진료데이터 시각화

In [13]:

```
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib
plt.rc('font', family='NanumGothic')
import numpy as np
import seaborn as sns
%matplotlib inline
import plotly.express as px
```

1. 년도별 데이터 확인

In [2]:

```
medi_2016 = pd.read_csv('../../lawdata/medical/NHIS_2016_Fixed.csv')
medi_2016.info()
medi_2016.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12540347 entries, 0 to 12540346
Data columns (total 8 columns):
 #   Column    Dtype 
---  ------    ----- 
 0   가입자일련번호   int64 
 1   성별코드      int64 
 2   연령대코드     int64 
 3   시도코드      int64 
 4   주상병코드     object
 5   요양개시일자    int64 
 6   시도        object
 7   질병코드      object
dtypes: int64(5), object(3)
memory usage: 765.4+ MB
```

Out[2]:

	가입자일련번호	성별코드	연령대코드	시도코드	주상병코드	요양개시일자	시도	질병코드
0	1	2	2	45	L209	20160817	전북	L
1	23	1	9	45	K769	20161031	전북	K
2	23	1	9	45	J042	20160920	전북	J
3	23	1	9	45	M511	20160405	전북	M
4	23	1	9	45	M511	20160330	전북	M

In [3]:

```
medi_2017 = pd.read_csv('../../lawdata/medical/NHIS_2017_Fixed.csv')
medi_2017.info()
medi_2017.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12568532 entries, 0 to 12568531
Data columns (total 8 columns):
 #   Column    Dtype 
---  ------    ----- 
 0   가입자일련번호   int64 
 1   성별코드      int64 
 2   연령대코드     int64 
 3   시도코드      int64 
 4   주상병코드     object
 5   요양개시일자    int64 
```

```
 6   시도        object
 7   질병코드      object
dtypes: int64(5), object(3)
memory usage: 767.1+ MB
```

Out[3]:

	가입자일련번호	성별코드	연령대코드	시도코드	주상병코드	요양개시일자	시도	질병코드
0	1	1	5	11	J060	20170316	서울	J
1	1	1	5	11	R51	20170123	서울	R
2	1	1	5	11	J0190	20170404	서울	J
3	1	1	5	11	J0190	20170407	서울	J
4	1	1	5	11	L309	20170516	서울	L

In [4]:

```
medi_2018 = pd.read_csv('../../lawdata/medical/NHIS_2018_Fixed.csv')
medi_2018.info()
medi_2018.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12974000 entries, 0 to 12973999
Data columns (total 8 columns):
 #   Column    Dtype
---  ------    -----
 0   가입자일련번호  int64
 1   성별코드     int64
 2   연령대코드    int64
 3   시도코드     int64
 4   주상병코드    object
 5   요양개시일자   int64
 6   시도        object
 7   질병코드     object
dtypes: int64(5), object(3)
memory usage: 791.9+ MB
```

Out[4]:

	가입자일련번호	성별코드	연령대코드	시도코드	주상병코드	요양개시일자	시도	질병코드
0	1	1	5	47	J209	20181120	경북	J
1	8	1	13	47	E148	20181106	경북	E
2	8	1	13	47	M6591	20180809	경북	M
3	8	1	13	47	E119	20180802	경북	E
4	8	1	13	47	K769	20180901	경북	K

2. 년도별 질병코드별 발병건수 시각화

2-1. 2016년 질병코드별 발병건수 시각화

In [6]:

```
def pie_chart(data, col, title = ''):
    cnt_df = data[col].value_counts().reset_index()
    fig = px.pie(cnt_df,
                 values = col,
                 names = 'index',
                 title = title,
                 template = 'seaborn',
                 color_discrete_sequence=px.colors.sequential.RdBu)
    fig.update_traces(rotation=90, pull=0.05, textinfo="value+percent+label")
    fig.show()
```

