망교통사고와 관련된 프로젝트를 결정하게된 배경과 목적을 설명하고 다음으로 데이터 시각화 및 분석에 사용된 데이터를 설명한 후 데이터 분석에서는 년도별 교통사고 건수, 가해자 법규 위반에 따른 사고 건수, 사고유형이 높은 2가지 유형의 사고가 높게 일어나는 시각(주야) 구분, 도로형태 별 사망자 및 부상자 비교의 분석 설명 후 각 분석에 따른 중간 결론 및, 남은 프로젝트 플랜 설명 순서로 진행된다.

## 2. 프로젝트에 사용된 데이터

### 2.1 데이터 소개/설명

시각화 및 분석에 사용된 자료는 도로교통공단의 2012년부터 2020년까지의 사망교통사고정보 데이터이다. 이 데이터는 <https://www.dploor.com/traffic/acc-death> 에서 제공하는 데이터로, 컬럼의 순서는 ‘사고년도’, ‘사망자수’, ‘부상자수’, ‘중상자수’, ‘경상자수’, ‘부상신고자수’, ‘시도’, ‘시군구’, ‘사고유형 대분류’, ‘사고유형 중분류’, ‘사고유형’, ‘가해자법규위반’, ‘도시형태 대분류’, ‘도로형태’, ‘가해당사자 종별’, ‘피해당사자 종별’, ‘주야구분’, ‘발생일’, ‘발생시간’, ‘발생요일’, ‘발생위치X좌표(EPSG5179)’, ‘발생위치Y좌표(EPSG 5179)’, ‘경도좌표’, ’위도좌표’으로 구성된다.

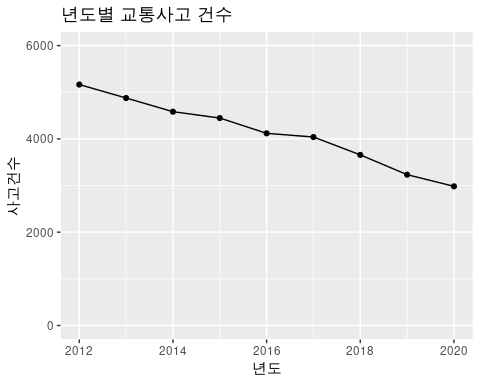
### 2.2 데이터 Dictionary

위의 컬럼 중 분석에 사용되어진 데이터는 ‘사고년도’, ‘사망자수’, ‘부상자수’, ‘사고유형’, ‘가해자법규위반’, ‘도로형태’, ‘주야구분’ 이다. ’사고년도’의 단위는 ’년’이고 2012년 부터 2020년 까지 나타내며, ’사망자수’는 ’명’으로 단위가 계산되고, 사고 후 병원에서 사망이 확인되어진 인원이다. ’부상자수’의 단위는 ’명’이고, 사고 후 병원에서 부상이 확인되어진 인원이다. ’사고유형’은 사고의 분류 유형으로, 사고 후 도로교통공단에서 분류한다. ’가해자법규위반’의 경우 사고가 발생된 사유로써, 사고 후 도로교통공단에서 분류되어진다. ’도로형태’의 경우 각 도로의 형태로, 사고가 일어난 도로를 도로교통공단에서 분류한다. ’주야구분’은 사고가 일어난 시간을 주간(09:00~18:00 9시간), 야간(18:00~익일 09:00 15시간)을 기준으로 분류한 시각이다.

## 3. 데이터 분석

### 3.1 분석 1. 년도별 사망교통사고 발생 건수

ggplot(data = count\_traffic\_accident\_by\_year, aes(x = 사고년도, y = count)) +   
 geom\_line() +   
 geom\_point() +   
 labs(title="년도별 교통사고 건수", x = "년도", y = "사고건수") +   
 scale\_y\_continuous(limits = c(0, 6000))



년도와 사건수의 상관관계분석에서 교통법 재재강화, 단속으로 인해 사건수가 2013년 5퍼센트, 2014년 6퍼센트로 2012년부터 2020년까지 약 57퍼센트가 감소됨을 확인했다.



