사망교통사고 정보 분석 및 시각화

프로젝트 중간 리포트

김강현, 조수빈, 최홍찬

11/12/2022

## 팀소개

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

## 1. 서론

1. 이 프로젝트는 전국에서 오는 저희 학교 학생들의 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.

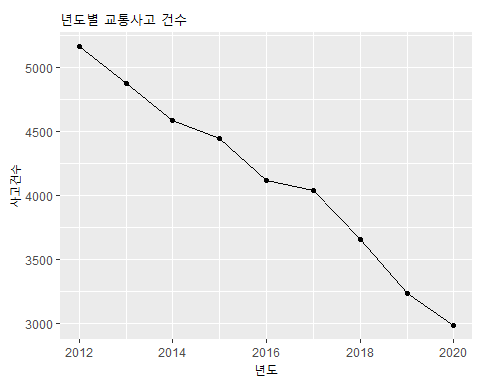
### 2.2 데이터 Dictionary

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

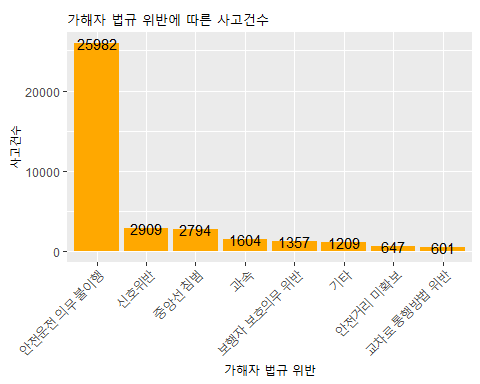
ggplot(data = count\_traffic\_accident\_by\_year, aes(x = 사고년도, y = count)) + geom\_line() + geom\_point() + labs(title="년도별 교통사고 건수", x = "년도", y = "사고건수")



년도와 사건수의 상관관계분석에서 교통법 재재강화, 단속으로 인해 사건수가 2013년 5.55퍼센트, 2014년 6.00퍼센트로 2012년부터 2020년까지 약 57퍼센트가 감소되는 유의미한 결과값을 확인하였다.

### 3.2 분석 2

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.

### 3.3 분석 3

traffic\_accident %>% group\_by(`사고유형`) %>% summarise(사망자수 = sum(사망자수), 부상자수 = sum(부상자수)) %>% arrange(desc(사망자수))

## # A tibble: 22 × 3  
## 사고유형 사망자수 부상자수  
## <chr> <dbl> <dbl>  
## 1 기타 8759 12650  
## 2 횡단중 7775 8421  
## 3 측면충돌 5397 9925  
## 4 공작물충돌 3705 5173  
## 5 정면충돌 2510 5946  
## 6 진행중 추돌 1845 4129  
## 7 전도전복 1715 2314  
## 8 추돌 1591 3749  
## 9 차도통행중 1584 1700  
## 10 도로이탈 추락 824 1030  
## # … with 12 more rows

tmp6 <- traffic\_accident %>% filter(사고유형 == '측면충돌' | 사고유형 == '횡단중')

사고유형별 사망자 부상자 를 비교분석한 결과 기타, 횡단중, 측면충돌 순으로 사망자가 많았고 기타의 경우 정확한 사고유형을 알 수 없으니 제외한다. 사망자수가 높은 횡단중과 측면충돌의 주야구분을 비교한 결과 아래 표를 보면

tmp2 <- count(tmp6, 사고유형 == '측면충돌' & 주야구분 == '주간')$n[[2]]  
tmp3 <- count(tmp6, 사고유형 == '측면충돌' & 주야구분 == '야간')$n[[2]]  
traffic\_accident %>% summarise(주야구분 = if(tmp2>tmp3) "주간" else "야간", 주간 = tmp2, 야간 = tmp3)

## # A tibble: 1 × 3  
## 주야구분 주간 야간  
## <chr> <int> <int>  
## 1 주간 3478 1690

측면충돌의 경우 전체사고 중 약 67%가 주간에 발생하였고

tmp4 <- count(tmp6, 사고유형 == '횡단중' & 주야구분 == '주간')$n[[2]]  
tmp5 <- count(tmp6, 사고유형 == '횡단중' & 주야구분 == '야간')$n[[2]]  
traffic\_accident %>% summarise(주야구분 = if(tmp4>tmp5) "주간" else "야간", 주간 = tmp4, 야간 = tmp5)

## # A tibble: 1 × 3  
## 주야구분 주간 야간  
## <chr> <int> <int>  
## 1 야간 2947 4780

횡단중의 경우 전체사고 중 약 62%가 야간에 발생하였다.