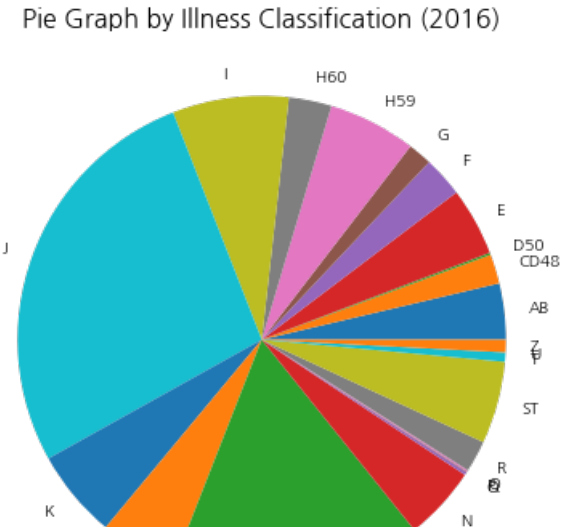
In [7]:

```
df = medi_2016.groupby('질병코드')['가입자일련번호'].count()
df = pd.DataFrame(df)

# plt.subplots(figsize=(20, 7))
pie_chart(medi_2016,'질병코드')

# plt.title('Pie Graph by Illness Classification (2016)', fontsize = 16)
#  plt.show()
```

2-2. 2017년 질병코드별 발병건수 시각화

In [8]:

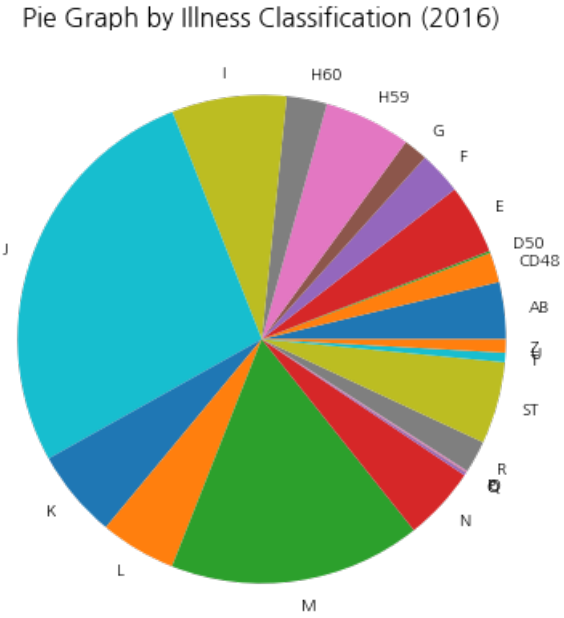
```
df = medi_2017.groupby('질병코드')['가입자일련번호'].count()
df = pd.DataFrame(df)

plt.subplots(figsize=(20, 7))
plt.pie(x = df['가입자일련번호'], labels= df.index)
plt.title('Pie Graph by Illness Classification (2016)', fontsize = 16)
plt.show()
```



2-3. 2018년 질병코드별 발병건수 시각화

In [9]:

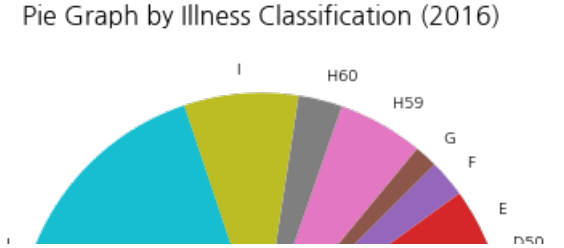
```
df = medi_2018.groupby('질병코드')['가입자일련번호'].count()
df = pd.DataFrame(df)

plt.subplots(figsize=(20, 7))
plt.pie(x = df['가입자일련번호'], labels= df.index)
plt.title('Pie Graph by Illness Classification (2016)', fontsize = 16)
plt.show()
```