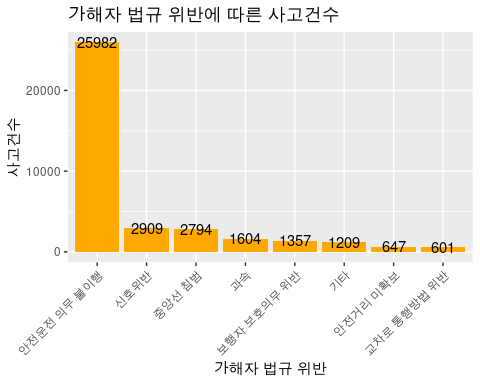
[그림1] 작년과 비교하여 교통사고가 줄어든게 체감이 되십니까?(자신이 본 교통사고 포함)

하지만 위 자료를 보면 학교 학생들의 설문조사 결과 여전히 교통사고감소를 체감하고 있는 학생들이 적었다.

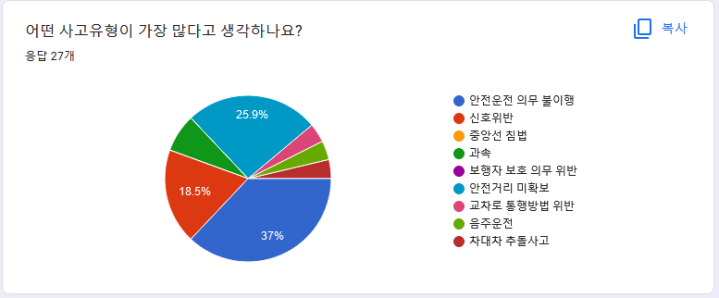
그래서 다음으로 어떤 유형의 사고건수가 가장 많은지를 알아봤다.

### 3.2 분석 2. 가해자 법규 위반에 따른 사고 건수

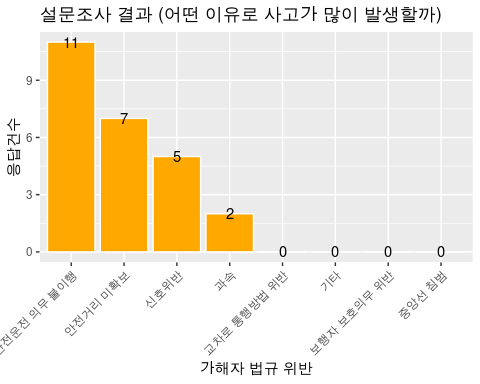
tmp1 <- traffic\_accident %>% group\_by(가해자법규위반) %>% summarise(사고건수 = n())  
ggplot(data=tmp1, aes(x=reorder(가해자법규위반, -사고건수), y=사고건수)) +   
 geom\_bar(stat = "identity", position=position\_dodge(), fill='#FFA800') +   
 geom\_text(aes(label=사고건수), color='black') +   
 theme(axis.text.x=element\_text(angle=45, hjust=1)) +   
 labs(title="가해자 법규 위반에 따른 사고건수", x='가해자 법규 위반')



가해자법규위반의 경우 안전운전 의무 불이행이 25982건로 사고건수가 가장 높았고 교차로 통행방법 위반이 601건으로 사고건수가 가장 낮았다.

 위 자료를 막대그래프로 표현하면

data <- data.frame(  
 사고유형=c('안전운전 의무 불이행', '신호위반', '중앙선 침범', '과속', '보행자 보호의무 위반', '기타', '안전거리 미확보', '교차로 통행방법 위반'),  
 응답건수=c(11, 5, 0, 2, 0, 0, 7, 0)  
)  
  
data %>% arrange(desc(응답건수)) %>%   
 ggplot(aes(reorder(x=사고유형, -응답건수),y=응답건수)) +  
 geom\_bar(stat="identity", color="white", fill='#FFA800', position=position\_dodge()) +   
 geom\_text(aes(label=응답건수), color='black') +   
 theme(axis.text.x=element\_text(angle=45, hjust=1)) +   
 labs(title="설문조사 결과 (어떤 이유로 사고가 많이 발생할까)", x='가해자 법규 위반')

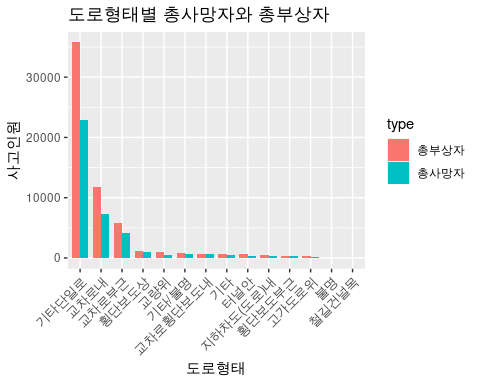


위 자료를 보면 학생들이 생각하는 교통사고의 가장 많은 사고유형은 안전운전 의무 불이행이었다.