또한

tmp6 %>% filter(사고유형 == '측면충돌') %>% group\_by(도로형태) %>% summarise(n = n()) %>% arrange(desc(n))

## # A tibble: 13 × 2  
## 도로형태 n  
## <chr> <int>  
## 1 교차로내 2720  
## 2 기타단일로 1691  
## 3 교차로부근 486  
## 4 교차로횡단보도내 57  
## 5 횡단보도상 55  
## 6 교량위 43  
## 7 지하차도(도로)내 25  
## 8 기타 24  
## 9 횡단보도부근 19  
## 10 기타/불명 16  
## 11 고가도로위 15  
## 12 터널안 13  
## 13 불명 4

측면충돌의 경우 교차로내, 기타단일로, 교차로부근에서 주로 발생하였고,

tmp6 %>% filter(사고유형 == '횡단중') %>% group\_by(도로형태) %>% summarise(n = n()) %>% arrange(desc(n))

## # A tibble: 13 × 2  
## 도로형태 n  
## <chr> <int>  
## 1 기타단일로 3645  
## 2 교차로내 1448  
## 3 교차로부근 1103  
## 4 횡단보도상 689  
## 5 교차로횡단보도내 450  
## 6 횡단보도부근 162  
## 7 기타/불명 83  
## 8 기타 65  
## 9 지하차도(도로)내 37  
## 10 교량위 32  
## 11 고가도로위 7  
## 12 터널안 4  
## 13 불명 2

횡단중의 경우기타단일로, 교차로내, 교차로부근에서 주로 발생하였다.

### 3.4 분석 4

traffic\_accident %>% group\_by(`도로형태`) %>% summarise(총사망자=sum(사망자수), 총부상자=sum(부상자수), 평균=sum(사망자수)/sum(부상자수)) %>% arrange(desc(총사망자))

## # A tibble: 14 × 4  
## 도로형태 총사망자 총부상자 평균  
## <chr> <dbl> <dbl> <dbl>  
## 1 기타단일로 22890 35772 0.640  
## 2 교차로내 7287 11698 0.623  
## 3 교차로부근 4108 5878 0.699  
## 4 횡단보도상 984 1125 0.875  
## 5 교차로횡단보도내 653 721 0.906  
## 6 기타/불명 591 855 0.691  
## 7 교량위 541 996 0.543  
## 8 기타 474 680 0.697  
## 9 횡단보도부근 301 361 0.834  
## 10 지하차도(도로)내 268 451 0.594  
## 11 터널안 253 660 0.383  
## 12 고가도로위 154 308 0.5   
## 13 불명 13 20 0.65   
## 14 철길건널목 13 15 0.867

사고전체로 보았을때는 주로 기타단일로, 교차로내, 교차로부근 순서로 많이 일어났다.

## 4. 중간 결론

분석3의 결과로 사망/부상률이 가장 높은 사고유형이 전체적인 사망교통사고분석2와 4랑 비슷하게 나오는 것으로 보아 분석3에 결과는 측면충돌, 횡단중에 국한되지 않고 다른 사고유형에도 적용이 될 것 같다. 그렇기때문에 운전 중 안전운전의무를 이행하는 것을 잊지 않도록 하는 장치가 기타단일로 및 교차로내/부근에 필요할 것 같다.

## 5. 프로젝트 플랜

남은 기간 동안은 각 분석 결과를 확인하며 놓친 부분이 있는지 확인하고 타당성을 검사한 뒤 발표자료를 만들 것이다.