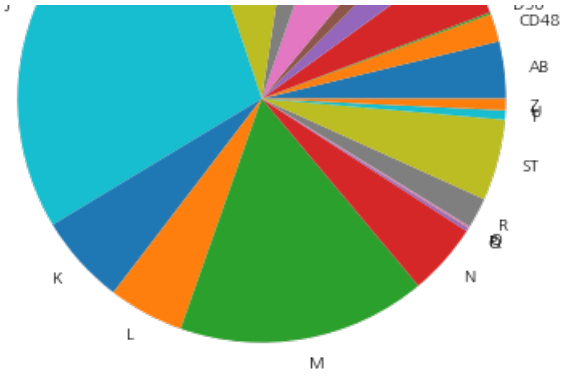


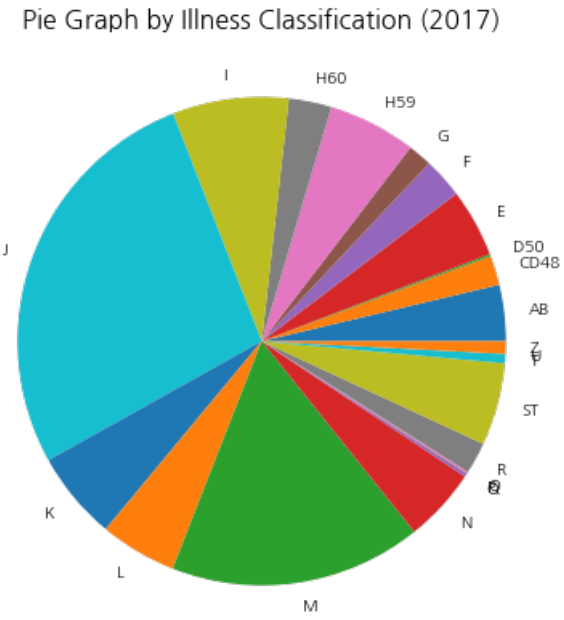
2-4. 반복문을 통한 년도별 질병코드별 발병건수 그래프 저장

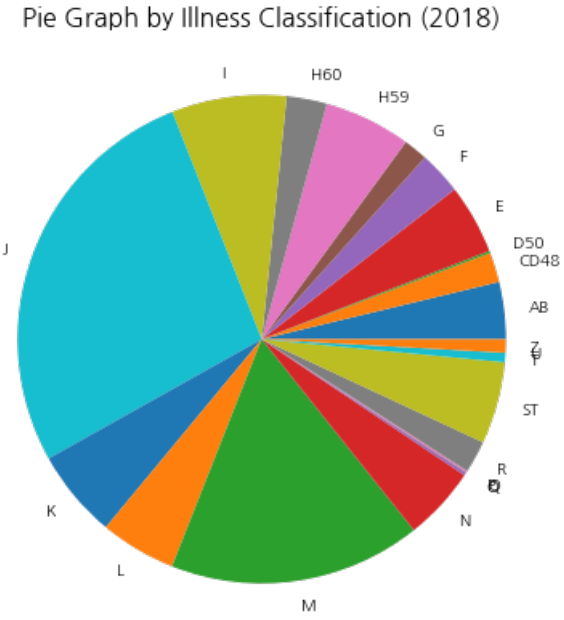
In [10]:

```
for i in [2016,2017,2018]:

    medi = pd.read_csv(f'../../lawdata/medical/NHIS_{i}_Fixed.csv')

    df = medi.groupby('질병코드')['가입자일련번호'].count()
    df = pd.DataFrame(df)

    plt.subplots(figsize=(20, 7))
    plt.pie(x = df['가입자일련번호'], labels= df.index)
    plt.title(f'Pie Graph by Illness Classification ({i})', fontsize = 16)

    plt.savefig(f'{i}년 질병코드별 발병건수')

    pass
```









3. 년도별 지역별 호흡기질환 발병건수 시각화

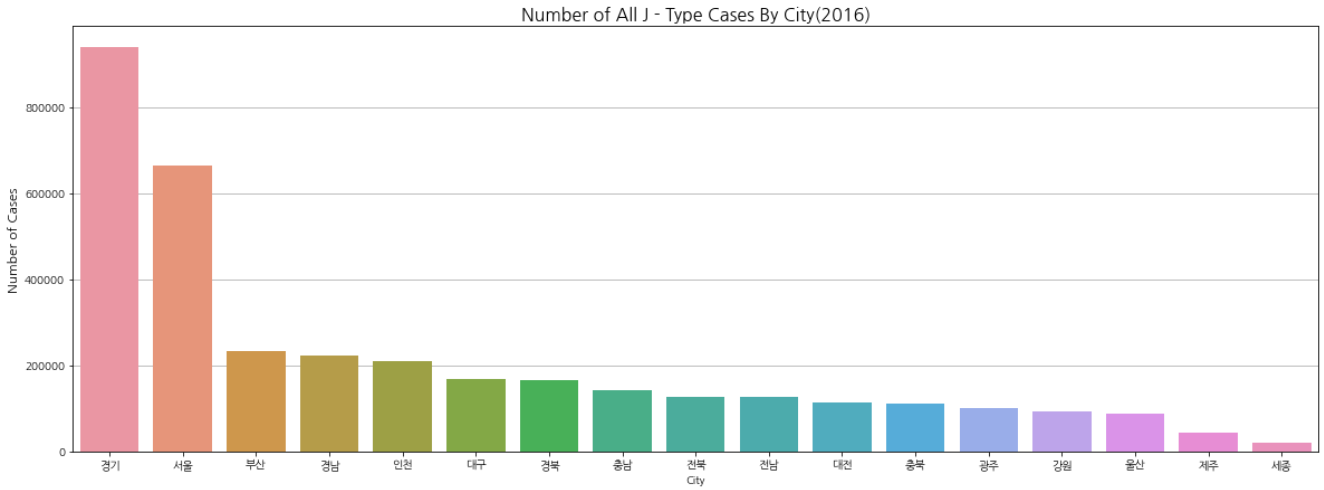
3-1. 2016년 지역별 호흡기질환 발병건수 시각화

In [21]:

```
medi_2016_j = medi_2016[medi_2016['질병코드']=='J']
medi_2017_j = medi_2017[medi_2017['질병코드']=='J']
medi_2018_j = medi_2018[medi_2018['질병코드']=='J']
```

In [22]:

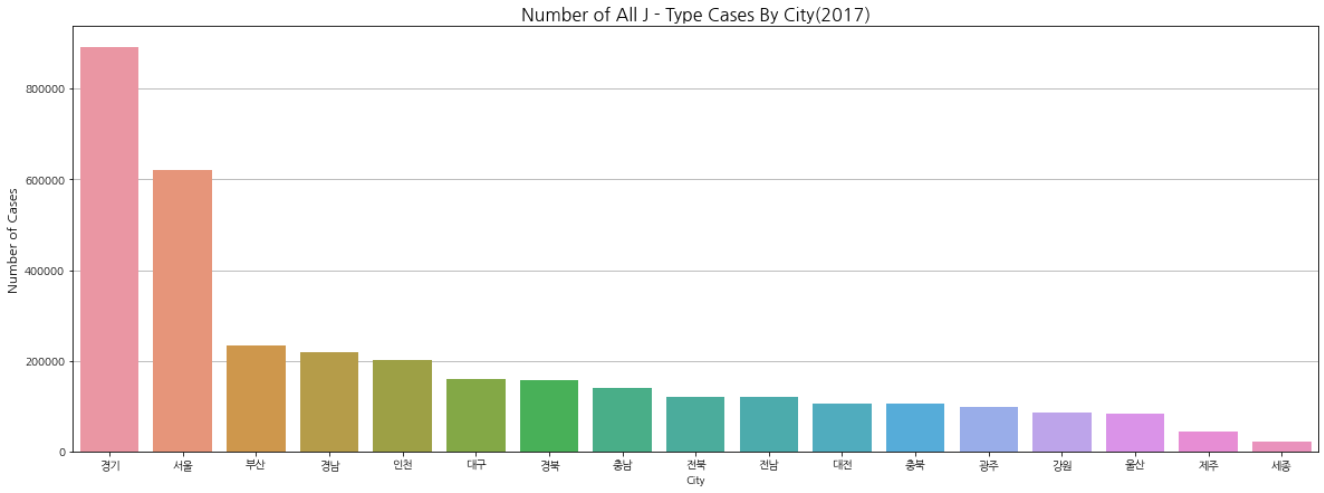
```
plt.subplots(figsize=(20, 7))
plt.grid(zorder = 1)
sns.countplot(x = '시도', orient = 'v', data = medi_2016_j, zorder = 2, order = medi_2016_j['시도'].v
alue_counts().index )

plt.xlabel('City', fontsize=10 )
plt.ylabel('Number of Cases', fontsize=12)
plt.title('Number of All J - Type Cases By City(2016)', fontsize=16)
plt.show()
```



3-2. 2017년 지역별 발병건수 시각화

In [23]:

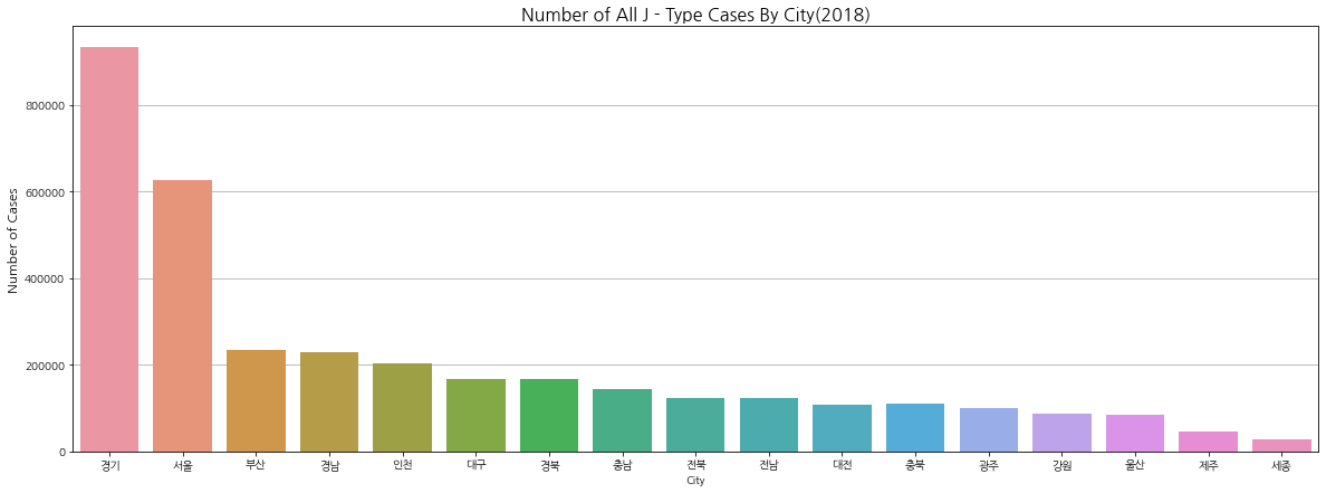
```
plt.subplots(figsize=(20, 7))
plt.grid(zorder = 1)
sns.countplot(x = '시도', orient = 'v', data = medi_2017_j, zorder = 2, order = medi_2016_j['시도'].v
alue_counts().index )

plt.xlabel('City', fontsize=10 )
plt.ylabel('Number of Cases', fontsize=12)
plt.title('Number of All J - Type Cases By City(2017)', fontsize=16)
plt.show()
```



3-3. 2018년 지역별 발병건수 시각화

In [24]:

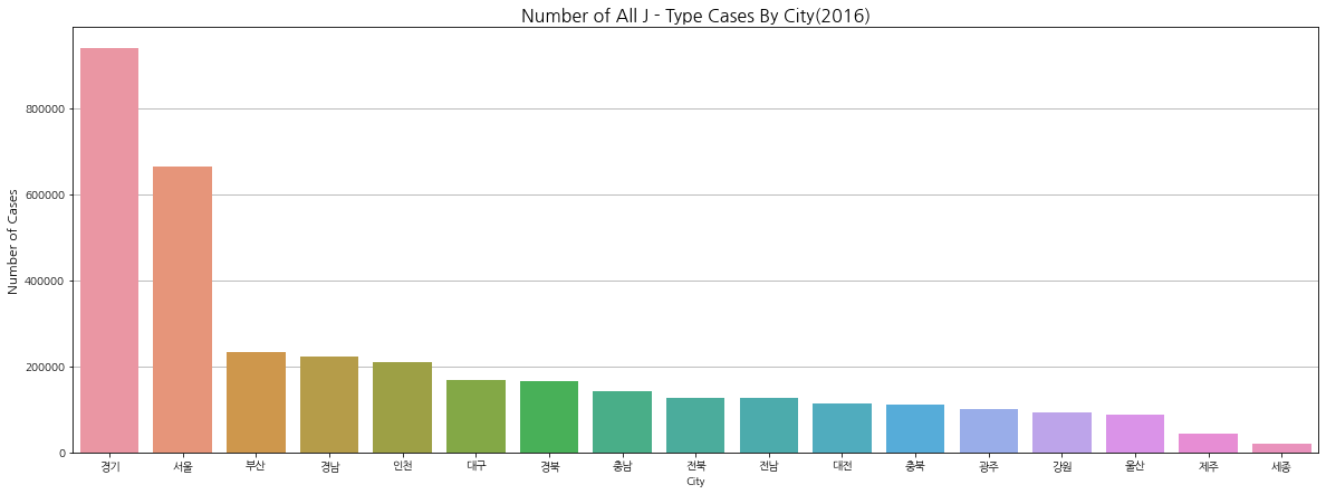
```
plt.subplots(figsize=(20, 7))
plt.grid(zorder = 1)
sns.countplot(x = '시도', orient = 'v', data = medi_2018_j, zorder = 2, order = medi_2016_j['시도'].v
alue_counts().index )

plt.xlabel('City', fontsize=10 )
plt.ylabel('Number of Cases', fontsize=12)
plt.title('Number of All J - Type Cases By City(2018)', fontsize=16)
plt.show()
```

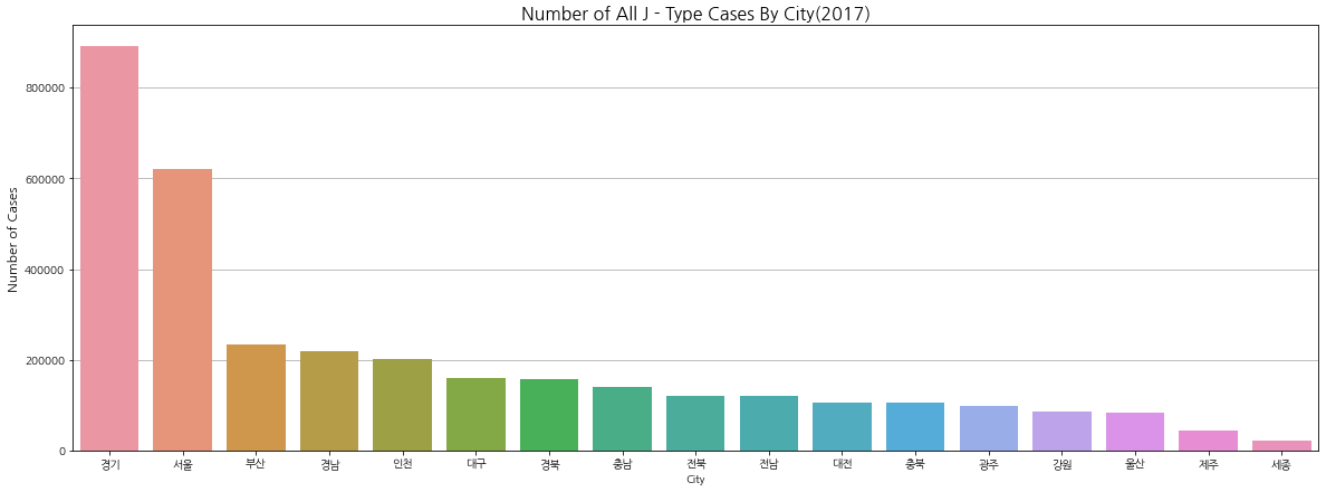


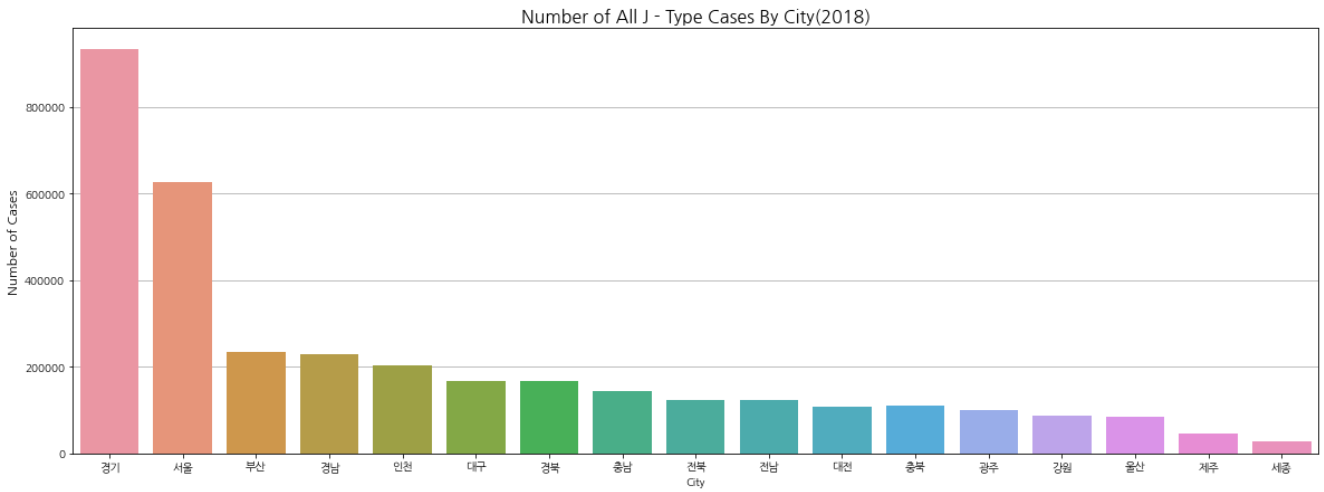
3-4. 반복문을 통한 년도별 호흡기 질환 발병건수 그래프 저장

In [25]:

```
for i in [2016,2017,2018]:

    medi = pd.read_csv(f'../../lawdata/medical/NHIS_{i}_Fixed.csv')
    medi_j = medi[medi['질병코드']=='J']

    plt.subplots(figsize=(20, 7))
    plt.grid(zorder = 1)
    sns.countplot(x = '시도', orient = 'v', data = medi_j, zorder = 2, order = medi_2016_j['시도'].va
lue_counts().index )

    plt.xlabel('City', fontsize=10 )
    plt.ylabel('Number of Cases', fontsize=12)
    plt.title(f'Number of All J - Type Cases By City({i})', fontsize=16)

    plt.savefig(f'{i}년 호흡기 질환 발병건수')

    pass
```




Number of All J - Type Cases By City(2017)
Number of Cases
800000
600000
400000
200000
0
경기
서울
부산
경남
인천
대구
경북
충남
전북
전남
대전
충북
광주
강원
울산
제주
세종
City


Number of All J - Type Cases By City(2018)
Number of Cases
800000
600000
400000
200000
0
경기
서울
부산
경남
인천
대구
경북
충남
전북
전남
대전
충북
광주
강원
울산
제주
세종
City