

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
file_path = '../data/carAccident.csv'
carAccident = pd.read_csv(file_path, encoding='cp949')
carAccident.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 228 entries, 0 to 227
Data columns (total 8 columns):
#   Column      Non-Null Count  Dtype
---  -
0   시도        228 non-null    object
1   시군구      228 non-null    object
2   발생건수    228 non-null    int64
3   사망자수    228 non-null    int64
4   부상자수    228 non-null    int64
5   중상        228 non-null    int64
6   경상        228 non-null    int64
7   부상신고    228 non-null    int64
dtypes: int64(6), object(2)
memory usage: 14.4+ KB
```

In [3]:

```
carAccident.describe()
```

Out[3]:

	발생건수	사망자수	부상자수	중상	경상	부상신고
count	228.000000	228.000000	228.000000	228.000000	228.000000	228.000000
mean	85.004386	1.517544	144.526316	29.478070	108.381579	6.666667
std	88.952717	1.845258	155.724353	29.468485	122.590788	8.205152
min	3.000000	0.000000	3.000000	1.000000	1.000000	0.000000
25%	21.000000	0.000000	31.750000	9.750000	21.750000	1.000000
50%	59.500000	1.000000	94.000000	20.000000	65.000000	4.000000
75%	114.000000	2.000000	194.500000	39.000000	145.250000	10.000000
max	543.000000	12.000000	921.000000	155.000000	726.000000	54.000000

In [4]:

```
carCity=carAccident.groupby('시도')
```

In [5]:

```
# 도시별 수치형 데이터
carAccident.groupby('시도').describe()
```

Out[5]:

	발생건수								사망자수		...	경
	count	mean	std	min	25%	50%	75%	max	count	mean	...	7!
시도												
강원	18.0	37.722222	50.094975	7.0	11.00	20.0	33.00	201.0	18.0	0.777778	...	4
경기	31.0	164.193548	132.610563	14.0	60.00	133.0	225.50	543.0	31.0	2.000000	...	29
경남	18.0	53.222222	61.900827	7.0	17.25	26.5	65.25	245.0	18.0	2.388889	...	0
경북	22.0	54.090909	62.628085	4.0	15.25	26.0	70.75	229.0	22.0	1.636364	...	3
광주	5.0	135.000000	78.727378	39.0	61.00	175.0	195.00	205.0	5.0	0.800000	...	30
대구	8.0	111.250000	51.861767	55.0	67.50	104.5	141.00	202.0	8.0	2.250000	...	10
대전	5.0	118.800000	70.022853	51.0	67.00	97.0	159.00	220.0	5.0	2.000000	...	20
부산	16.0	46.437500	26.872461	16.0	30.00	43.5	52.25	117.0	16.0	0.625000	...	0
서울	25.0	114.240000	66.022269	44.0	79.00	97.0	119.00	373.0	25.0	1.080000	...	10
세종	1.0	98.000000	NaN	98.0	98.00	98.0	98.00	98.0	1.0	0.000000	...	10
울산	5.0	113.400000	42.559370	64.0	95.00	103.0	127.00	178.0	5.0	2.800000	...	10
인천	10.0	93.800000	67.791838	12.0	25.50	97.0	145.75	199.0	10.0	0.400000	...	10
전남	22.0	41.272727	43.274519	7.0	15.75	23.5	33.25	150.0	22.0	1.136364	...	0
전북	14.0	48.785714	68.924907	3.0	9.00	31.5	50.00	266.0	14.0	1.214286	...	0
제주	2.0	161.000000	77.781746	106.0	133.50	161.0	188.50	216.0	2.0	1.000000	...	20
충남	15.0	87.200000	110.823025	9.0	31.50	52.0	97.00	458.0	15.0	2.666667	...	10
충북	11.0	80.181818	123.301921	11.0	17.00	18.0	80.50	433.0	11.0	1.818182	...	0

17 rows × 48 columns

In [6]:

시도 별 사망자수의 합계.

```
carDeathTotal = carAccident.groupby('시도').사망자수.sum()  
carDeathTotal
```

Out[6]:

시도

강원 14

경기 62

경남 43

경북 36

광주 4

대구 18

대전 10

부산 10

서울 27

세종 0

울산 14

인천 4

전남 25

전북 17

제주 2

충남 40

충북 20

Name: 사망자수, dtype: int64

In [7]:

```
#시도 별 사망자수의 합계. 사망자 순으로 가장 많은 곳
citySum=carCity.sum() #시도 별 합계
citySum.sort_values(by=['사망자수'], ascending=False)
```

Out[7]:

	발생건수	사망자수	부상자수	중상	경상	부상신고
시도						
경기	5090	62	8723	1698	6529	496
경남	958	43	1519	476	979	64
충남	1308	40	2132	544	1524	64
경북	1190	36	1898	480	1322	96
서울	2856	27	5005	963	3770	272
전남	908	25	1472	297	1138	37
충북	882	20	1503	339	1108	56
대구	890	18	1473	281	1109	83
전북	683	17	1182	220	949	13
울산	567	14	913	207	650	56
강원	679	14	1163	227	899	37
부산	743	10	1237	281	876	80
대전	594	10	1092	132	932	28
인천	938	4	1630	308	1269	53
광주	675	4	1301	128	1143	30
제주	322	2	551	102	395	54
세종	98	0	158	38	119	1

In [8]:

