

¹College of Transdisciplinary Studies, DGIST

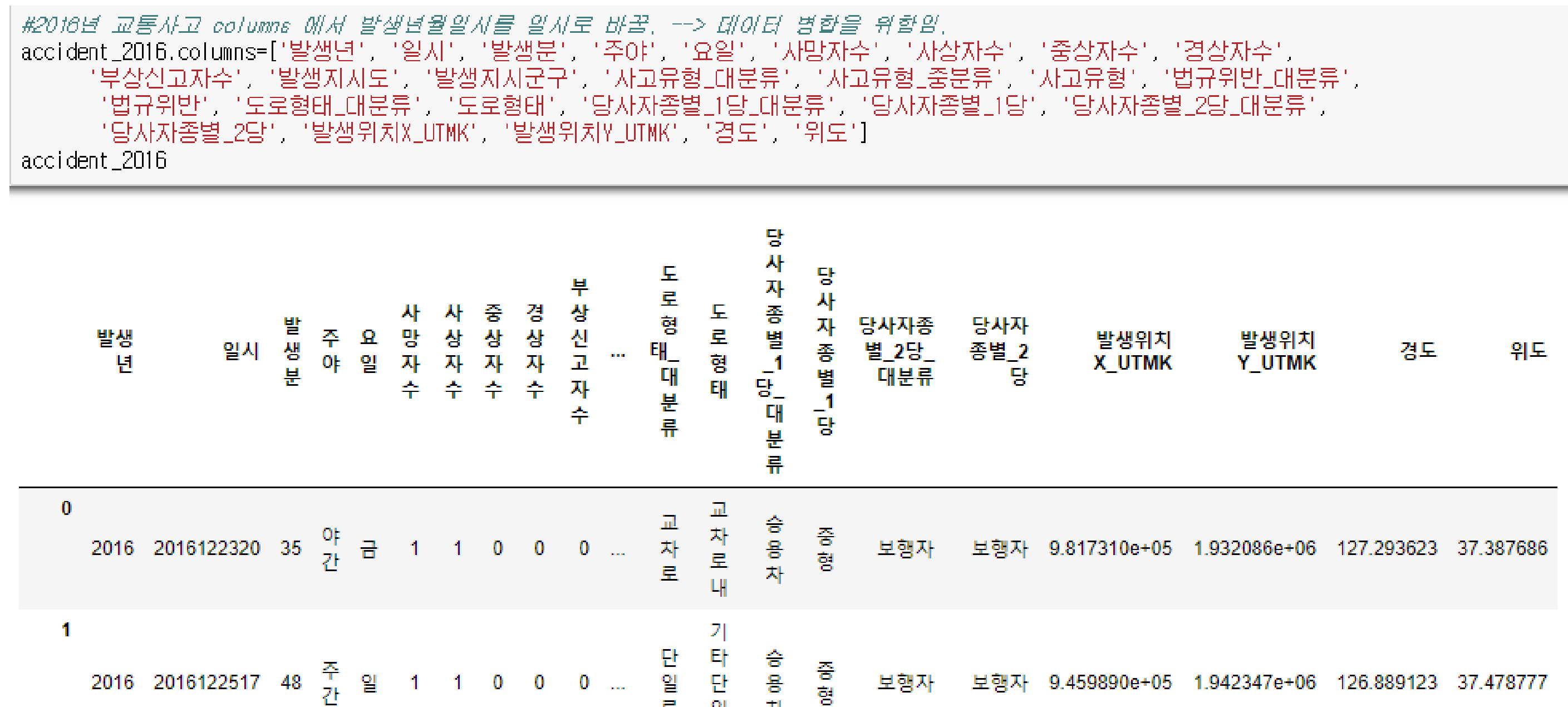
5 rows x 27 columns

교통사망사고 분석 및 대응방안

Team04 남윤철¹, 이승훈¹, 장용주^{1*}

¹College of Transdisciplinary Studies, DGIST

2. 불러온 데이터 중에서 데이터인 사망교통사고 데이터와 날씨 데이터를 병합하기 위해 교통사고 columns에서 ‘발생년월일시’를 ‘일시’로 바꾸었습니다. (columns 값을 일치시켜 병합하기 위함)



3. 먼저 병합할 기준이 될 ‘일시’ 칼럼의 형식을 맞추기 위해 교통사고의 일시 column과 날씨 데이터의 일시 column의 형식을 ‘yyyymmddhh’로 통일합니다. 그리고 날씨 데이터가 시군구별로 나와있지 않고 지점별로 나와있기 때문에 지점별 발생지시도로 정리한 ‘weather_merge_key.csv’로 날씨데이터에 ‘발생지시도’ column을 추가하고 이제 날씨데이터의 ‘발생지시도’와 교통사고의 ‘발생지시도’를 merge 하여 두 데이터를 병합하였습니다.

```
# 날씨 데이터와 교통사고 데이터를 병합하기 위한 과정
accident_2016['datetime'] = accident_2016['일시'].apply(lambda x: pd.to_datetime(str(x), format="%Y%m%d%H"))
#pd.to_datetime(accident_2016['일시'], format="%Y%m%d%H")
accident_2016.datetime.head()

weather_2016=pd.read_csv('2016_weather.csv', quotechar='\"', encoding='CP949')
weather_2016.head()

weather_2016['datetime'] = weather_2016['일시'].apply(lambda x: pd.to_datetime(str(x), format="%Y-%m-%d %H:%M"))
weather_2016.head()

accident_2016.loc[:, ['발생지시도', '사상자수', 'datetime']]
key_column=pd.read_csv('weather_merge_key.csv', quotechar='\"', encoding='CP949')
key_column.head()

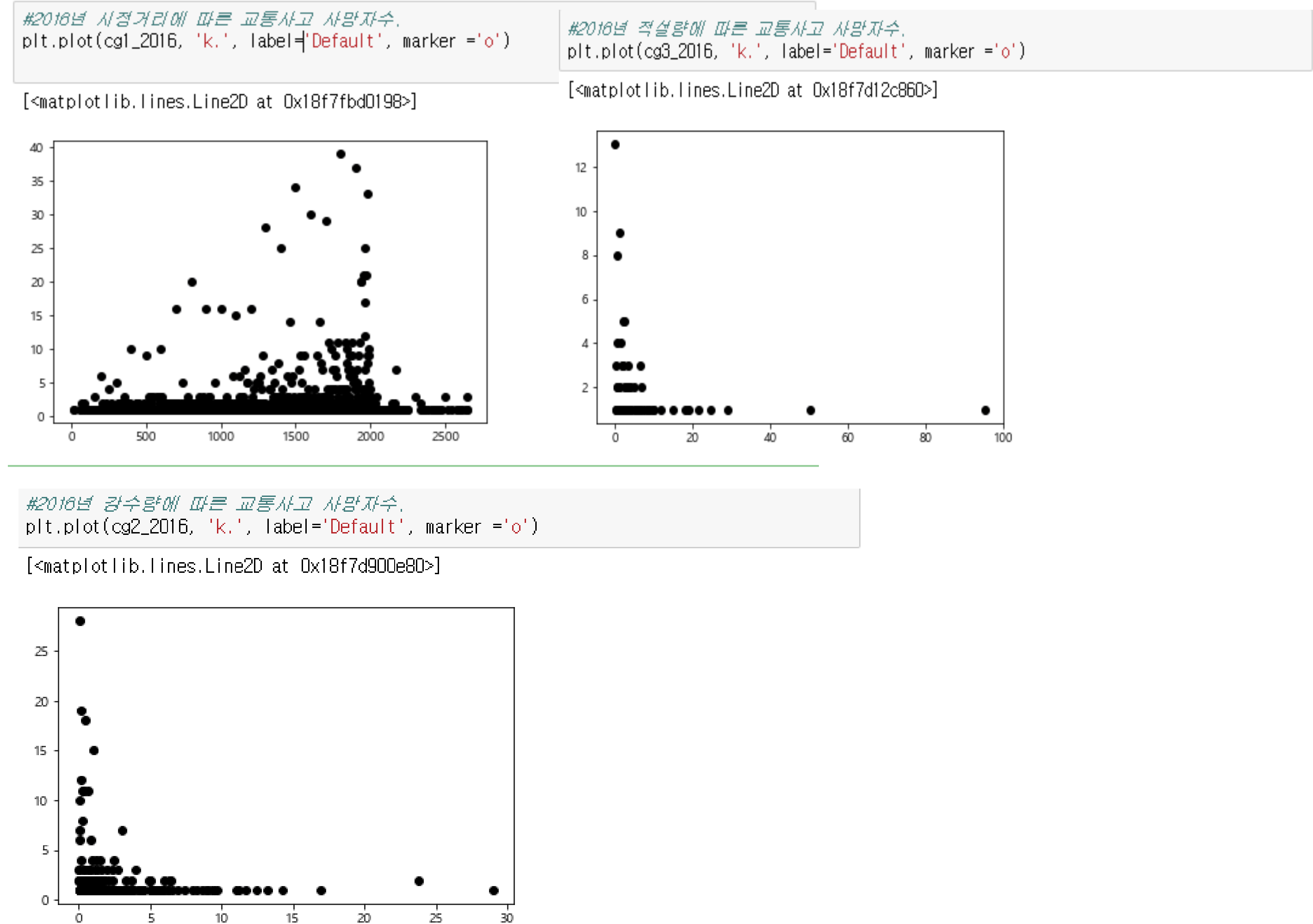
weather_location_2016 = pd.merge(weather_2016, key_column, how='left', on='지점')
weather_location_2016.head()

weather_location_temp_2016=weather_location_2016.loc[:, ['발생지시도', 'datetime', '강수량(mm)', '적설(cm)', '전운량(10분위)', '시정(10m)']]

# 날씨 데이터와 교통사고 데이터를 병합하기 위한 과정
weather_location_sort_2016=weather_location_temp_2016.groupby(['발생지시도', 'datetime']).mean()
weather_location_sort_2016.reset_index(inplace=True)
weather_location_sort_2016

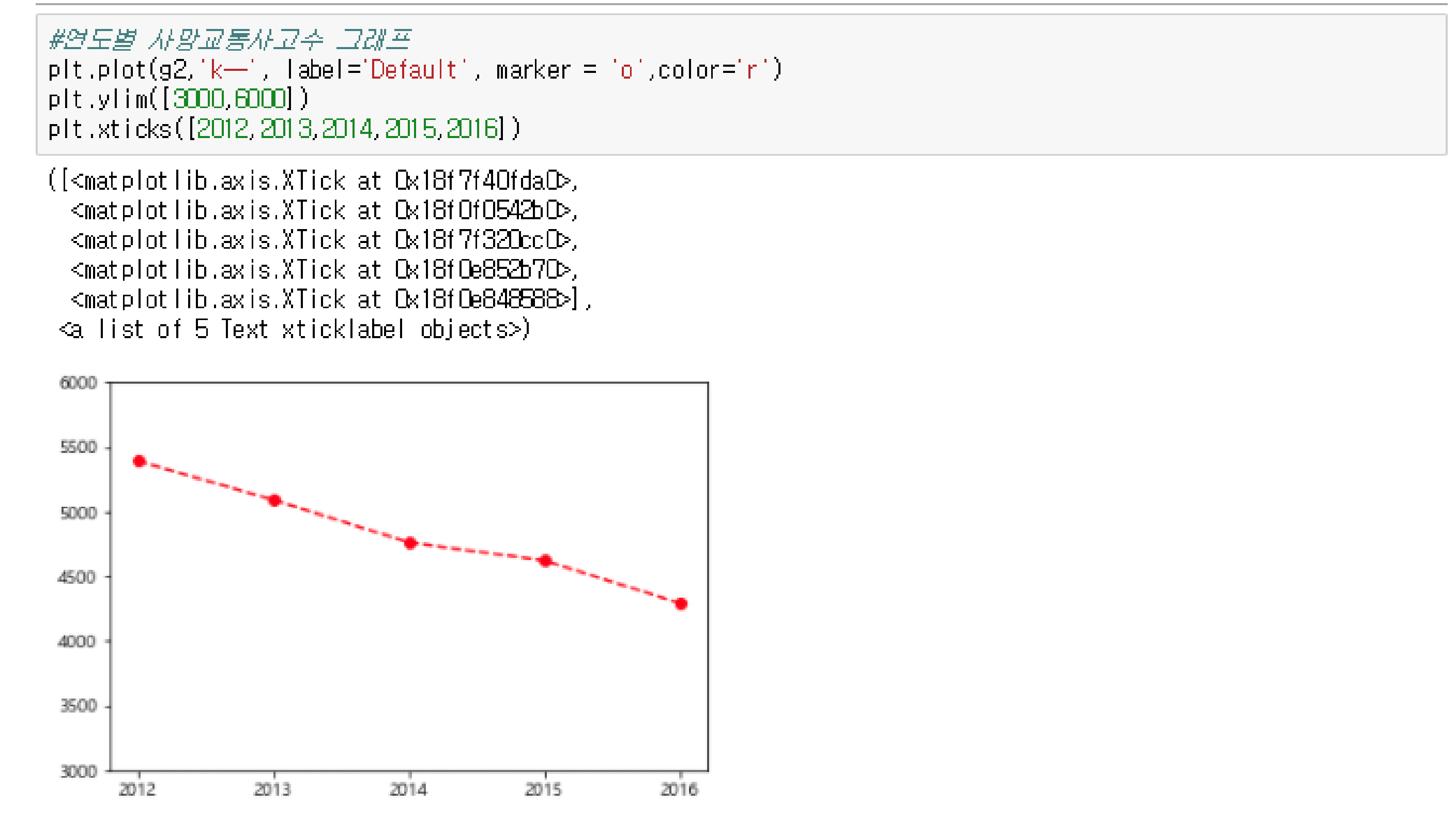
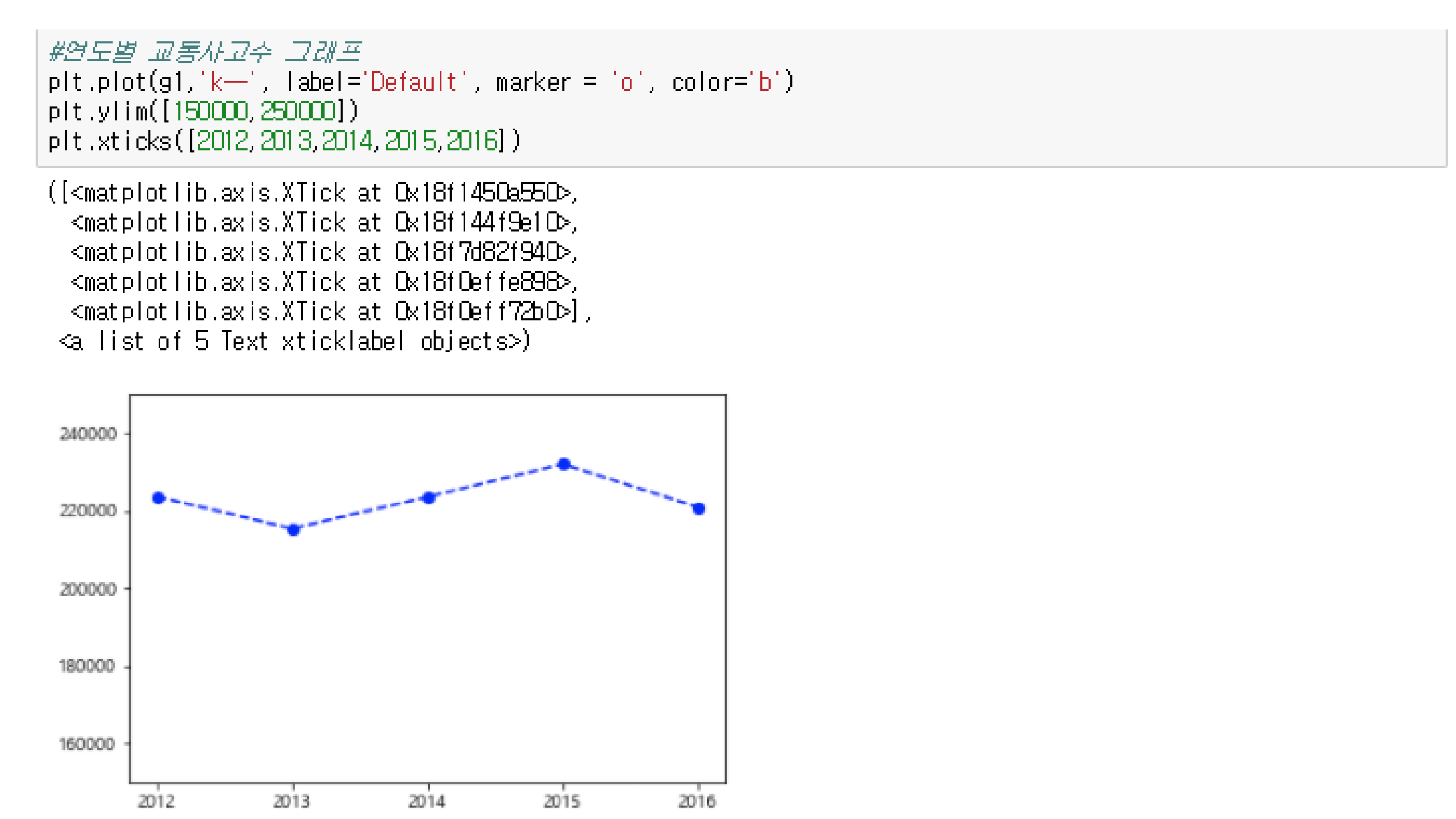
weather_casualty_2016 = pd.merge(accident_2016, weather_location_sort_2016, how='left', on=['발생지시도', 'datetime'])
combined_2016=weather_casualty_2016.loc[:, ['datetime', '시정(10m)', '강수량(mm)', '적설(cm)', '도로형태', '사상자수']]
combined_2016.head()
```

4. 병합한 데이터를 통해 시정거리, 강수량과 적설량에 대한 그래프 그림.



시정거리, 강수량과 적설량의 그래프를 그려보았는데 날씨와 교통사고 사건 수 사이에 일정한 상관관계가 있을 거라 예상하였지만 의미있는 관계를 찾아내지는 못했습니다. 교수님이 주신 질문에서 강수량이 많을 때와 적설량이 많을 때 사고가 많이 발생하지 않은 점은 비가 오는 날과 눈이 오는 날이 적기 때문에 사고가 많이 나지 않았기 때문입니다. 그래서 만약에 적설량과 강수량의 데이터를 의미있게 사용하려면 강수량 마다 혹은 적설량 마다 그에 해당하는 일시 목록을 만들고 그 목록에 속하는 일시에 사고가 난 수를 비교해 본다면 기후와 교통사고의 수의 관계를 보다 정확하게 비교 할 수 있음을 예상합니다.

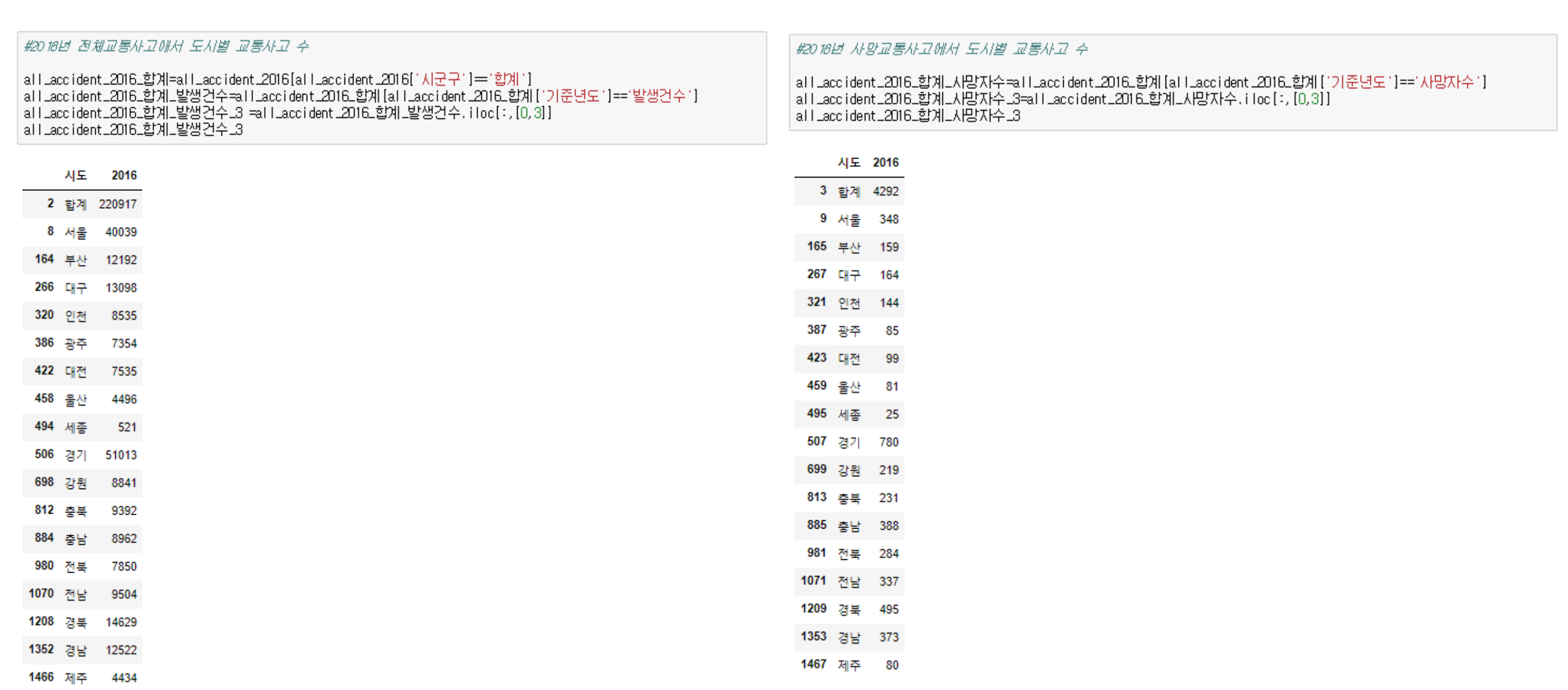
5.교통사고수의 경향성을 알기 위해 연도별 교통사고수 그래프와 연도별 사망교통사고수 그래프를 나타내었습니다.



연도별 교통사고수, 연도별 교통사망사고수 분석한 이유

연도별 전체 교통사고수는 딱히 경향성이 없었고, 교통사망사고에 대한 그래프에서는 점점 사망교통사고수가 감소함을 알 수 있습니다. 교통사망사고가 줄어든다고 하더라도 발표(최종 ppt)에 제시했듯이 OECD국가에서 인구별 교통사망사고수가 10으로 높은 수치를 보임을 알 수 있습니다. 그래서 교통사고에 대한 분석을 통해 사망교통사고를 줄일 수 있으면 좋지 않을까 한 계기가 이 그래프를 그리고 난 이후였습니다.

6. 2016년 전체교통사고에서 도시별 교통사고 수와 교통사망사고에서 도시별 교통사고수 데이터를 뽑아내어 특정 도시서 발생하는 교통사고수를 비교해 보았습니다.



지역별 교통사고의 수는 인구가 제일 많은 경기가 제일 높았으며 그 다음으로는 서울, 경북, 대구, 부산 순이었습니다. 지역별 교통사망사고의 수 또한 경기, 경북, 서울 순으로 인구가 많을수록 그 수가 크다는 것을 알 수 있었습니다. 단순히 특정지역에서 교통사고가 많이 일어난다고 볼 수 없다고 판단하였고 그래서 지역보다는 자동차 종류에 따라 분류를 하면 특정한 값이 나올 수도 있다고 생각해서 대형차인 화물차에서 사망사고가 많이 일어난다고 예상했고, 이에 관한 사고유형, 법규위반, 도로형태, 주야에 대해서 데이터를 뽑아봤습니다.(다음장)

교통사망사고 분석 및 대응방안

Team04 남윤철¹, 이승훈¹, 장용주^{1*}

¹College of Transdisciplinary Studies, DGIST

#2019년 사망교통사고에서 당사자종별, 1당, 2당분류에서 화물차에 대해 읽기

```
hnameu=Caccident_2016.loc[:,['사고유형','당사자종별,1당,2당분류','법규위반','도로형태','주야']]
hnameu[hnameu['당사자종별,1당,2당분류']=='화물차']
h[h['주야']=='주간']
```

사고유형	당사자종별, 1당, 2당분류	법규위반	도로형태	주야
26	정면충돌	화물차 교차로 통행방법 위반	교차로부근	주간
264	횡단충돌	화물차 안전운전 의무 불이행	교차로부근	주간
308	공작물충돌	화물차 안전운전 의무 불이행	기타단일로	주간
311	정면충돌	화물차 안전운전 의무 불이행	기타단일로	주간
352	드라이브 주박	화물차 안전운전 의무 불이행	기타단일로	주간
371	횡단충돌	화물차 보행자 보호의무 위반	교차로내	주간
428	횡단충돌	화물차 안전운전 의무 불이행	기타단일로	주간
436	측면직각충돌	화물차 교차로 통행방법 위반	교차로내	주간
505	진행중 추돌	화물차 안전운전 의무 불이행	교차로부근	주간
636	기타	화물차 앞지르기 금지위반	교차로내	주간
642	공작물충돌	화물차 안전운전 의무 불이행	교차로부근	주간
794	측면직각충돌	화물차 안전운전 의무 불이행	교차로내	주간
795	기타	화물차 안전운전 의무 불이행	기타단일로	주간
854	측면직각충돌	화물차 안전운전 의무 불이행	기타단일로	주간
877	전도전복	화물차 안전운전 의무 불이행	교황학	주간
896	정면충돌	화물차 중앙선 침범	기타단일로	주간
904	진행중 추돌	화물차 안전운전 의무 불이행	교차로부근	주간
912	전도전복	화물차 안전운전 의무 불이행	기타단일로	주간
1065	측면직각충돌	화물차 과속	기타단일로	주간
1385	횡단충돌	화물차 보행자 보호의무 위반	교차로부근	주간
1326	기타	화물차 안전운전 의무 불이행	기타단일로	주간
1327	정면충돌	화물차 안전운전 의무 불이행	기타단일로	주간
1335	측면직각충돌	화물차 안전운전 의무 불이행	기타단일로	주간
1350	횡단충돌	화물차 보행자 보호의무 위반	교차로내	주간
1555	횡단충돌	화물차 안전운전 의무 불이행	교차로내	주간
1559	전도전복	화물차 안전운전 의무 불이행	기타단일로	주간

#2019년 사망교통사고에서 당사자종별, 1당, 2당분류에서 화물차에 대해 읽기

```
hnameu=Caccident_2016.loc[:,['사고유형','당사자종별,1당,2당분류','법규위반','도로형태','주야']]
hnameu[hnameu['당사자종별,1당,2당분류']=='화물차']
h[h['주야']=='주간']
```

사고유형	당사자종별, 1당, 2당분류	법규위반	도로형태	주야
268	전도전복	화물차 안전운전 의무 불이행	교차로내	야간
283	정면충돌	화물차 안전운전 의무 불이행	기타단일로	야간
438	밀가장차단구역통행중	화물차 안전운전 의무 불이행	기타단일로	야간
443	밀가장차단구역통행중	화물차 안전운전 의무 불이행	기타단일로	야간
469	보도통행중	화물차 보행자 보호의무 위반	교차로부근	야간
512	전도전복	화물차 안전운전 의무 불이행	기타단일로	야간
724	측면직각충돌	화물차 안전운전 의무 불이행	교차로내	야간
749	전도전복	화물차 안전운전 의무 불이행	기타단일로	야간
809	전도전복	화물차 안전운전 의무 불이행	교황학	야간
1112	횡단충돌	화물차 안전운전 의무 불이행	기타단일로	야간
1473	차도통행중	화물차 안전운전 의무 불이행	기타단일로	야간
1584	횡단충돌	화물차 안전운전 의무 불이행	교차로내	야간
1609	측면직각충돌	화물차 안전운전 의무 불이행	교차로내	야간
1679	공작물충돌	화물차 안전운전 의무 불이행	기타단일로	야간
1704	정면충돌	화물차 안전운전 의무 불이행	교차로부근	야간
1743	공작물충돌	화물차 안전운전 의무 불이행	교차로부근	야간
1912	측면직각충돌	화물차 중앙선 침범	교차로부근	야간
2062	정면충돌	화물차 중앙선 침범	기타단일로	야간
2112	전도전복	화물차 안전운전 의무 불이행	기타단일로	야간
2165	측면직각충돌	화물차 안전운전 의무 불이행	교차로내	야간
2181	정면충돌	화물차 중앙선 침범	교황학	야간
2228	진행중 추돌	화물차 안전운전 의무 불이행	기타단일로	야간
2229	횡단충돌	화물차 안전운전 의무 불이행	기타단일로	야간
2267	진행중 추돌	화물차 안전운전 의무 불이행	기타단일로	야간
2323	진행중 추돌	화물차 안전운전 의무 불이행	기타단일로	야간
2447	진행중 추돌	화물차 안전운전 의무 불이행	기타단일로	야간

위의 데이터는 교통사망사고에서 화물차의 야간, 주간에 대한 데이터입니다. 야간에 운전자가 졸음운전을 하는 경우가 많다고 생각하여 야간에 많은 교통사망사고가 있을 줄 알았는데 야간보다 주간에 더 많은 교통사망사고가 났음을 알 수 있었습니다. 이를 분석하기 위해서는 교통량 별 사고수를 시간대를 나누어 분석한다면 유의미한 관계를 얻을 수 있을 것이라 생각합니다. 그래서 교통량을 병합하기 위해서 관련 데이터를 찾아보았습니다. 교통량 데이터와 사고를 엮기 위해서 필요한 정보는 사고당시 시간과 사고당시 위치입니다. 하지만 교통량데이터에서는 위치정보가 도로번호로 주어져 있고 또한 도로번호는 국도와 고속도로에만 부여된 정보라서 그 이외의 도로에 대한 정보를 분석할 수 없고 또한 도로번호는 사고지점의 위치에 대한 정보를 포함하지만 정확히 대응하는 것이 아니기 때문에 의미있는 결과를 얻을 수 없다고 판단했습니다.

또한 화물차, 소형차, 중형차 등 자동차 종류에 대한 교통사망사고에 대한 분석을 한다면 대한민국에 있는 화물차, 소형차, 중형차의 수에 대한 데이터를 통해 각 차종수에 따른 사고 비율을 구하는 방향을 구상해 보았으나, 적절한 데이터 셋을 찾지 못하였습니다.

7.사망교통사고에서 도로형태 데이터 및 그래프

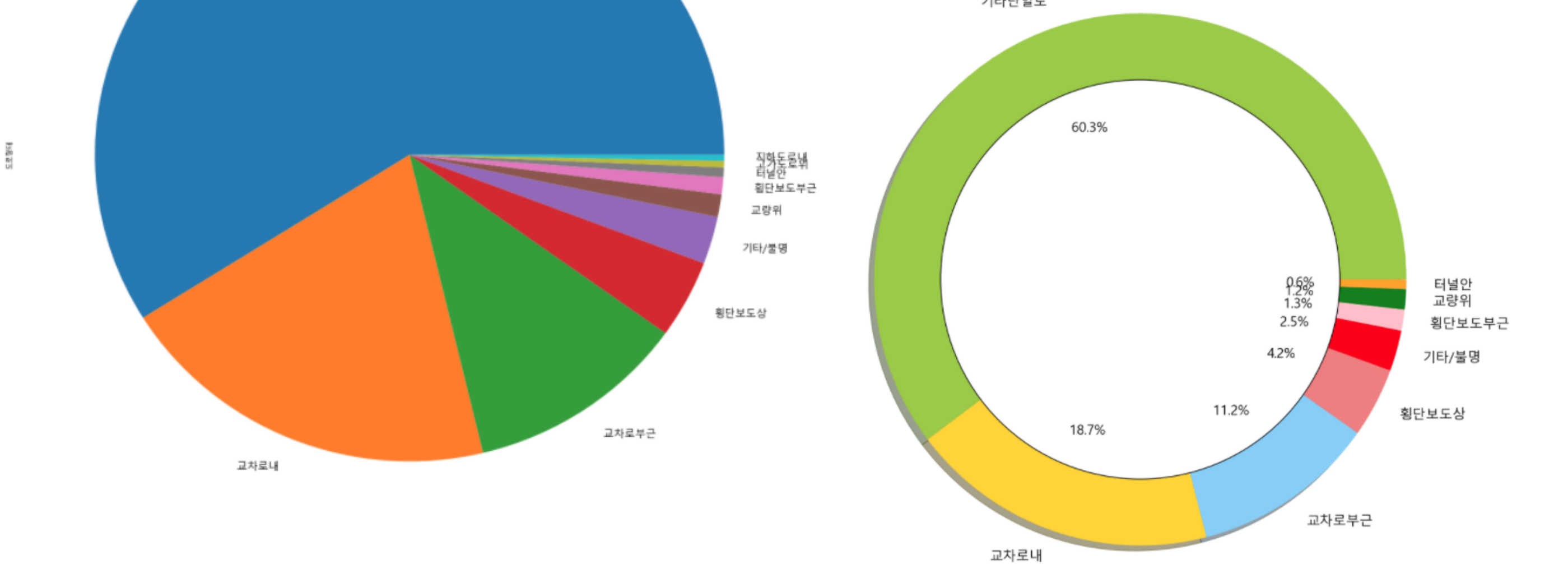
#2012~2016 사망사고 사건종에서 도로형태의 수에 따라 읽기

```
accident_2012~2016.도로형태.value_counts()
```

```
기타단일로      13882
교차로내        4319
교차로부근      2552
횡단보도상      974
기타/불명       571
횡단보도부근    294
교황학          283
터널안         133
고가도로위      85
지하도로내     52
건널목          5
Name: 도로형태, dtype: int64
```

#2012~2016 사망사고 사건종에서 도로형태와 수를 density over로 나타내기

```
labels = ['기타단일로', '교차로내', '교차로부근', '횡단보도상', '기타/불명', '횡단보도부근', '교황학', '터널안']
size = [13882, 4319, 2552, 974, 571, 294, 283, 133]
colors = ['yellowgreen', 'gold', 'lightslateblue', 'lightcoral', 'red', 'pink', 'green', 'orange']
explode = (0, 0, 0, 0, 0, 0, 0, 0)
plt.figure(figsize=(15,10))
ax=plt.subplot(111,projection='aea',r=19.0)
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True)
centre_circle=plt.Circle((0,0),0.75,color='black',fc='white',linewidth=2.5)
fig=plt.gcf()
fig.gca().add_artist(centre_circle)
plt.axis('equal')
plt.show()
```



앞서 적설량, 강수량, 시정거리, 차 종류, 도시별 등 다양한 관점에서 데이터를 분류해 보았으나 교통량, 자동차 종류의 비율 등 데이터를 엮기에는 접점이 없었어 교통사고 수와 유의미한 관계를 가지는 특징을 찾지 못했습니다. 이 외에 사망교통사고에서 도로형태에 대한 데이터로 분석을 해보았습니다. 분석결과 위의 사진과 같이 기타단일로에서 사고가 난 경우가 압도적으로 많았으며, 교차로내, 교차로부근, 횡단보도부근 차례로 사고가 많이 발생하였습니다. 단순히 이 데이터로는 의미있는 값을 뽑아내지 못했고, 의미있는 데이터를 뽑기 위해 사고유형별로 데이터를 분석해보았습니다.

8.사고유형별 사망자수 데이터 및 그래프

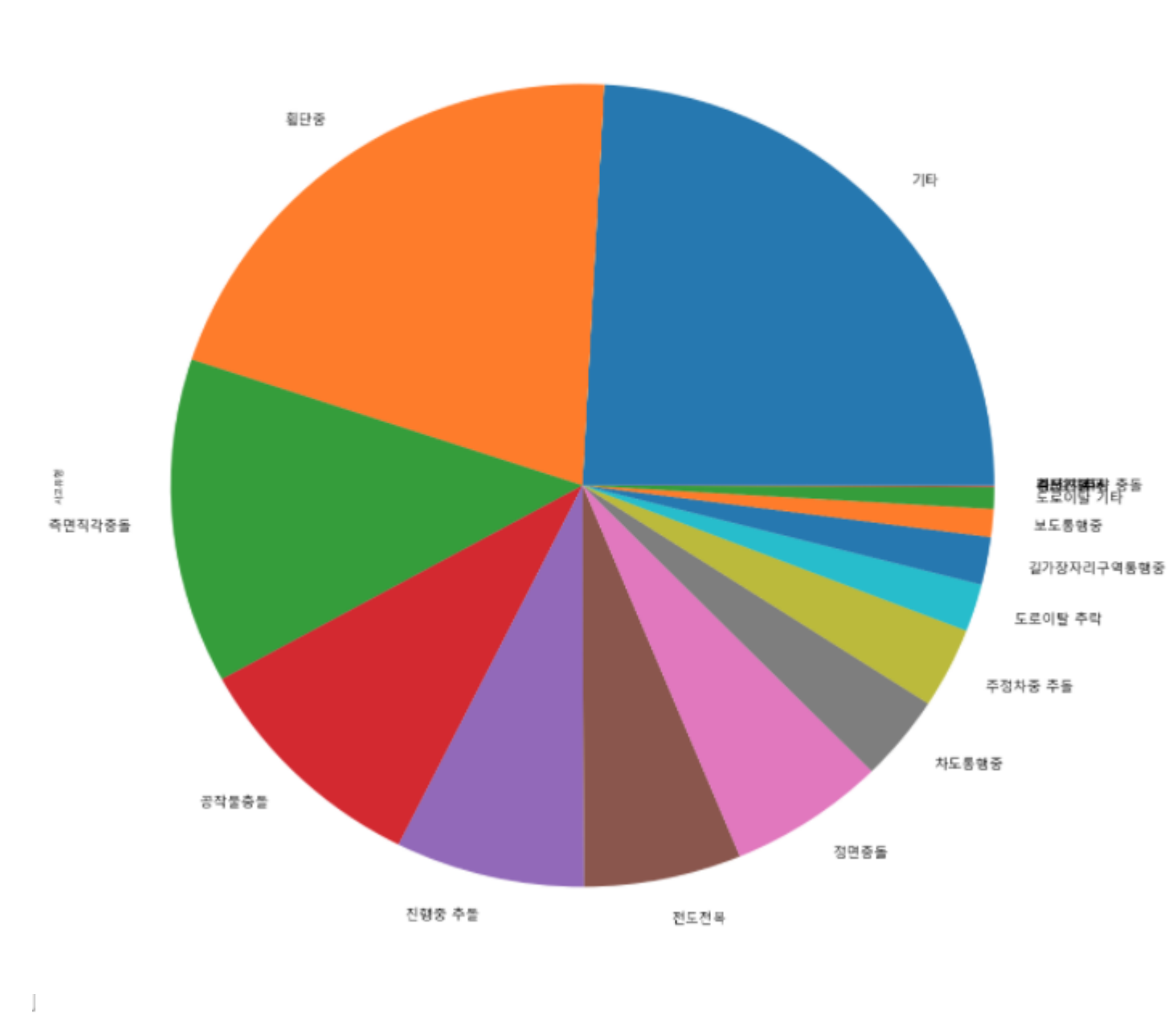
#2012~2016년도 사망교통사고에서 사고유형별 사망자수 읽기

```
accident_2012~2016.사고유형.value_counts()
```

```
기타      5532
횡단충돌  4815
측면직각충돌  3037
공작물충돌  2228
진행중 추돌  1724
전도전복  1438
정면충돌  1427
차도통행중  803
주정차중 추돌  754
도로미탈 추락  445
밀가장차단구역통행중  445
보도통행중  260
도로미탈 기타  204
추/정차차량 충돌  9
경보기 무시  2
직전전행  1
Name: 사고유형, dtype: int64
```

#2016년도 사망교통사고 사고유형 pie 그래프

```
accident_pie.사고유형 = 사고유형.accident.plot(kind='pie',figsize=(20,20),fontsize=15,subplots=True)
```



9.사고유형 & 도로형태 별, 사망자수 3차원 그래프

2012~2016년 사망교통사고 도로형태와 사고유형을 3차원 plot으로

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

# Load data
accident = pd.read_csv('accident_2012~2016.csv')

# Filter for fatal accidents
fatal_accident = accident[accident['사망자수'] > 0]

# Group by road type and accident type
grouped = fatal_accident.groupby(['도로형태', '사고유형'])

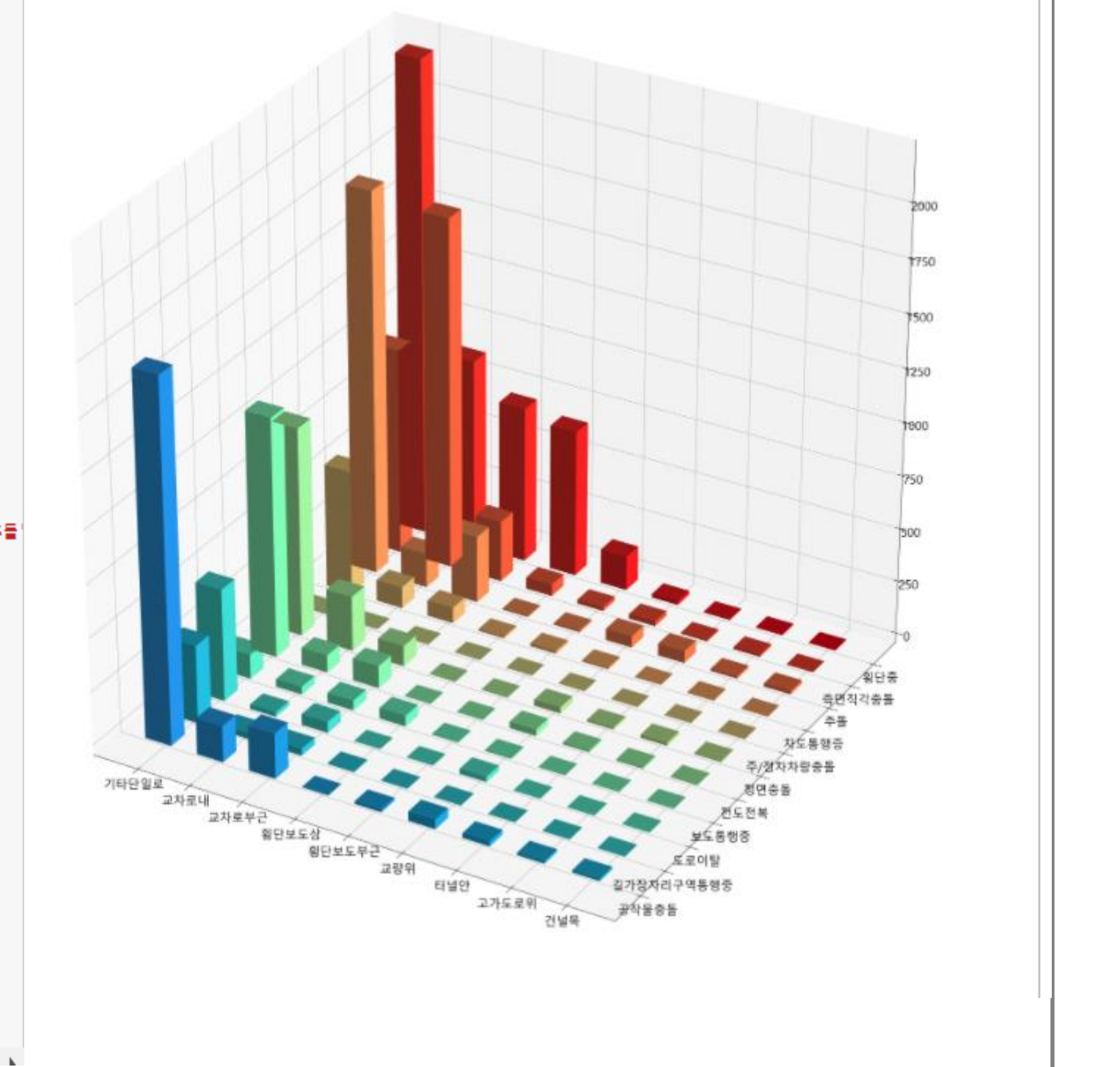
# Count the number of deaths
counts = grouped.agg({'사망자수': 'sum'})

# Melt the data for 3D bar chart
melted = counts.reset_index().melt(id_vars=['도로형태', '사고유형'], var_name='Category', value_name='Deaths')

# Create 3D bar chart
fig = plt.figure(figsize=(20,20))
ax = fig.gca().projection='3d'

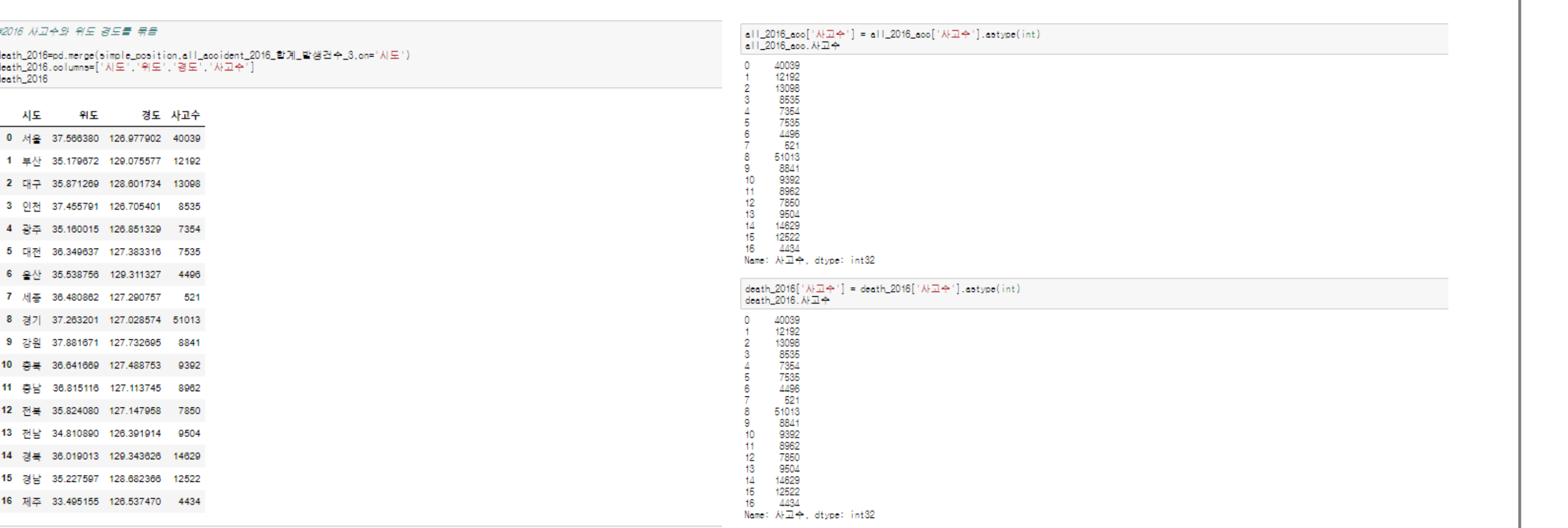
# Add labels for each bar
for i in range(len(melted)):
    label = '%d' % melted['Deaths'][i]
    ax.text(melted['도로형태'][i], melted['사고유형'][i], melted['Deaths'][i], label)

# Show the plot
plt.show()
```



위의 3차원 그래프는 도로형태와 사고유형을 엮은 데이터의 교통사망사고 수를 나타낸 것입니다. 특히 도로형태는 기타단일로, 사고유형은 횡단 중일 때가 아주 많은 교통사망사고가 발생했음을 알 수 있습니다. 이를 통해 각 사고가 어디서 발생하는지 Bubble map과 Dot map을 나타내어 확인한다면 사고가 자주 발생한 지역에 대해 도로형태를 분석하여 개선할 수 있는 부분을 찾을 수 있다고 생각했습니다.

10. 교통사고 발생지점에 대한 도시 별 Bubble Map



Bubble Map을 사용하기 위해 도시 별 위도, 경도, 사망자수의 데이터셋을 만들고, astype을 통해 형변환을 진행하였습니다.

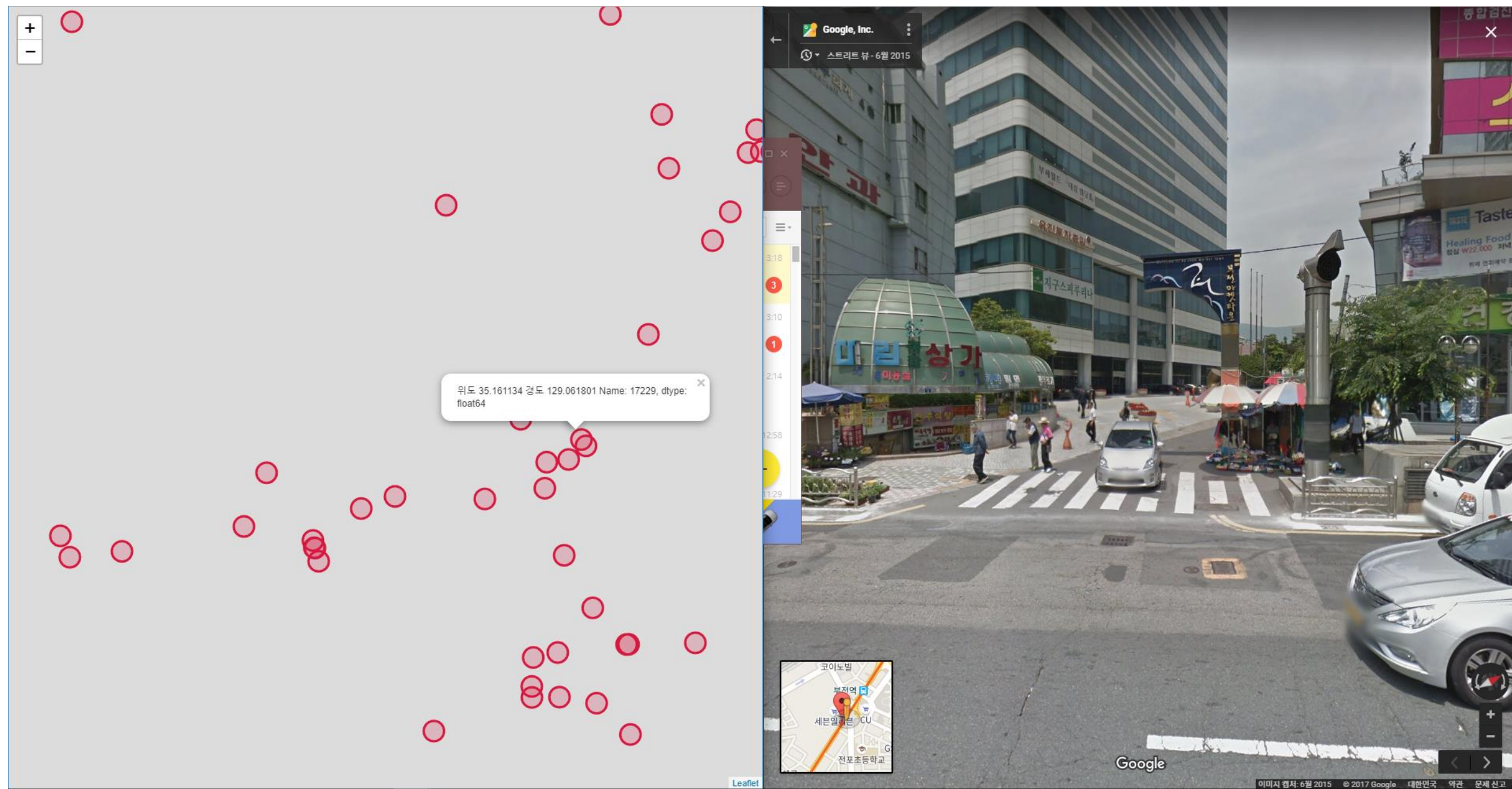
교통사망사고 분석 및 대응방안

Team04 남윤철¹, 이승훈¹, 장용주^{1*}

¹College of Transdisciplinary Studies, DGIST

분석 결과 및 대응 방안

앞서 분석한 결과를 통해 사고가 많이 난 지역에 대해 도로분석을 진행하였습니다.



주간	안전운전 의무 불이행	단일로	기타단일로	승용차	경형
야간	과속	단일로	기타단일로	승용차	중형
야간	안전운전 의무 불이행	단일로	기타단일로	승용차	중형
야간	안전운전 의무 불이행	단일로	횡단보도상	승용차	중형
야간	안전운전 의무 불이행	단일로	기타단일로	승용차	중형

Dot map을 토대로 찾은 이 지역은 총 5개의 교통사망사고가 발생하였습니다. 위의 사진을 보면 교차로 앞 시야를 방해하는 노점상과 큰 나무가 있는 것을 알 수 있었습니다. 이는 충분히 횡단보도 근처에 있는 사람들이 차가 안보일수도 있으며 차를 운전하는 사람 또한 횡단보도에 있는 사람을 보지 못할 수 있습니다. 또한 저렇게 사람들이 많이 다니는 교차로에는 차가 과속을 하지 못하게 과속방지턱이 있어야 하는데 과속방지턱이 없어 차의 과속을 막을 방법이 없습니다. 또한 5개의 교통사고 중 4가지가 야간에 발생한 교통사고인데 이 도로에는 가로등이 없어 야간에 매우 어둡다는 것을 알 수 있습니다.



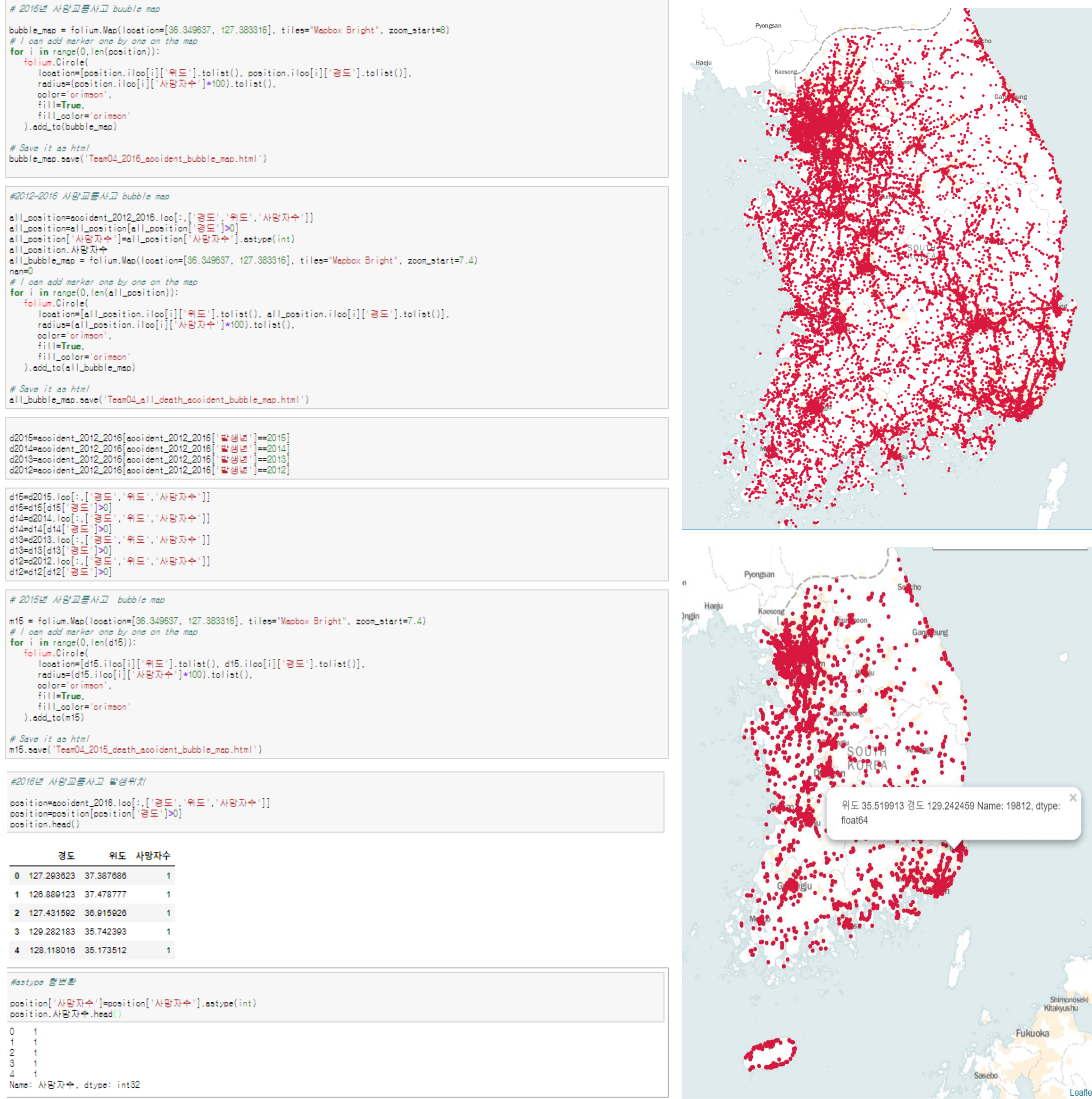
앞에서 문제점을 다루었는데 이를 개선하기 위해서는 교차로 앞 과속방지턱을 설치하고, 교차로 앞 시야를 확보해야 할 것입니다. 마지막으로 야간에 어두워 길이 잘 보이지 않기 때문에 교차로 부근 가로등 또한 설치해야 합니다.

분석 의의 및 발전 가능성

교통사고 데이터를 다양한 관점으로 분석을 시도한 결과, 도로유형이 '기타단일로' 이면사 사고유형이 '횡단 중' 일 때가 가장 많음을 파악하고 이에 해당하는 사건의 위치를 지도 위에서 인접한 거리 기준으로 분류해 교통사고가 밀집 발생한 곳을 특정 지을 수 있었습니다. 이번 분석의 의의는 특정 유형의 사고나 특정 유형의 도로에서 사고를 전국적으로 조회할 수 있고 이 지점들의 지리적인 문제점을 유형화하여 대처방안을 수립할 수 있다는 점입니다. 이를 확장하면 '기타단일로' 에서 '횡단 중' 발생한 교통사고 뿐 아니라 특정 도로유형에서 특정 사고유형이 일어나는 지점의 주변 도로상황을 파악하여 신호등을 추가 설치하거나 가로등을 추가 설치하는 등 실질적인 방안을 제안할 수 있습니다.

또한 추후 교통사고 데이터의 축적으로 특정 지점에서 일어난 사고수가 n개 이상인 지점을 선택할 수 있는 알고리즘을 제작을 통해 모든 도로유형 & 사고유형 별 교통사망사고가 일어나는 지점에 대한 정보를 얻을 수 있습니다. 이를 통해 신호등이 없어서 사고가 일어난 지점 수와 같이 지역적 정보에 대한 교통사고의 수를 뽑아 낼 수 있을 것입니다.

11. 교통사고 발생지점에 대한 Dot Map



위의 코드는 기존의 Bubble Map을 통해 원의 크기를 줄여 만든 코드와 Dot Map입니다. Dot Map에 있는 점을 누르면 위도, 경도를 알 수 있도록 설정해두었습니다. 지도를 확대할 수도 있기 때문에 지도를 확대한 후 같은 위도, 경도에 점이 여러 개 찍혀있다면 이는 그 도로에서만 많은 교통사고가 발생했다는 것을 알 수 있습니다. 이를 통해 그 도로에서 무엇이 문제인지 알아보고, 개선점에 대해서 분석해보려 한다. 실제로 확대하여 보면 같은 지점에서 원 두 개 겹친 경우의 수는 그렇게 많지 않습니다. 하지만 같은 지점에 교통사망사고가 더 많이 일어난다면 그곳에는 지역적인 문제가 있을 것이라 예상하였고 그래서 같은 지점에서 세 번 이상 겹치는 경우의 수를 세고 각 지점에는 어떤 문제가 있는지 로드뷰를 통해 살펴보기로 하였습니다.