### 3.3 분석 3. 도로형태별 사망·부상률이 가장 높은 도로형태

tmp2 <- traffic\_accident %>% group\_by(`도로형태`) %>%   
 summarise(총사망자=sum(사망자수), 총부상자=sum(부상자수)) %>%   
 arrange(desc(총사망자)) %>%   
 pivot\_longer(c(`총사망자`, `총부상자`), names\_to = "type", values\_to ="n")  
  
ggplot(data=tmp2, aes(x=reorder(`도로형태`, -n), y=n, fill=type)) +   
 geom\_bar(position = position\_dodge(0.8), stat = "identity", size=.4, width=0.8) +  
 theme(axis.text.x=element\_text(angle=45, hjust=1, vjust=1, size=10)) +   
 labs(title="도로형태별 총사망자와 총부상자", x='도로형태', y = '사고인원')

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
## ℹ Please use `linewidth` instead.



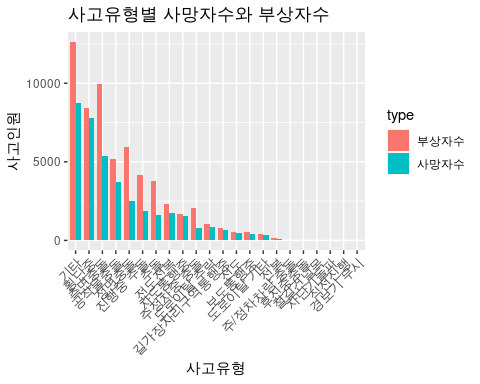
사고전체로 보았을 때는 주로 기타단일로, 교차로내, 교차로부근 순서로 많이 일어났다.

### 3.4 분석 4. 교통사고 사고 유형에 따른 사망자, 부상자 비교 분석을 통한 관계성 파악

tmp3 <- traffic\_accident %>% group\_by(`사고유형`) %>%   
 summarise(사망자수 = sum(사망자수), 부상자수 = sum(부상자수)) %>%   
 pivot\_longer(c(`사망자수`, `부상자수`), names\_to = "type", values\_to ="n")  
tmp3

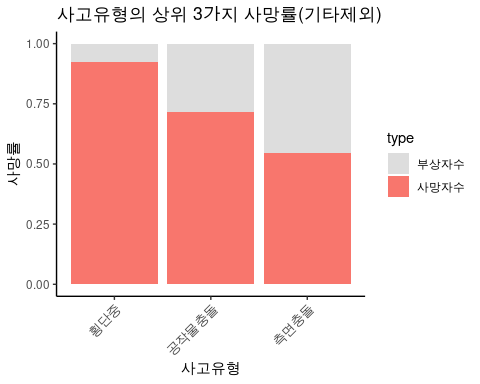
## # A tibble: 44 × 3  
## 사고유형 type n  
## <chr> <chr> <dbl>  
## 1 경보기 무시 사망자수 1  
## 2 경보기 무시 부상자수 1  
## 3 공작물충돌 사망자수 3705  
## 4 공작물충돌 부상자수 5173  
## 5 기타 사망자수 8759  
## 6 기타 부상자수 12650  
## 7 길가장자리구역통행중 사망자수 680  
## 8 길가장자리구역통행중 부상자수 760  
## 9 도로이탈 기타 사망자수 318  
## 10 도로이탈 기타 부상자수 437  
## # … with 34 more rows

ggplot(data=tmp3, aes(x=reorder(사고유형, -n), y=n, fill=type)) +   
 geom\_bar(position = position\_dodge(0.8), stat = "identity", size=.4, width=0.8) +  
 theme(axis.text.x=element\_text(angle=45, hjust=1, vjust=1, size=10)) +   
 labs(title="사고유형별 사망자수와 부상자수", x='사고유형', y = '사고인원')



사고유형별 사망자 부상자 중상자 경상자를 비교분석한 결과 기타, 횡단중, 측면충돌 순으로 사망자가 많았고 기타의 경우 정확한 사고유형을 알 수 없으니 제외한다.

tmp4 <- traffic\_accident %>% group\_by(`사고유형`) %>%   
 summarise(사망자수 = sum(사망자수),   
 부상자수 = sum(부상자수) - sum(사망자수),  
 평균 = sum(부상자수)/sum(사망자수)\*100) %>%   
 arrange(desc(사망자수)) %>%   
 filter(사고유형 == '횡단중' | 사고유형 == '측면충돌' | 사고유형 == '공작물충돌') %>%   
 pivot\_longer(c(`사망자수`, `부상자수`), names\_to = "type", values\_to ="n")  
  
  
ggplot(tmp4) +  
 geom\_bar(aes(x=reorder(`사고유형`, 평균), y = n, fill = type),   
 stat='identity', position='fill') +  
 theme\_classic() +  
 theme(axis.text.x=element\_text(angle=45, hjust=1, vjust=1, size=10)) +   
 labs(title="사고유형의 상위 3가지 사망률(기타제외)", x='사고유형', y = '사망률') +   
 scale\_fill\_manual(values = c("#dddddd", "#F8766D"))

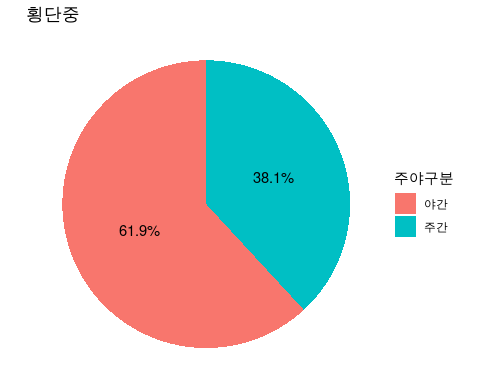


위 자료는 사고 유형별 사망자수와 부상자수의 상위 3개를 추려 사망률을 백분률로 나타낸 것으로

횡단중은 사망률이 약 90%에 육박하는 사망률을 기록하였다.

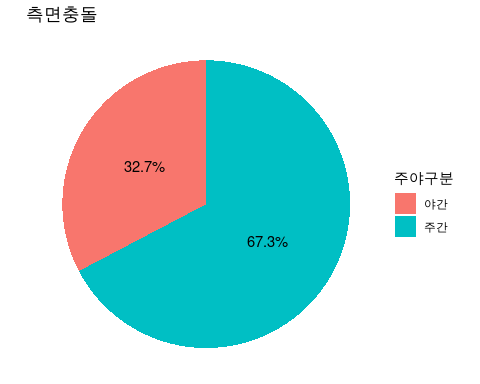
횡단중과 측면충돌의 주야구분을 확인한 결과,

tmp5 <- traffic\_accident %>% filter(사고유형 == '횡단중')  
tmp6 <- tmp5 %>%  
 group\_by(주야구분) %>%  
 summarise(n=n()) %>%   
 mutate(pct=n/sum(n)\*100)  
ggplot(tmp6, aes(x='', y=pct, fill=주야구분)) +  
 labs(title="횡단중") +  
 geom\_bar(stat='identity', color=NA) +  
 theme\_void() +  
 coord\_polar('y', start=0) +  
 geom\_text(aes(label=paste0(round(pct,1), '%')), position=position\_stack(vjust=0.5))



횡단중의 경우 전체사고 중 약 53%가 야간에 발생했고,

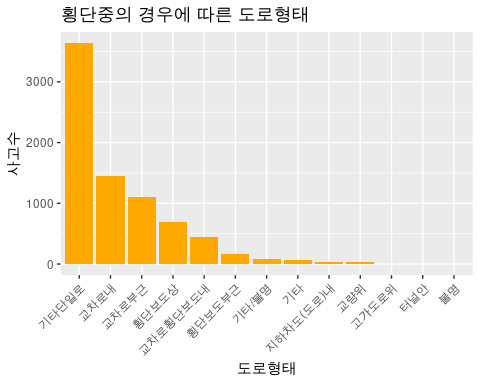
tmp7 <- traffic\_accident %>% filter(사고유형 == '측면충돌')  
tmp8 <- tmp7 %>%  
 group\_by(주야구분) %>%  
 summarise(n=n()) %>%   
 mutate(pct=n/sum(n)\*100)  
ggplot(tmp8, aes(x='', y=pct, fill=주야구분)) +  
 labs(title="측면충돌") +  
 geom\_bar(stat='identity', color=NA) +  
 theme\_void() +  
 coord\_polar('y', start=0) +  
 geom\_text(aes(label=paste0(round(pct,1), '%')), position=position\_stack(vjust=0.5))



안전거리 미확보의 경우 전체사고 중 약 67%가 주간에 발생했다.

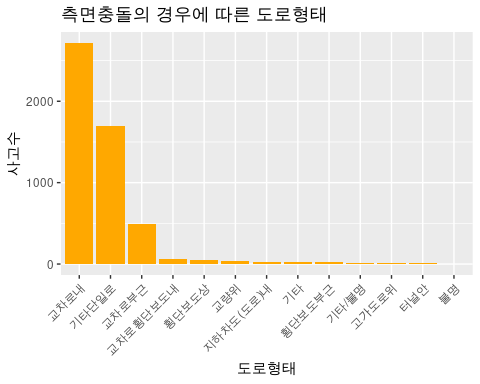
사망자수가 높은 횡단중과 측면충돌의 도로형태를 확인한 결과,

tmp9 <- traffic\_accident %>% filter(사고유형 == '횡단중') %>% group\_by(도로형태) %>%  
 summarise(사고수 = n()) %>% arrange(desc(사고수))  
ggplot(data=tmp9, aes(x=reorder(도로형태, -사고수), y=사고수)) +   
 geom\_bar(stat = "identity", position=position\_dodge(), fill='#FFA800') +   
 theme(axis.text.x=element\_text(angle=45, hjust=1)) +   
 labs(title="횡단중의 경우에 따른 도로형태", x='도로형태')



횡단중의 경우 교차로내, 기타단일로, 교차로부근에서 주로 발생했고,

tmp10 <- traffic\_accident %>% filter(사고유형 == '측면충돌') %>%   
 group\_by(도로형태) %>% summarise(사고수 = n()) %>% arrange(desc(사고수))  
ggplot(data=tmp10, aes(x=reorder(도로형태, -사고수), y=사고수)) +   
 geom\_bar(stat = "identity", position=position\_dodge(), fill='#FFA800') +  
 theme(axis.text.x=element\_text(angle=45, hjust=1)) +   
 labs(title="측면충돌의 경우에 따른 도로형태", x='도로형태')



측면충돌의 경우 교차로내, 기타단일로, 교차로부근에서 주로 발생했다.

## 4. 결론

분석 1.

* 해가 지날 수록 사고 건수는 줄어들고 있다.

분석 2.

* TOP3. 안전운전 의무 불이행, 신호위반, 중앙선 침범
* BOTTOM3. 교차로 통행방법 위반, 안전거리 미확보, 기타

분석 3.

* TOP3. 기타 단일로, 교차로 내, 교차로 부근

분석 4.

* 사망자가 가장 많은 사고 유형: 횡단중
* 부상자가 가장 많은 사고유형: 측면충돌

분석 1에서 해가 지날수록 교통사고가 줄어든다는 결론을 얻었지만 체감할 만큼 긍정적인 변화는 없었다는 현실이 아쉽다.

분석 2에서는 가해자가 위반한 법규가 안전운전 의무 불이행, 신호위반, 중앙선침법이라는 결론을 얻었다. 이 과정에서 학생들이 중앙선 침범에 대한 경각심을 다른 법규에 비하여 크게 가지지 않는다는 결과를 얻었다.

분석3에서는 기타단일로, 교차로내, 교차로부근에서 사고가 많이 일어난다는 결과를 얻을 수 있었으며, 이는 예상했던 결과와 크게 다르지 않았다.

마지막 분석4에서는 주야구분에서는 상관관계를 찾을 수 없었지만 가해자 법규위반, 도로형태에서 거의 비슷한 양상을 보였으며, 이를 통해 다른 사고유형도 크게 다르지 않다는 결과를 예측할 수 있었다.

따라서 분석 1 ~ 4의 결과와 분석 2에서 안전 운전의무 불이행이 압도적으로 높았던 점을 토대로 안전운전의무 불이행에 대한 경각심을 키울 수 있는 장치가 도로형태를 가리지 않고 필요하다는 결론을 내렸다.

## 5. Further Study

데이터셋을 확대하여 사망 교통사고 뿐만아니라 부상을 포함한 모든 교통사고에 대한 데이터를 확보하여 분석 하는것과, 결론에 장치가 필요하다고 한 것에 대해 교통사고를 줄이는 방법에 대한 데이터를 확보하여 효율적인 장치를 찾아내는 것을 후에 진행할 목표로 한다.