```
#시도의 수
numOfCity=len(carAccident['시도'].unique())
numOfCity
```

Out[8]:

17

In [9]:

```
carAccident['시도'].unique()
```

Out[9]:

```
array(['서울', '부산', '경기', '강원', '충북', '충남', '전북', '전남', '경북', '경남', '제주',  
      '대구', '인천', '광주', '대전', '울산', '세종'], dtype=object)
```

In [10]:

```
#시도 별 발생건수가 가장 많은 곳은?  
citySum.sort_values(by=['발생건수'], ascending=False)
```

Out[10]:

	발생건수	사망자수	부상자수	중상	경상	부상신고
시도						
경기	5090	62	8723	1698	6529	496
서울	2856	27	5005	963	3770	272
충남	1308	40	2132	544	1524	64
경북	1190	36	1898	480	1322	96
경남	958	43	1519	476	979	64
인천	938	4	1630	308	1269	53
전남	908	25	1472	297	1138	37
대구	890	18	1473	281	1109	83
충북	882	20	1503	339	1108	56
부산	743	10	1237	281	876	80
전북	683	17	1182	220	949	13
강원	679	14	1163	227	899	37
광주	675	4	1301	128	1143	30
대전	594	10	1092	132	932	28
울산	567	14	913	207	650	56
제주	322	2	551	102	395	54
세종	98	0	158	38	119	1

In [11]:

```
# 도시별 부상자수
citySum.sort_values(by=['부상자수'], ascending=False)
```

Out[11]:

	발생건수	사망자수	부상자수	중상	경상	부상신고
시도						
경기	5090	62	8723	1698	6529	496
서울	2856	27	5005	963	3770	272
충남	1308	40	2132	544	1524	64
경북	1190	36	1898	480	1322	96
인천	938	4	1630	308	1269	53
경남	958	43	1519	476	979	64
충북	882	20	1503	339	1108	56
대구	890	18	1473	281	1109	83
전남	908	25	1472	297	1138	37
광주	675	4	1301	128	1143	30
부산	743	10	1237	281	876	80
전북	683	17	1182	220	949	13
강원	679	14	1163	227	899	37
대전	594	10	1092	132	932	28
울산	567	14	913	207	650	56
제주	322	2	551	102	395	54
세종	98	0	158	38	119	1

In [12]:

```
#시/도를 list로
result=citySum.index.tolist()
city_name_list = carAccident[' 시도'].unique().tolist()
x_pos = np.arange(len(city_name_list))
```

In [15]:

```

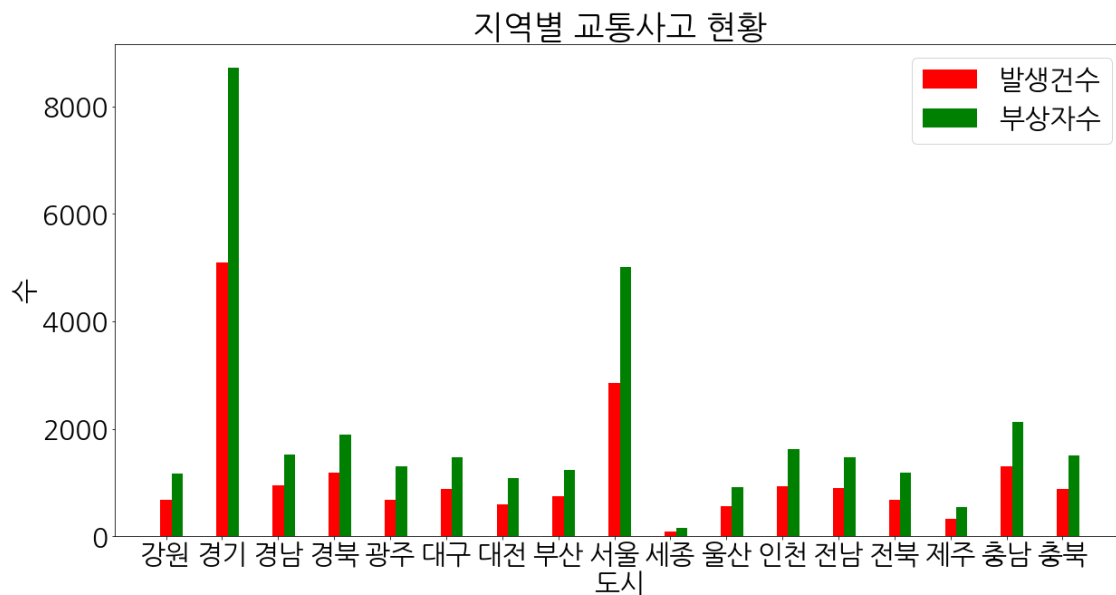
#발생건수
totalCase = carAccident.groupby('시도')['발생건수'].sum().tolist()
#사망자수
totalDeath = carAccident.groupby('시도')['사망자수'].sum().tolist()
#부상자수
totalInjury = carAccident.groupby('시도')['부상자수'].sum().tolist()

