

빅데이터 분석 개별사례 조사

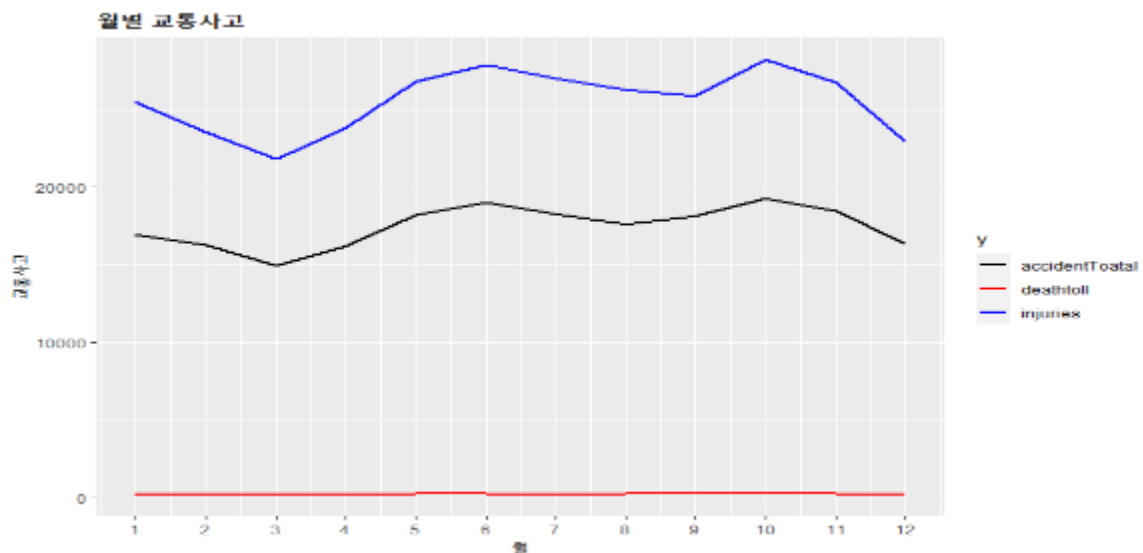
-교통사고 통계-

1. 데이터 정보

- 시도 시·군·구별로 2020년 동안의 교통사고 통계
- 경찰에서 조사, 처리한 교통사고에 대한 통계 정보로 인적 피해가 있는 사고만 집계
- 교통사고분석시스템의 데이터를 바탕으로
- <https://www.data.go.kr/data/15070297/fileData.do>

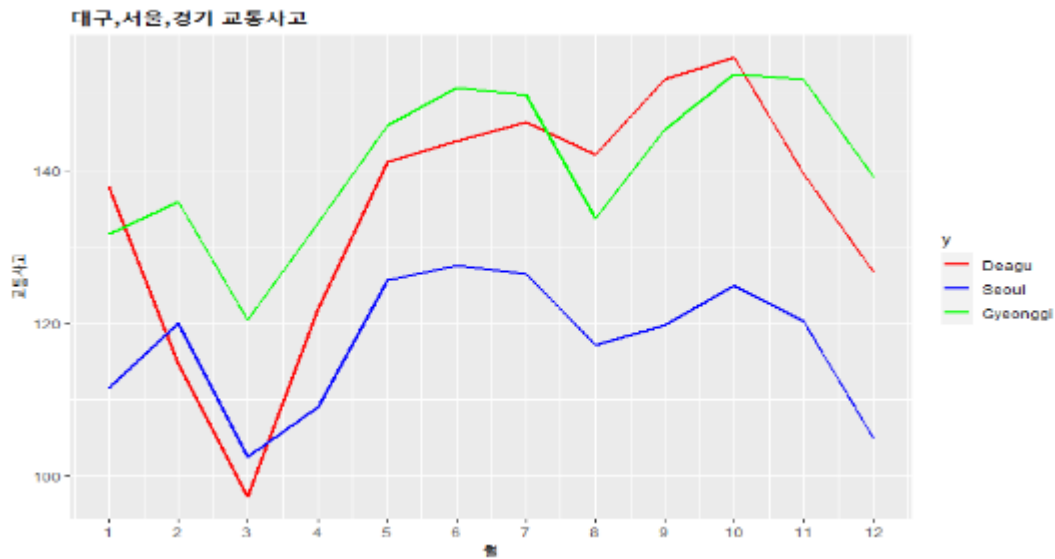
2. 데이터 분석

1)



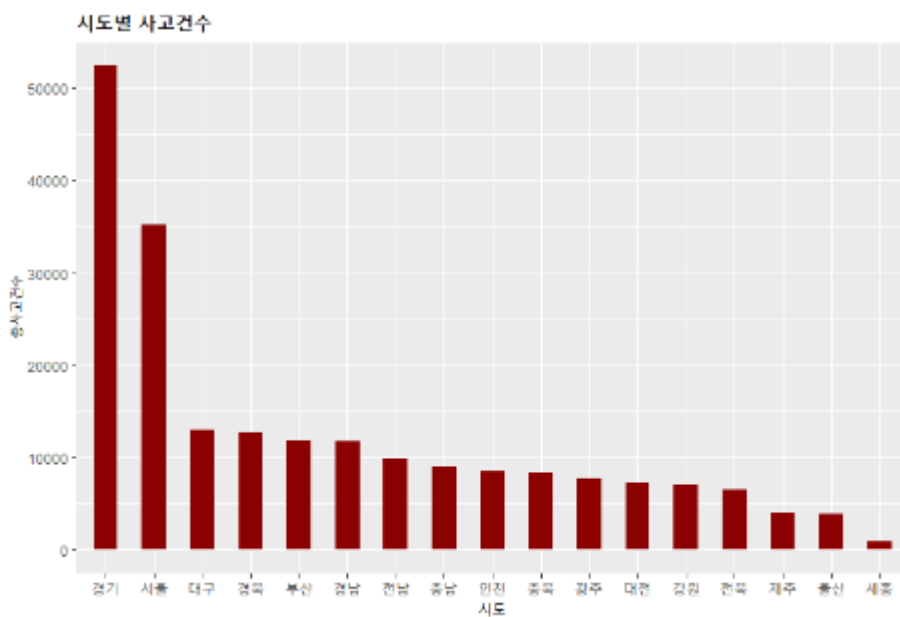
전국의 교통사고를 월별로 나타낸 것이다. 파란색 선은 부상자, 검정색 선은 교통사고건수, 빨간색 선은 사망자를 의미한다. 1~3월의 그래프가 비교적 급하게 하강하는 모습을 볼 수 있다. 11~12월은 이 이후의 데이터가 없기에 제외한다. 2020년 2월, 3월의 시기를 생각하면 코로나 1차 대유행이던 시기이다. 코로나로 이동수가 줄었기에 그에 따라 교통사고 수, 교통사고 부상자가 줄어든 것으로 보인다.

2)



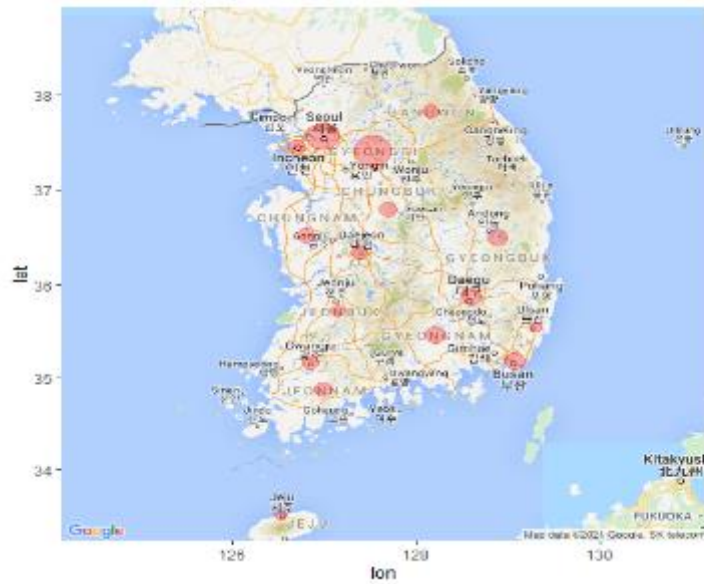
코로나 1차 대유행 지역인 대구와 멀리 떨어진 서울과 경기도의 교통사고건수를 비교한 그래프이다. 대구가 비교적 급격하게 하락한 모습을 보였기에 코로나로 인해 교통이동수가 줄어들었고 이에 교통사고수도 줄었다는 의견은 어느 정도 타당성이 있다.

3)



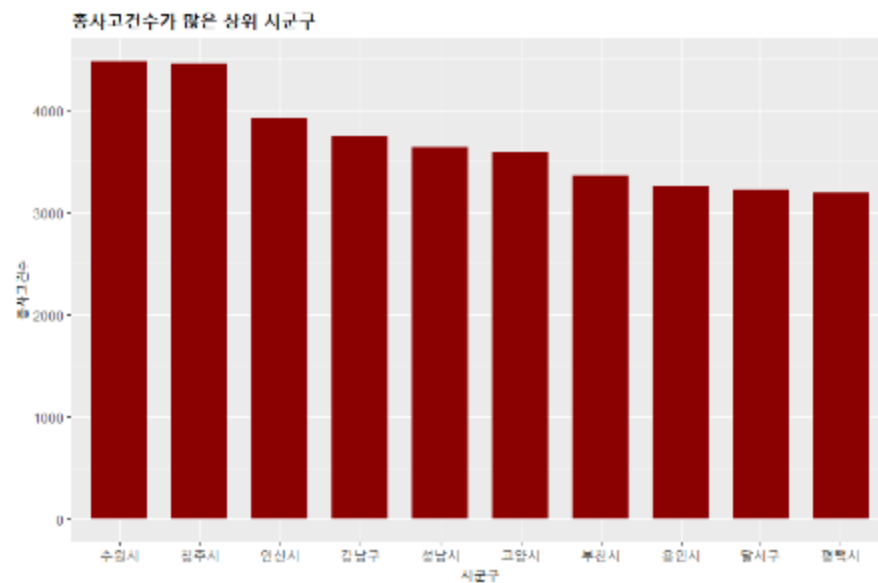
시도별로 사건건수를 차례대로 나열한 그래프이다. 인구수가 많은 경기와 서울의 교통사고수가 많다. 인구수가 높은 순위대로 교통사고가 많은 것은 아니지만 인구수가 높은 시도들이 교통사고 발생 상위권에 있다. 경기와 서울을 제외한 지역끼리 차이는 있지만 미미하게 보인다. 세종은 출범한지 얼마 되지 않은 특별자치시이기 때문에 그 수가 작게 나타난 것으로 보인다.

4)



위의 시도별 사고건수 그래프를 지도로 표현한 것이다. 고속도로가 많은 경기와 서울에 교통사고가 집중 되어있다.

5)



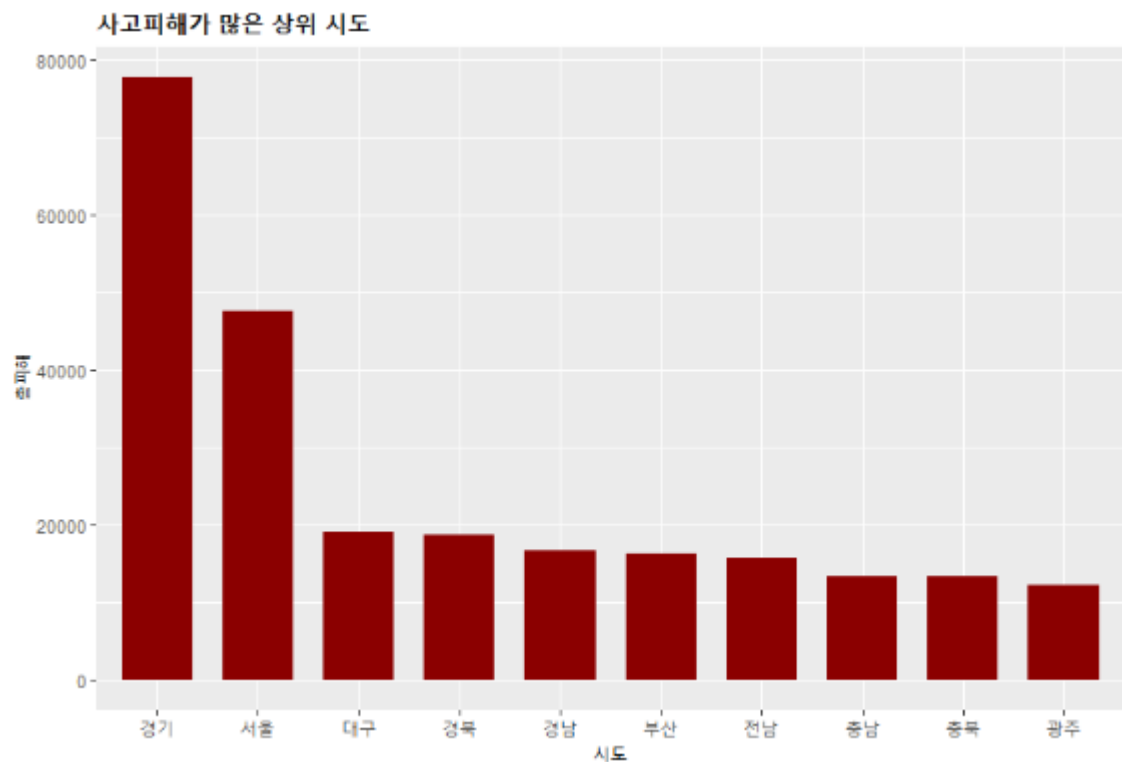
교통 사고수에 대해 시·군·구별로 세분화한 막대 그래프이다. 상위 10개의 시·군·구를 표시한 것으로 대부분 경기도에 속한다. 경기도에 속하지 않은 시·군·구로는 청주시, 강남구, 달서구가 있다. 강남구는 서울에 속하므로 서울과 경기를 제외한 지방에서 사고가 많은 시·군·구는 충청북도의 청주시, 대구의 달서구가 있다. 하지만 대구는 광역시로 대도시이고 청주시는 세종과 대전에 밀집되었으며 경기도와 비교적 가깝게 위치하였다. 인구수가 많은 도시 또는 그 도시 주변에 사고가 많은 것을 데이터로 확인하였다.

6)



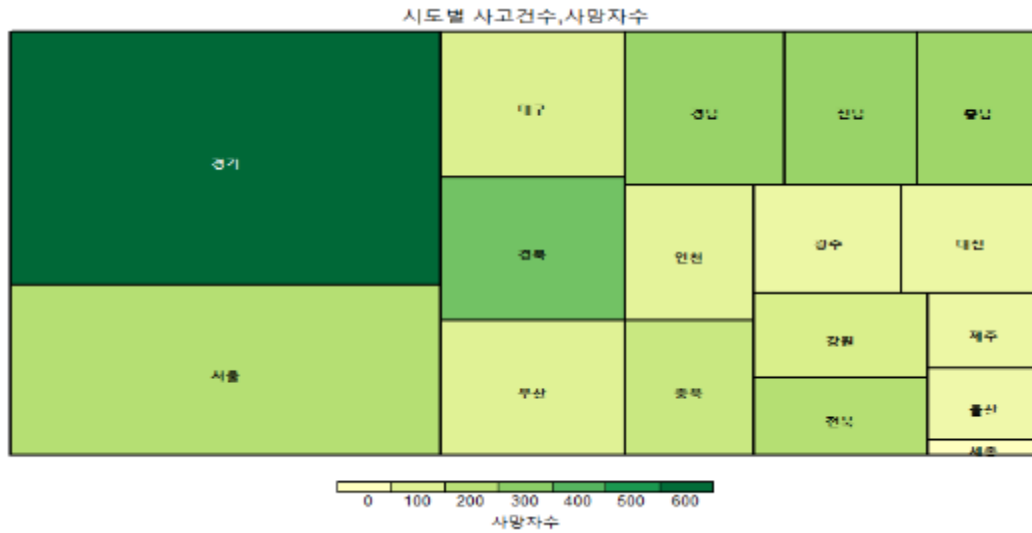
경상북도를 한정하여 교통사고수를 지도에 표현한 것이다. 경상북도에 발생했을지라도 광역시 주변 지역에 많이 발생한 것을 볼 수 있다. 대부분 대구 주변에 사고가 많지만 포항과 경주에도 사고가 많은 것을 볼 수 있다. 포항과 경주가 경상북도 도시의 인구수 순위 각각 1위,4위이다 2,3위는 구미시 경산시로 대구 주변에 위치하였다.

7)



교통사고에 비한 사망자와 부상자를 표현한 막대 그래프이다. 상위 10개의 시도이며 교통사고가 많이 일어난 시도의 순위와 비슷하다.

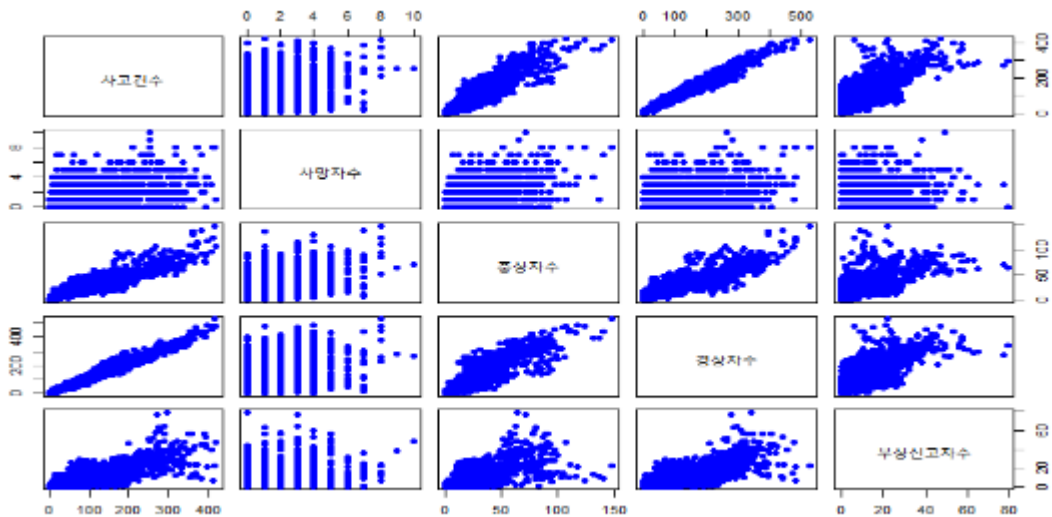
8)



시도별로 교통사고수와 사망자수에 대한 트리맵이다. 크기는 교통사고수, 색깔은 사망자수를 의미하며 교통사고수가 많을수록 사망자가 많다. 교통사고가 발생해야 그에 따른 사망자가 나오기 때문에 유의미한 결과 값은 아니다.

9)

사고건수	1.0000000	0.4238921	0.9233442	0.9851235	0.8015420
사망자수	0.4238921	1.0000000	0.4665928	0.4176615	0.2721347
중상자수	0.9233442	0.4665928	1.0000000	0.8839306	0.6971281
경상자수	0.9851235	0.4176615	0.8839306	1.0000000	0.7621941
부상신고자수	0.8015420	0.2721347	0.6971281	0.7621941	1.0000000



사고건수에 대한 중상자, 경상자, 부상신고자의 상관관계가 높은 편이지만 사고건수가 원인이 되어 발생하기 때문에 유의미하지 않다. 중상자수와 경상자수, 경상자수와 부상신고자수 또한 같은 맥락을 가지기 때문에 이 자료로 유의미한 상관분석이 불가하다.

3. 결론

이 자료로 통해 코로나19사태로 인해 이동수가 줄어 교통량이 줄어든 모습을 알 수 있었다. 인구수가 많은 도시 또는 그 주변에 교통사고와 그에 따른 피해가 큰 것 또한 알 수 있었지만 유의미한 상관분석은 알 수는 없었다.