index = np.arange(numOfCity)
bar_width = 0.2
plt.rc('font', family='NanumGothic') #한글을 그래프에 표시 방법1

rects1 = plt.bar(index, totalCase, bar_width, color='r', label = '발생건수')
rects2 = plt.bar(index + bar_width, totalInjury, bar_width, color='g', label = '부상자수')
#rects3 = plt.bar(index + bar_width *2, totalDeath, bar_width, color='b', label = '사망자수')

plt.xticks(index, result)
plt.legend()
plt.xlabel('도시')
plt.ylabel('수')
plt.title('지역별 교통사고 현황')
plt.rcParams['figure.figsize'] = [20, 10] # [width, height] (inches) 그래프 크기
plt.rc('font', size=30)
plt.show()

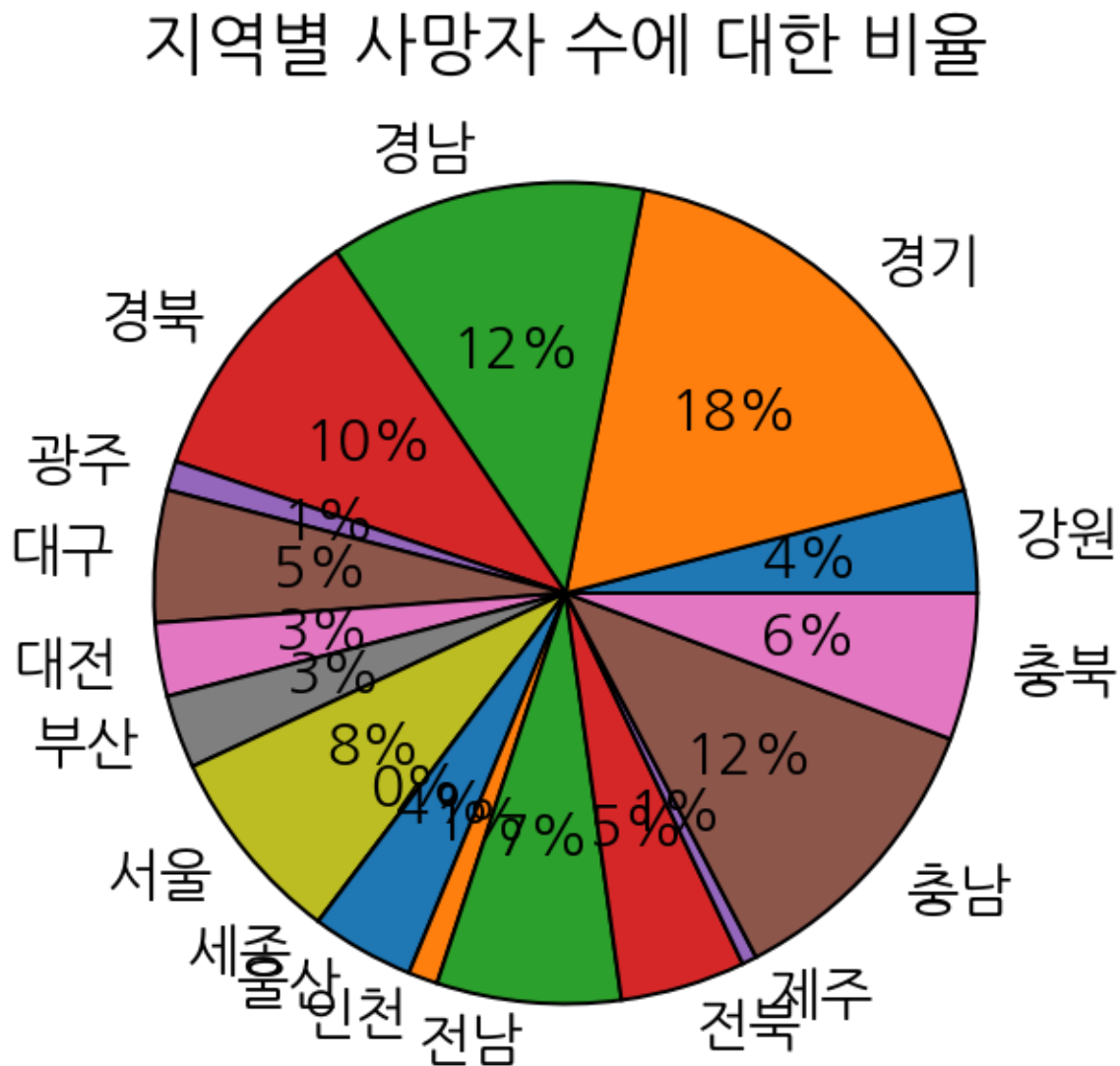
```



In []:

In [14]:

```
plt.pie(totalDeath, labels=result, autopct='%0.0f%%', wedgeprops={"linewidth":2, "edgecolor":"black"},
plt.title('지역별 사망자 수에 대한 비율')
plt.show()
```



In []: