

코로나 확진자 추이에 따른 해상 수출 데이터 분석

해상운임과 물동량에 관해

데이터 수집

전세계 코로나 데이터는 다른 팀원분들이 구해놓은 것을 활용

해상운임 데이터와 컨테이너 수출실적 데이터는 동적 크롤링을 적당히 이용하여 추출

출처 : 관세청 - <https://unipass.customs.go.kr/ets/>

한국관세물류협회 - <https://www.kcla.kr/web/inc/html/4-2.asp>

데이터 전처리

필요 데이터를 날짜를 기준으로 정리하여 병합

```
fin_df = pd.concat([pdf1,pdf2],axis=1).fillna(0).sort_index()  
fin_df.head()
```

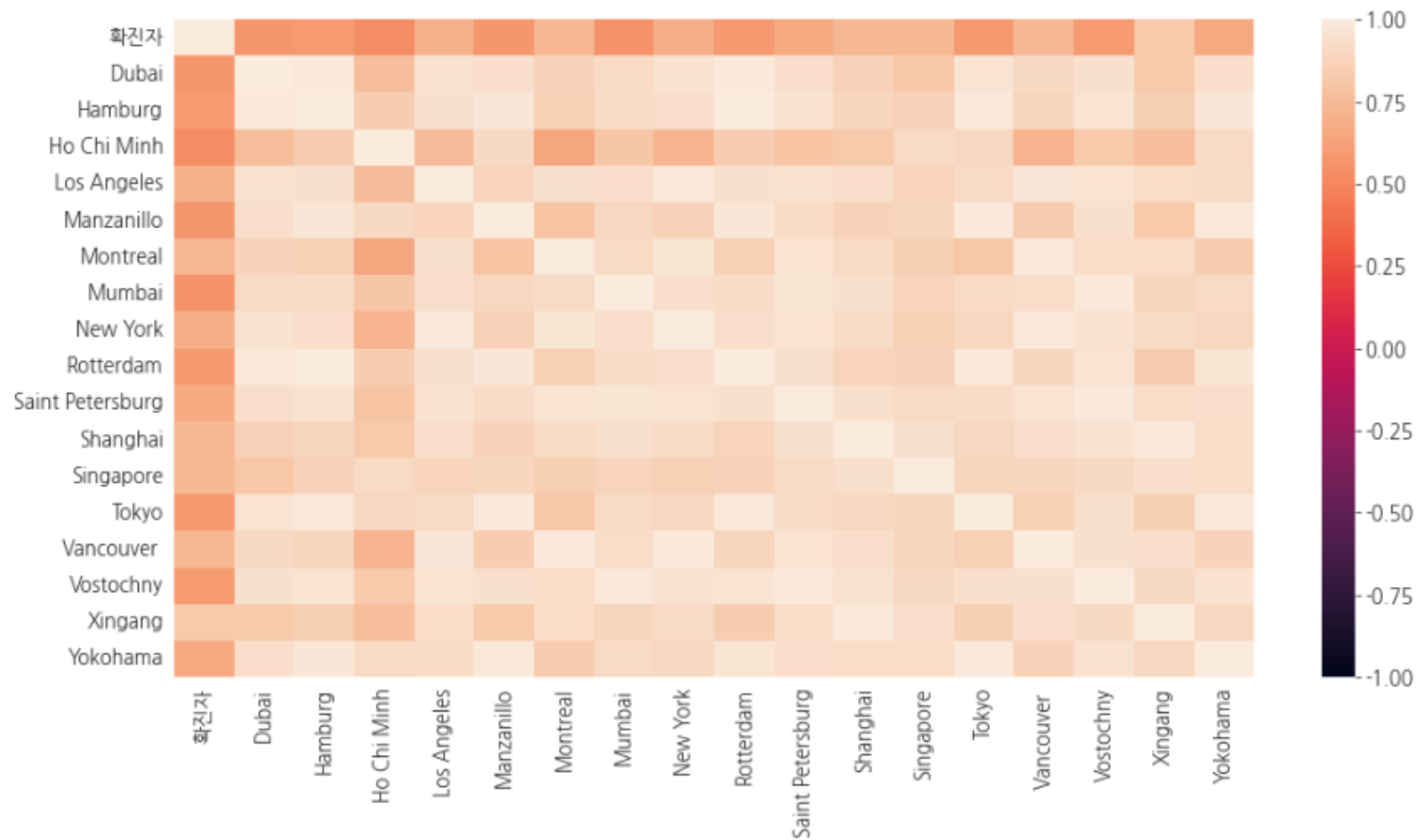
년월	확진자	Dubai	Hamburg	Ho Chi Minh	Los Angeles	Manzanillo	Montreal	Mumbai	New York	Rotterdam	Saint Petersburg	Shanghai	Singapore	Tokyo	Vancouver	Vosto
2019.01	0.0	566.7	1200.0	300.0	2253.3	2333.3	3243.3	1000.0	3246.7	1200.0	1766.7	56.7	260.0	203.3	1886.7	
2019.02	0.0	583.3	1183.3	300.0	2260.0	2300.0	3213.3	1000.0	3220.0	1183.3	1733.3	56.7	260.0	203.3	1876.7	
2019.03	0.0	600.0	1150.0	300.0	2126.7	2133.3	3100.0	1000.0	3086.7	1150.0	1666.7	56.7	260.0	203.3	1743.3	
2019.04	0.0	600.0	1066.7	290.0	2126.7	2183.3	3100.0	1000.0	3086.7	1066.7	1550.0	56.7	260.0	203.3	1743.3	
2019.05	0.0	600.0	1116.7	290.0	2180.0	2266.7	3213.3	1000.0	3140.0	1116.7	1633.3	56.7	260.0	203.3	1796.7	

데이터 분석

코로나 전후 확진자 수와 해상운임 간의 상관계수를 구한 뒤 heatmap으로 표현

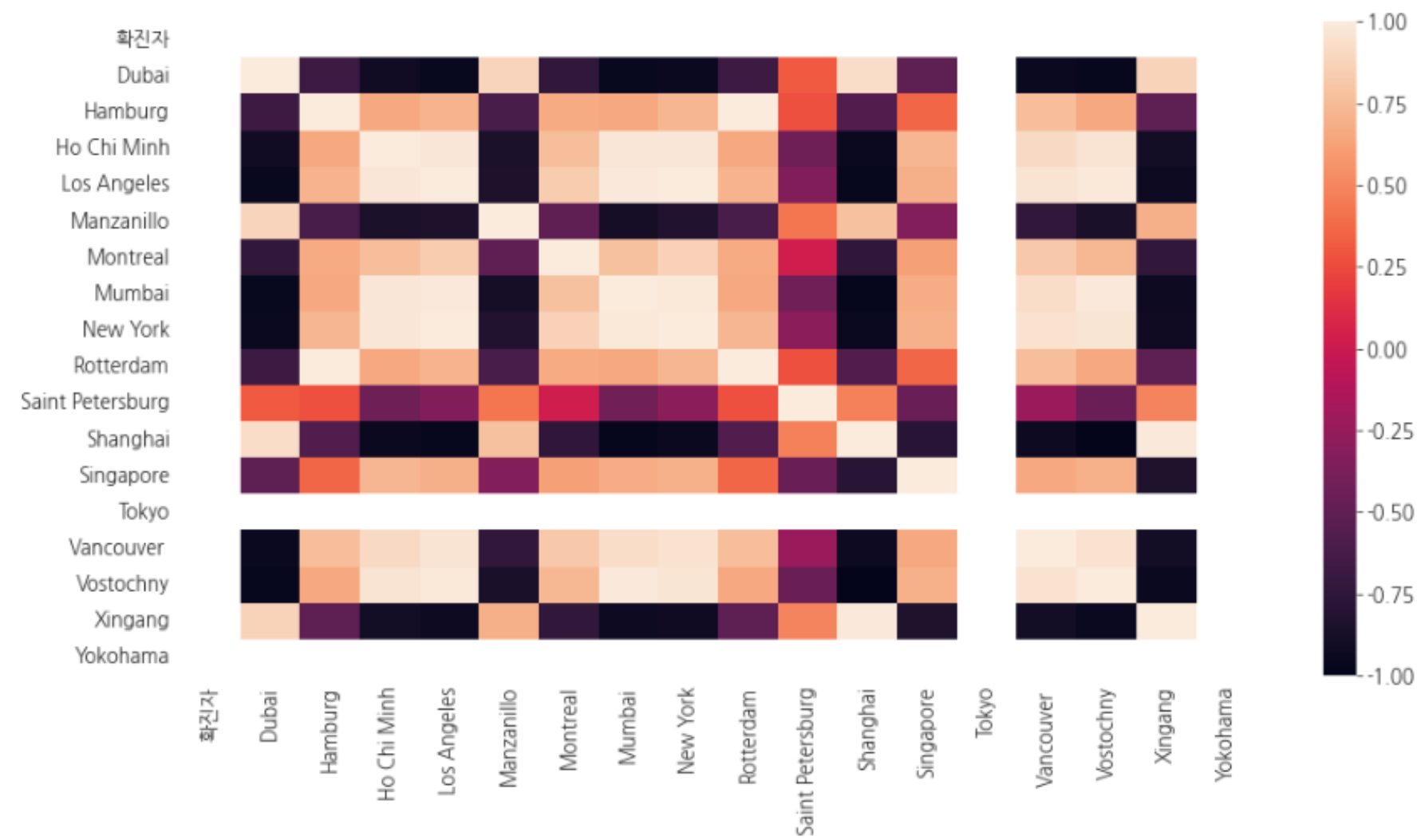
```
# 코로나 이후 상관계수의 heatmap  
plt.figure(figsize=(16,8))  
sns.heatmap(fin_df_after_corona.corr(),vmin=-1,vmax=1)
```

<AxesSubplot :>

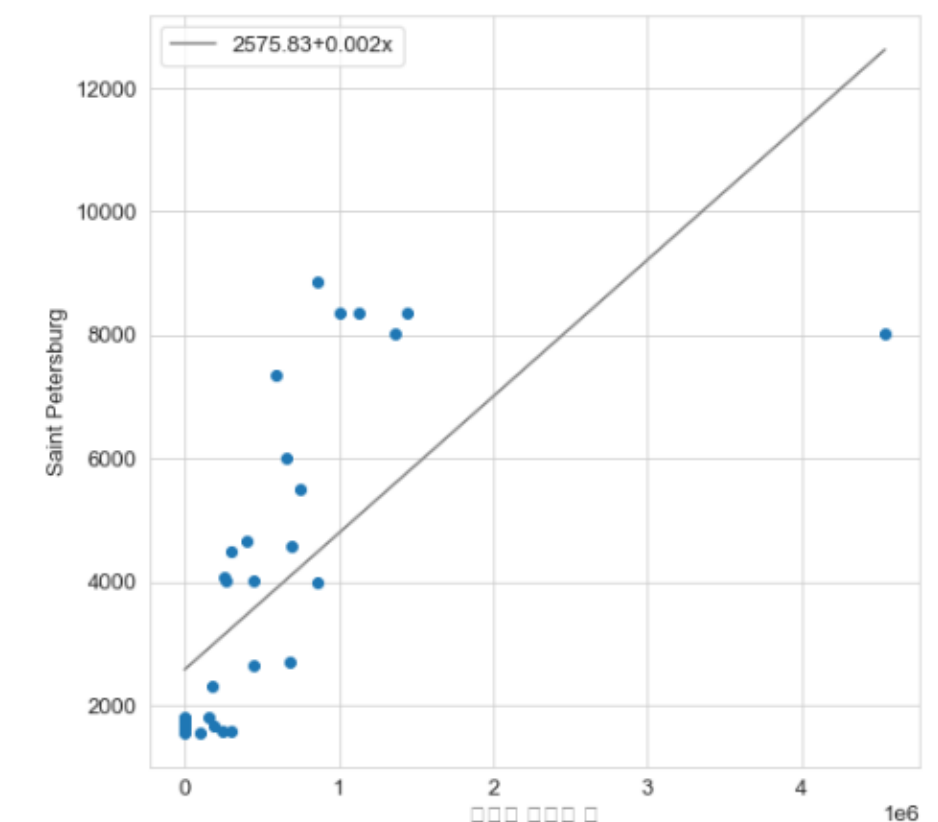
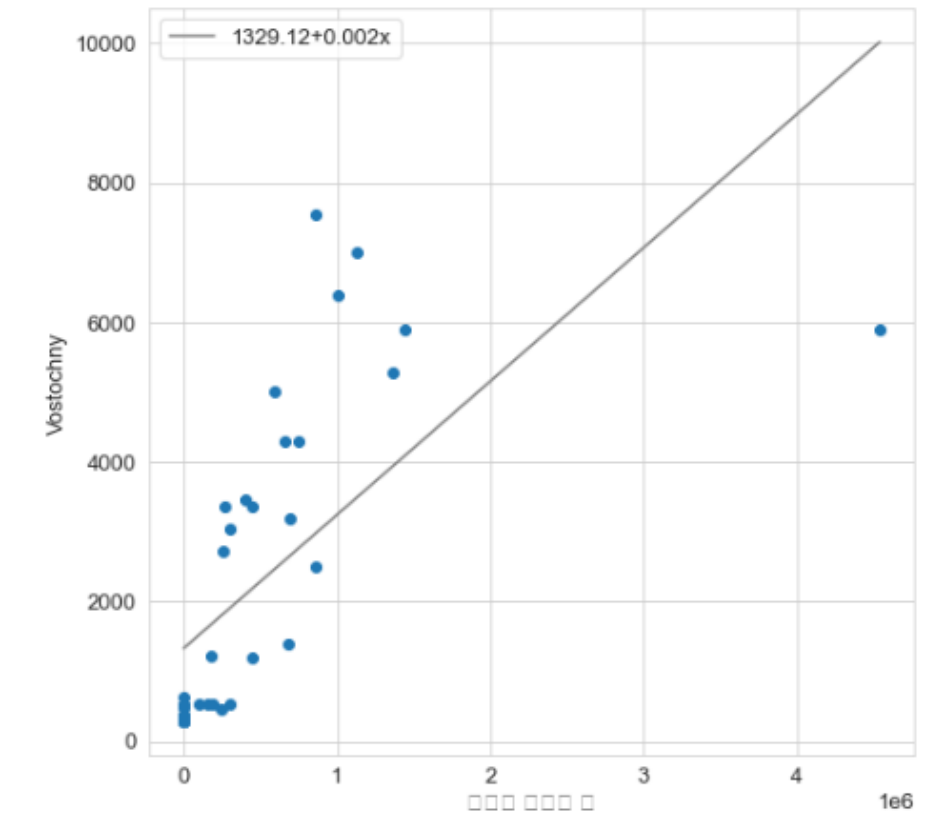
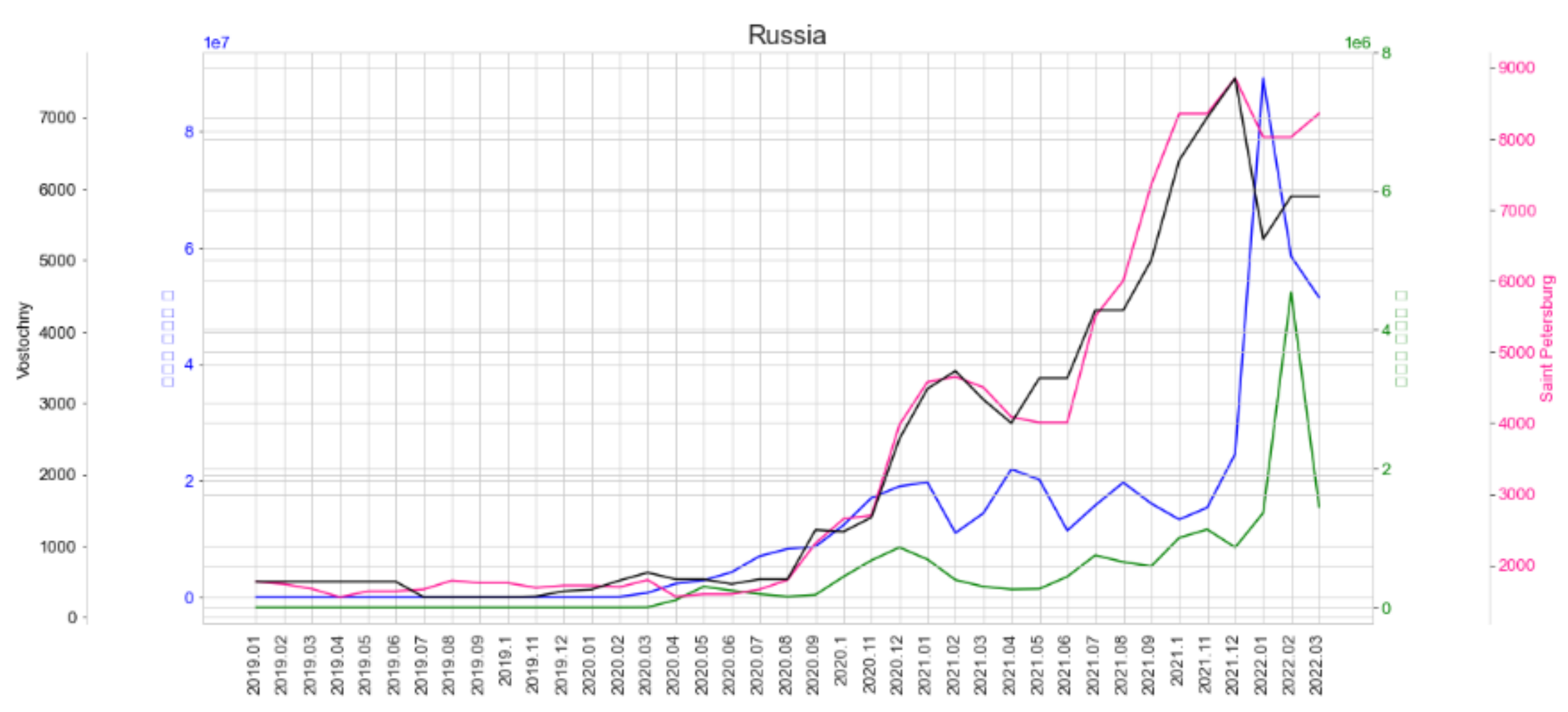


```
# 코로나 이전 상관계수의 heatmap  
plt.figure(figsize=(16,8))  
sns.heatmap(fin_df_before_corona.corr(),vmin=-1,vmax=1)
```

<AxesSubplot :>



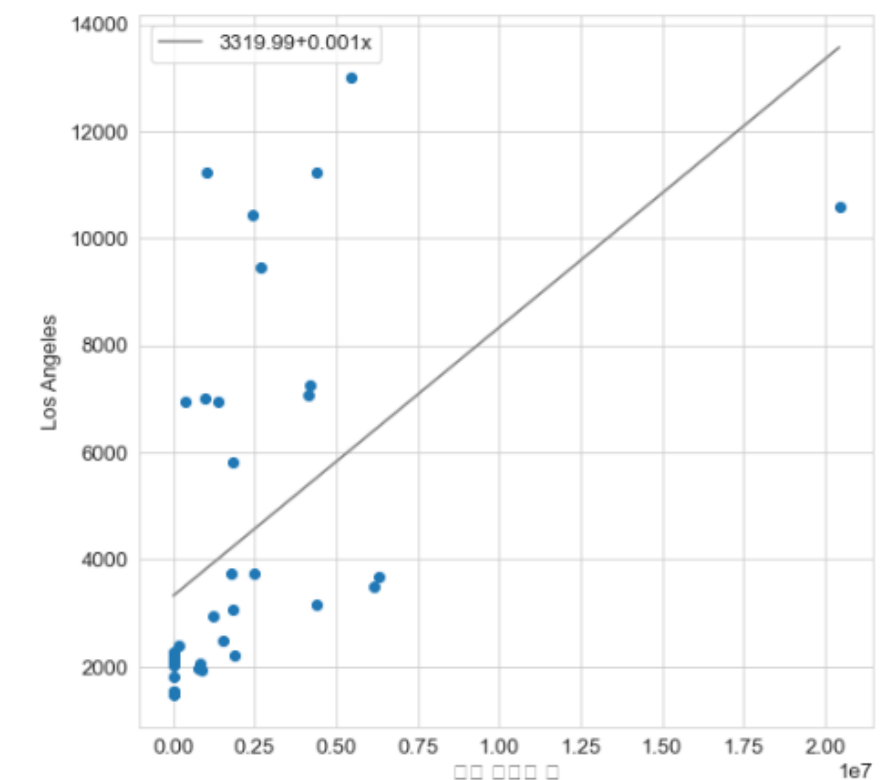
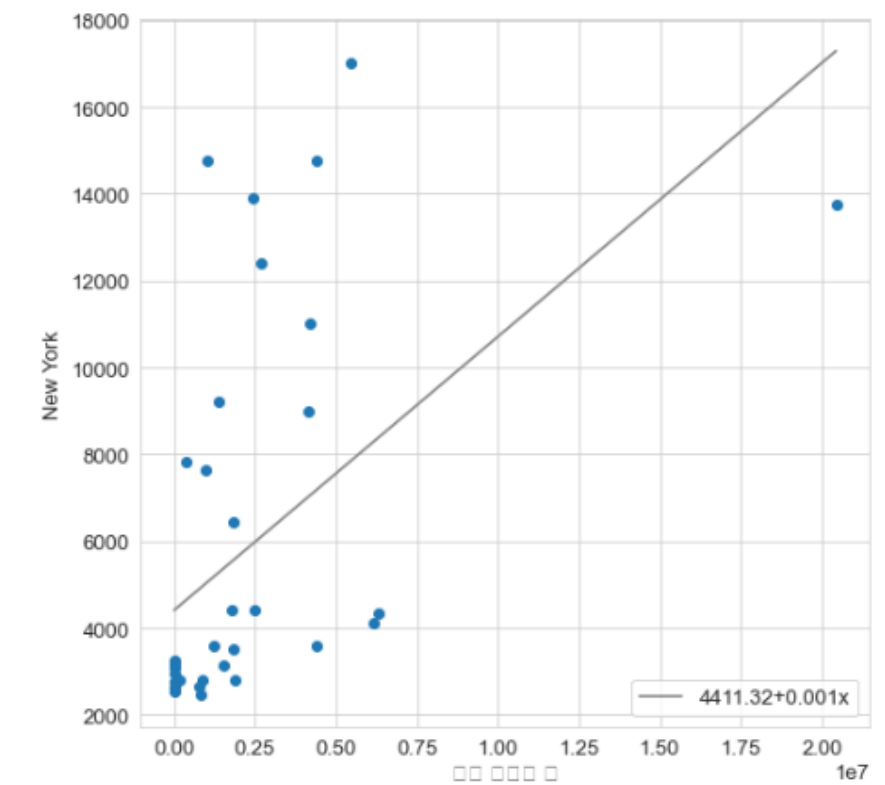
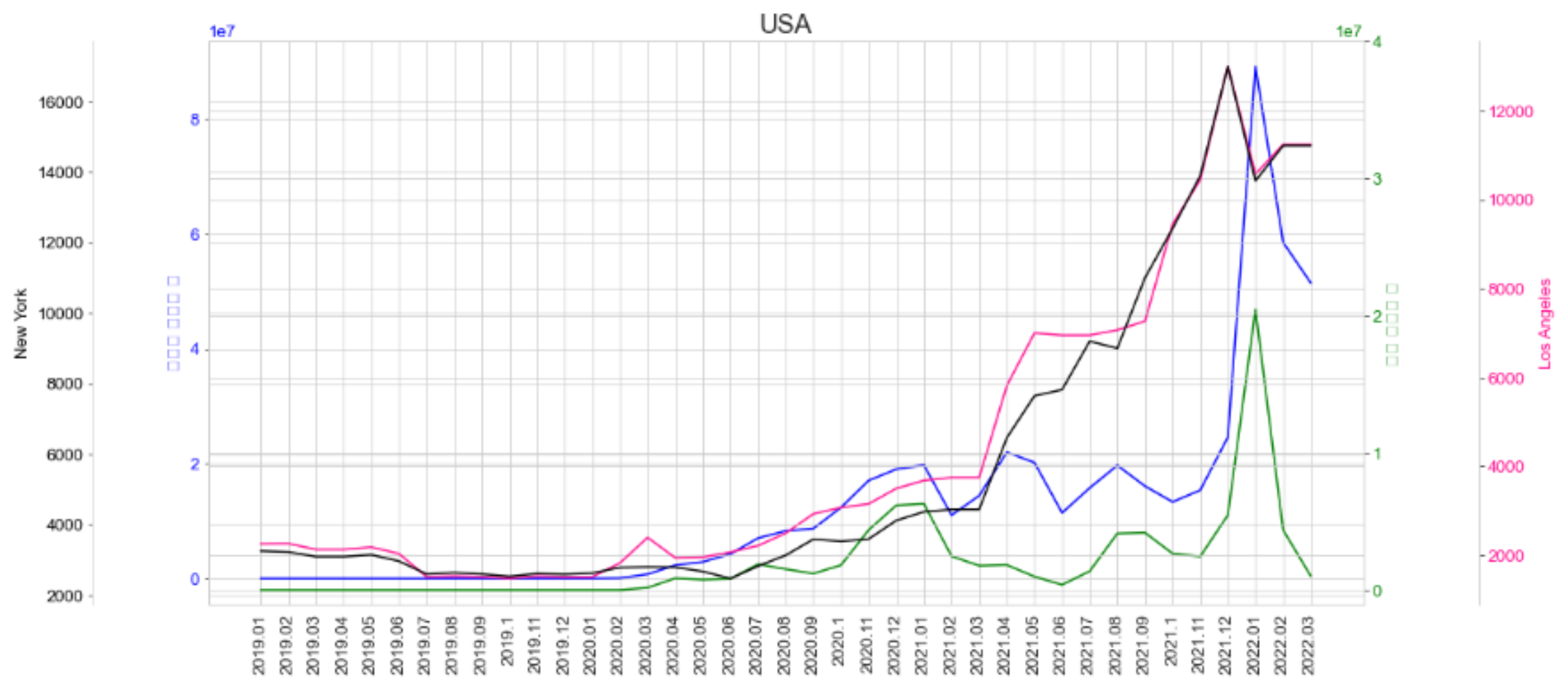
1)해상운임 데이터가 있는 국가에 대한 확진자 수와 해상운임의 시계열 그래프와 상관분석



```
# 피어슨 상관계수
from scipy.stats import *
print(pearsonr(fin_rus['확진자'],fin_rus['Saint Petersburg']))
print(pearsonr(fin_rus['확진자'],fin_rus['Vostochny']))
```

(0.6943510501409145, 9.445941854054339e-07)
(0.6664019328015185, 3.6367830224322452e-06)

1)해상운임 데이터가 있는 국가에 대한 확진자 수와 해상운임의 시계열 그래프와 상관분석

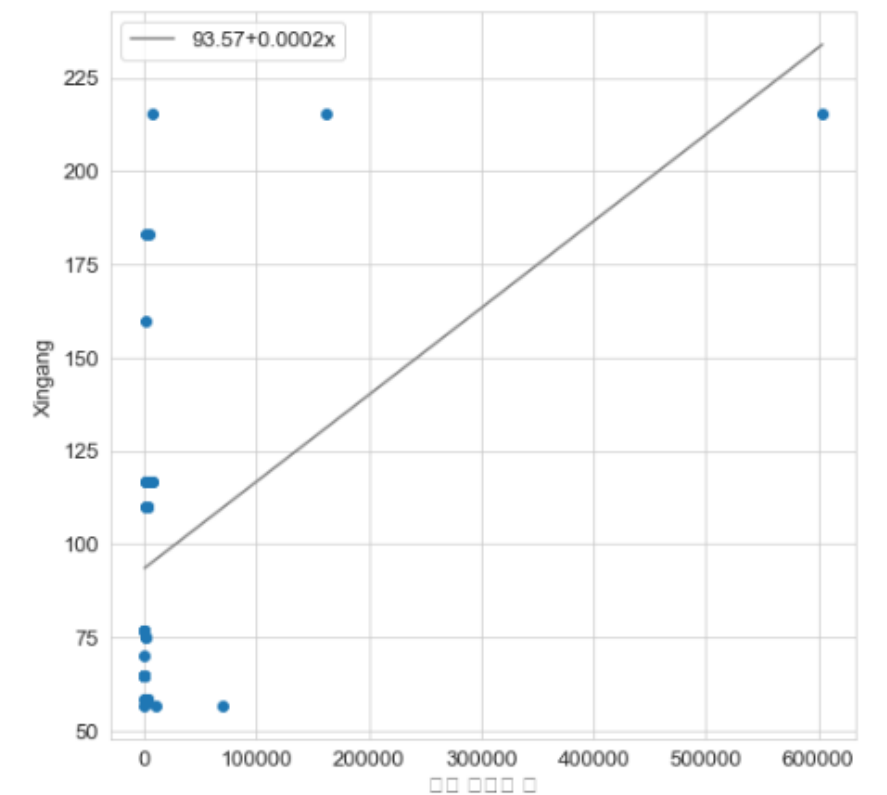
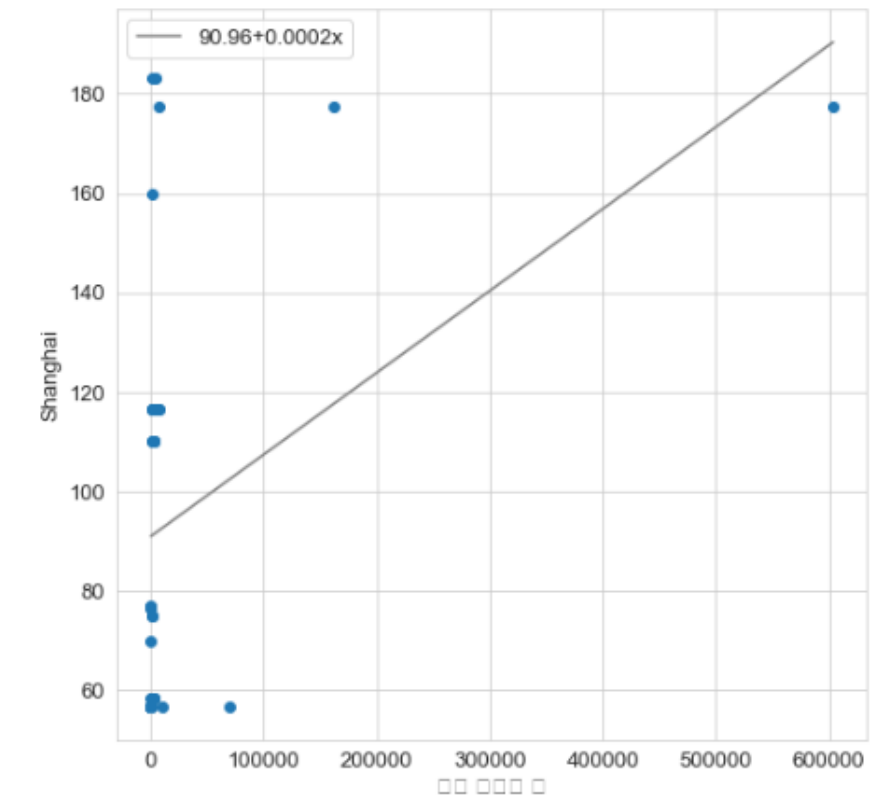
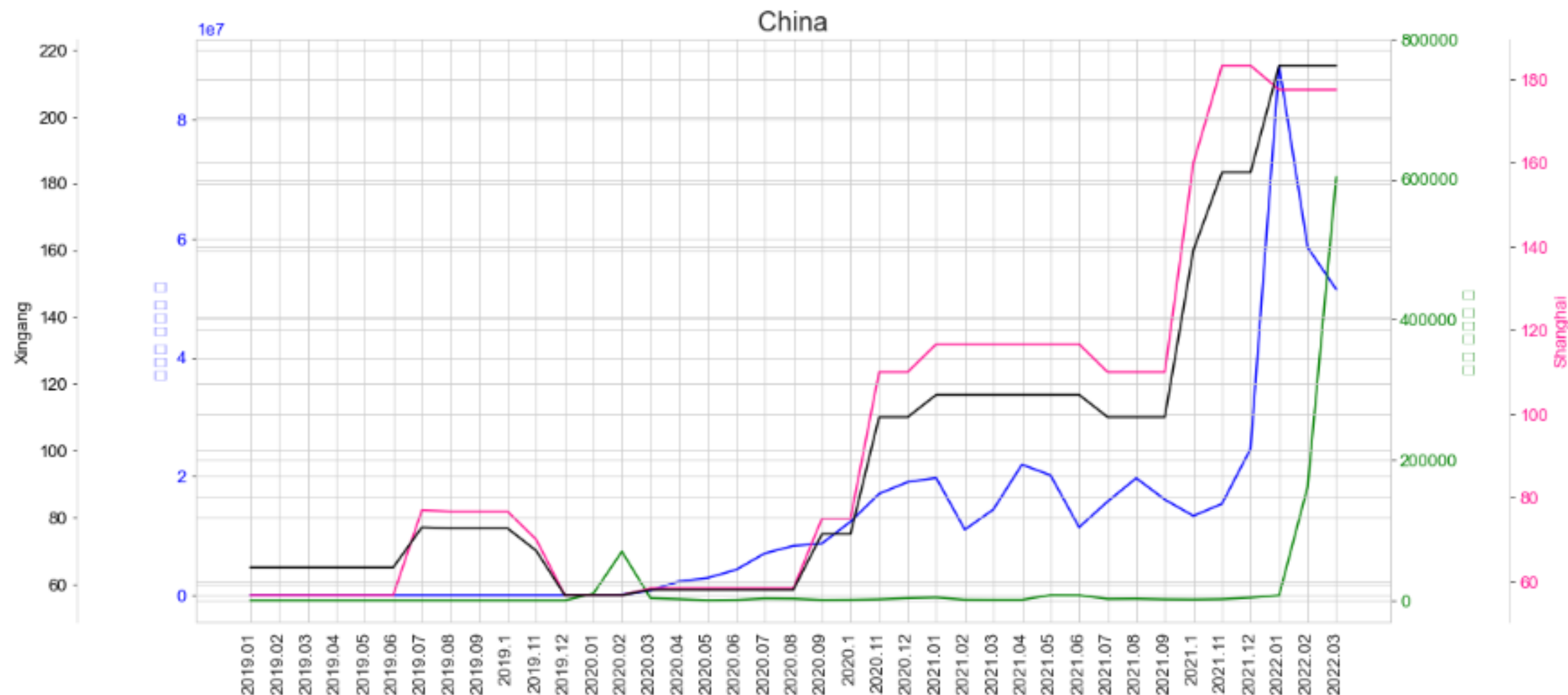


```
# 피어슨 상관계수
from scipy.stats import *
print(pearsonr(fin_usa['확진자'],fin_usa['Los Angeles']))
print(pearsonr(fin_usa['확진자'],fin_usa['New York']))
```

(0.5243199733775055, 0.0006121236583028439)

(0.5145844917762191, 0.0008036802950783147)

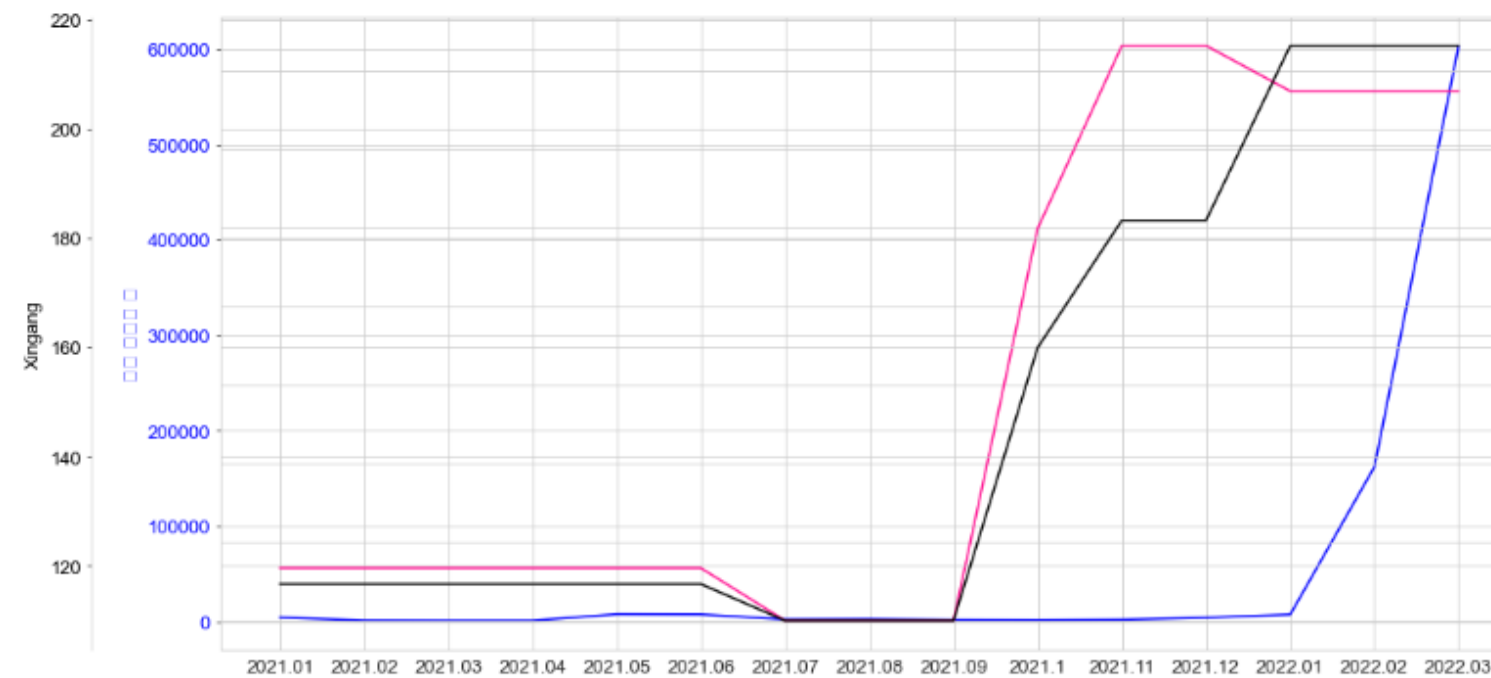
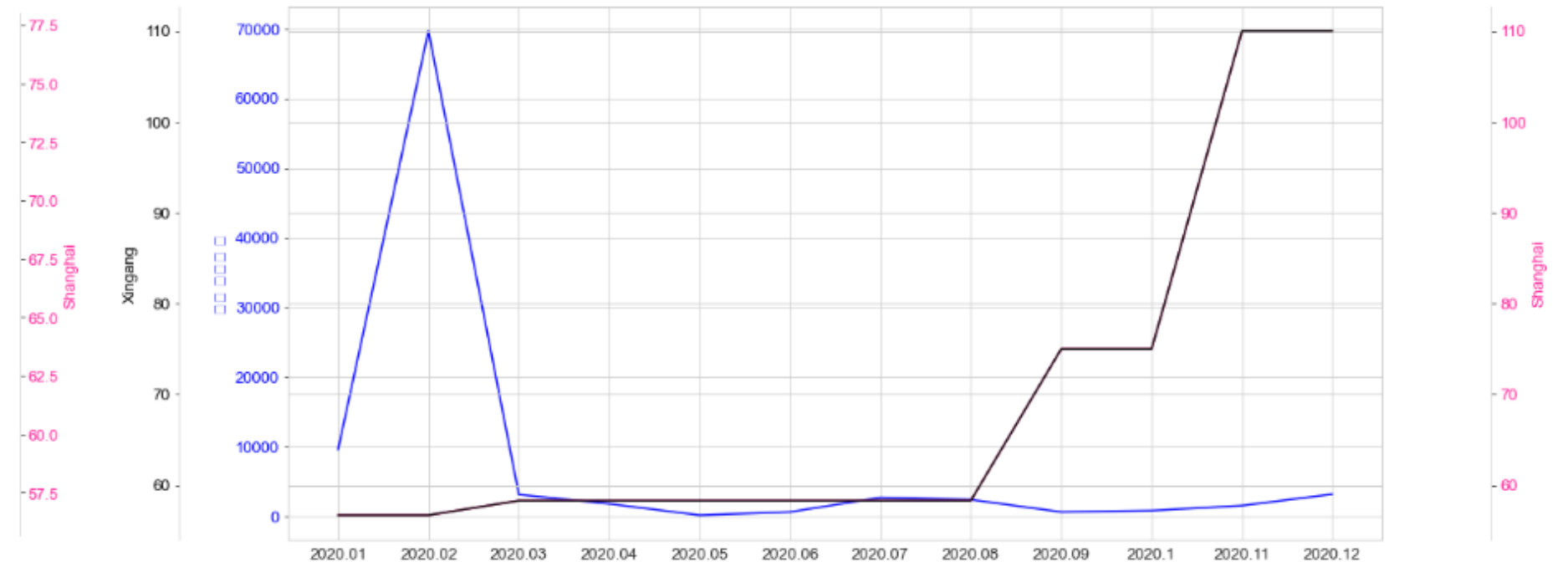
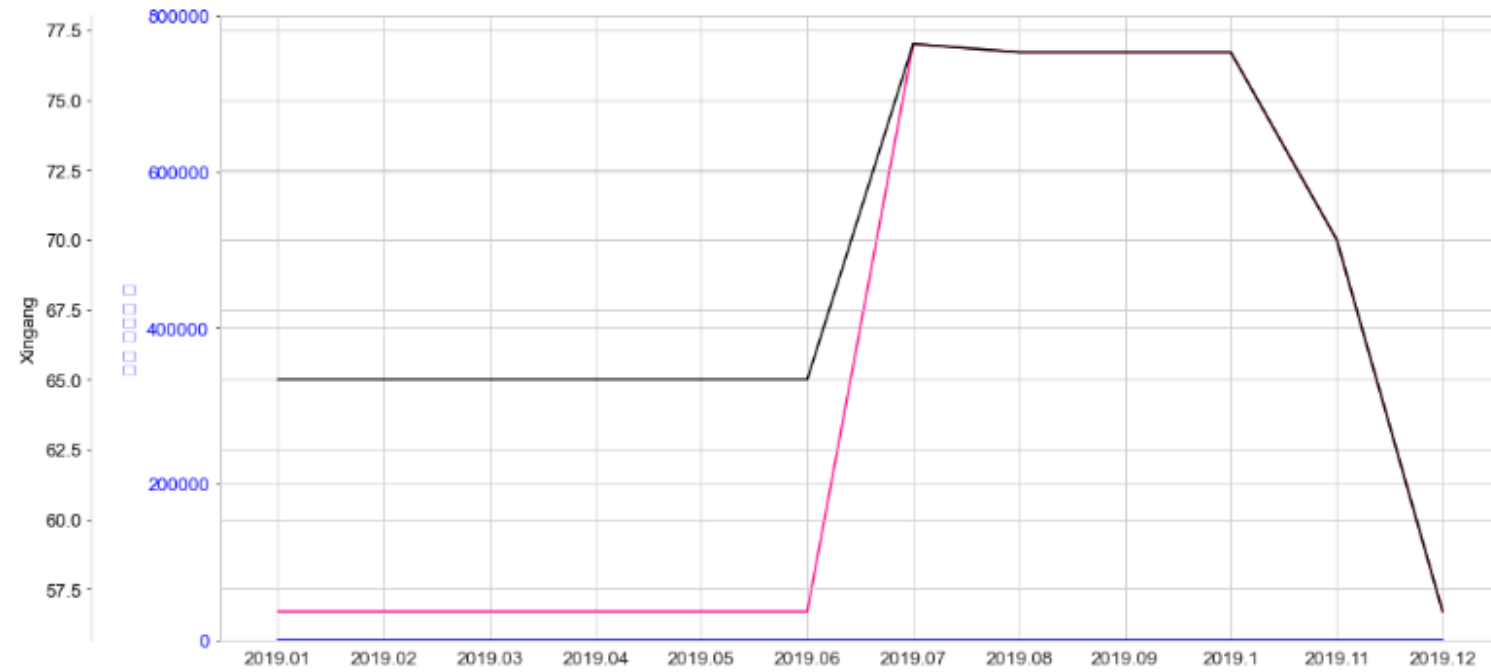
1)해상운임 데이터가 있는 국가에 대한 확진자 수와 해상운임의 시계열 그래프와 상관분석



```
# 피어슨 상관계수
from scipy.stats import *
print(pearsonr(fin_chi['확진자'],fin_chi['Shanghai']))
print(pearsonr(fin_chi['확진자'],fin_chi['Xingang']))
```

(0.3860426963011513, 0.015210646580514827)
(0.48211261526610844, 0.0018841343913174008)

연 단위로 분할



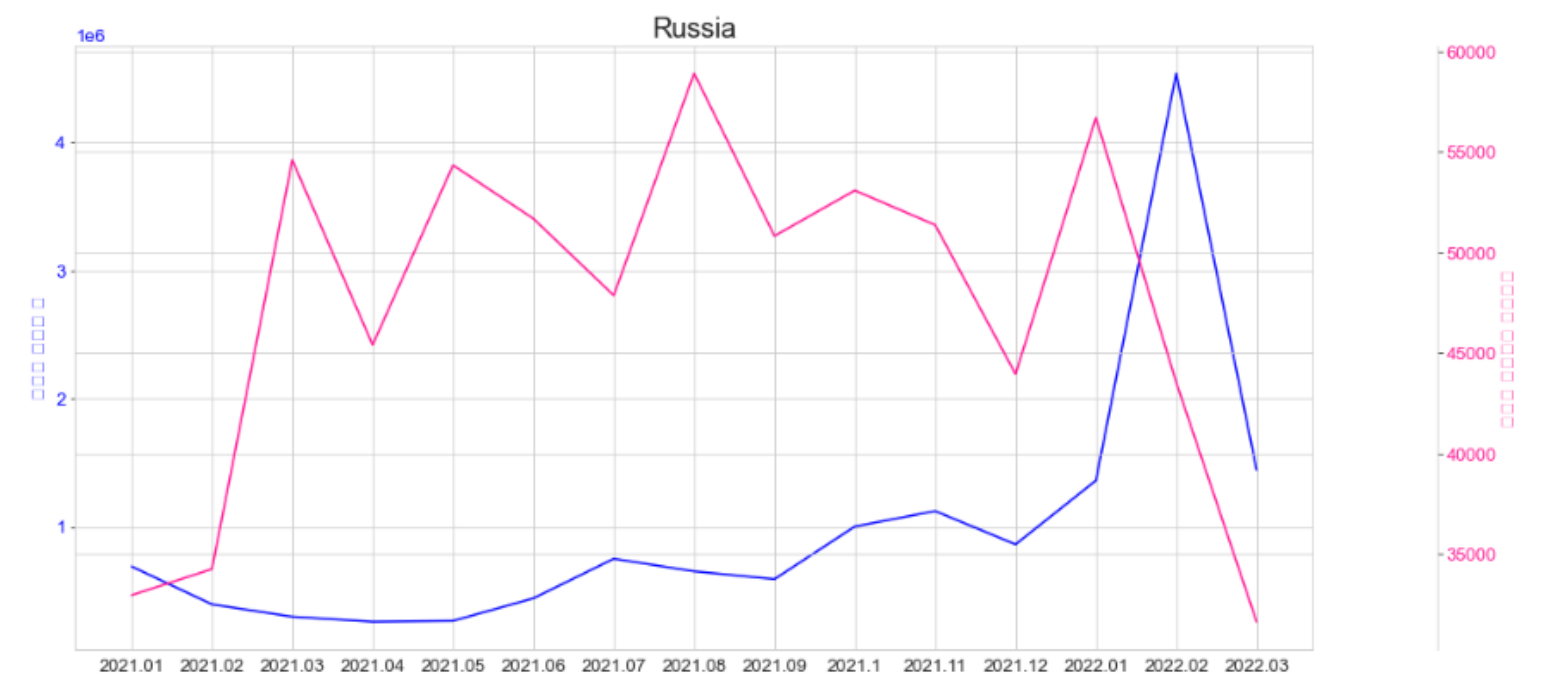
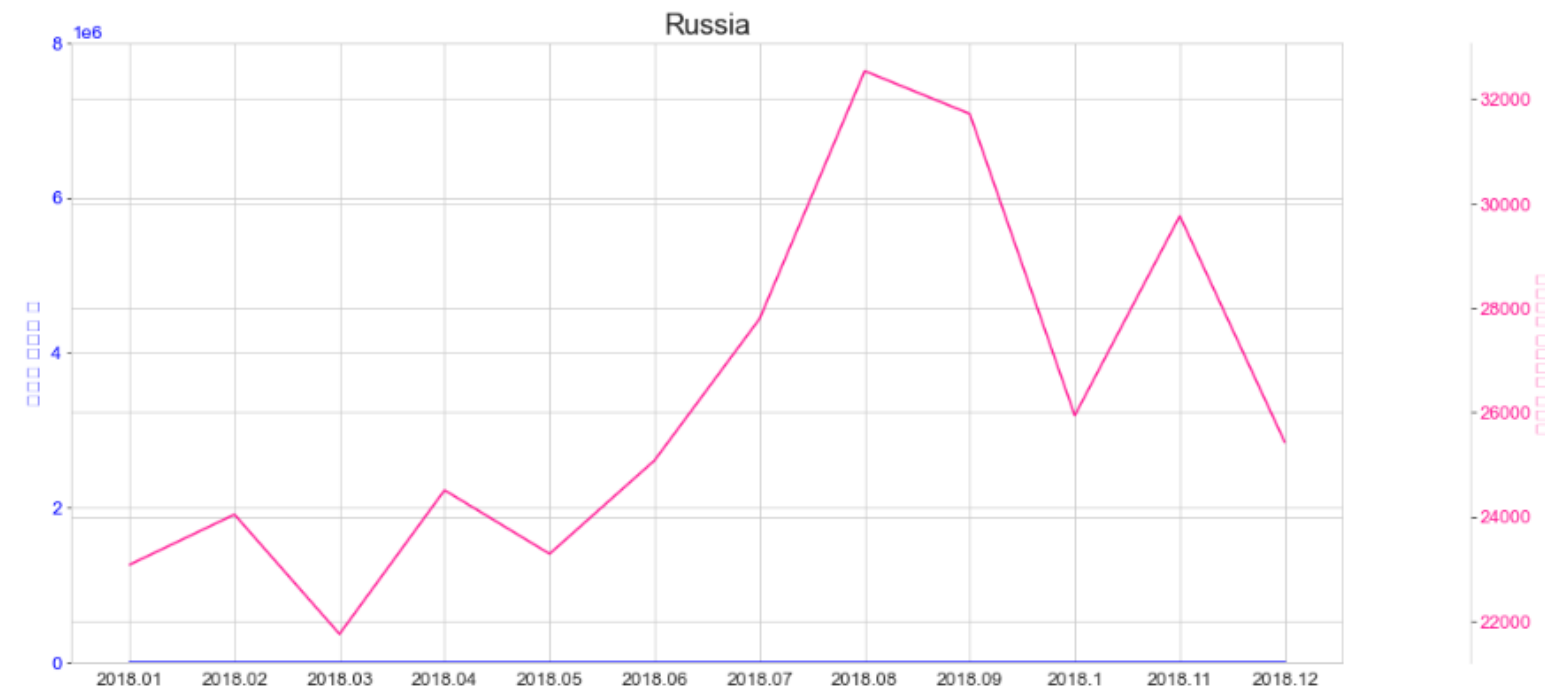
```
# 피어슨 상관계수
from scipy.stats import *
print(pearsonr(fin_chi2020['확진자'],fin_chi2020['Shanghai']))
print(pearsonr(fin_chi2020['확진자'],fin_chi2020['Xingang']))
```

(-0.21541474273681327, 0.5013373478249834)
 (-0.21541474273681327, 0.5013373478249834)

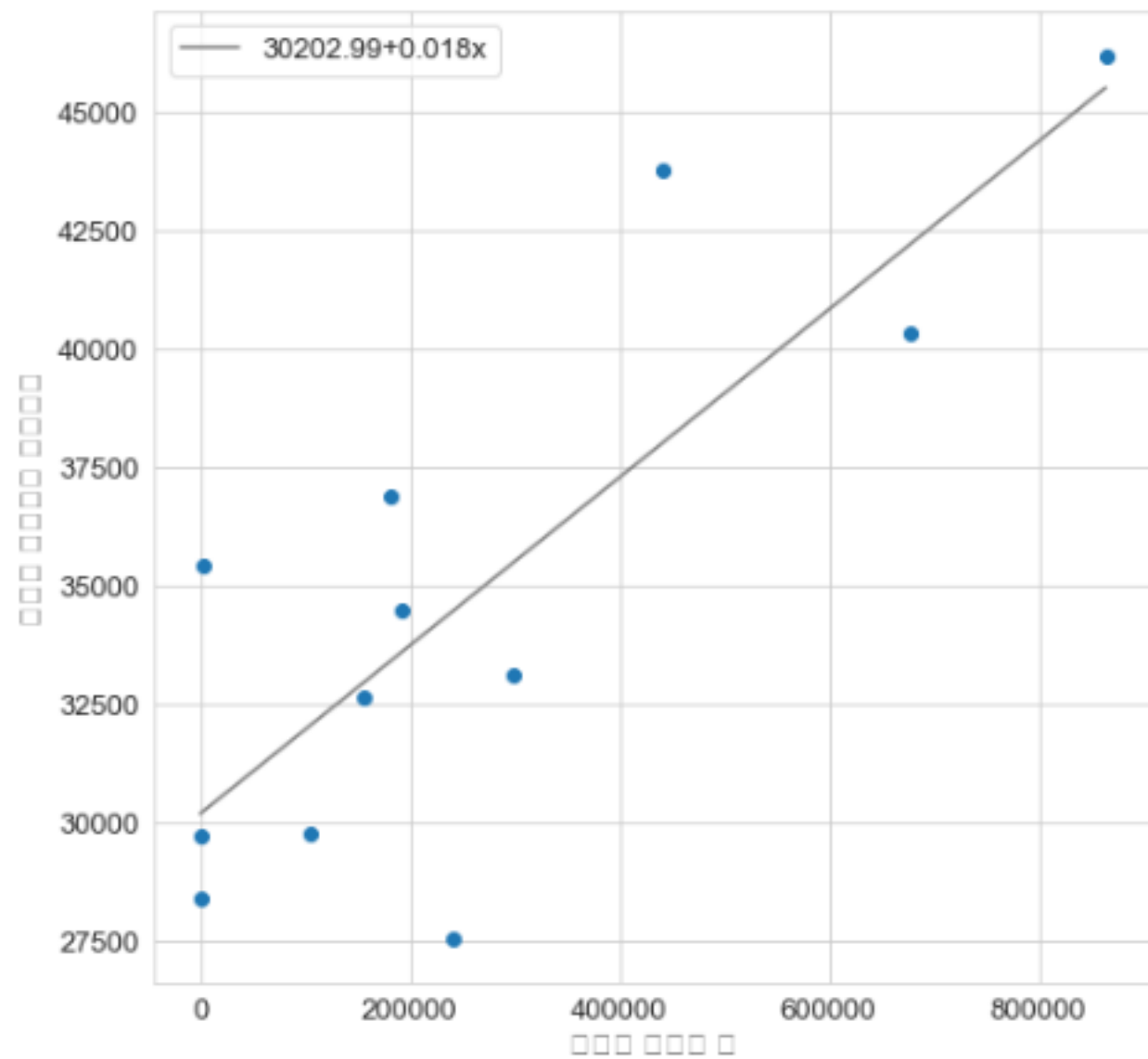
```
# 피어슨 상관계수
from scipy.stats import *
print(pearsonr(fin_chi2021['확진자'],fin_chi2021['Shanghai']))
print(pearsonr(fin_chi2021['확진자'],fin_chi2021['Xingang']))
```

(0.4133411558301705, 0.12566419036066045)
 (0.5455920528732259, 0.03540353251456931)

2) 확진자 수와 컨테이너 수출실적의 연단위 시계열 그래프와 상관분석

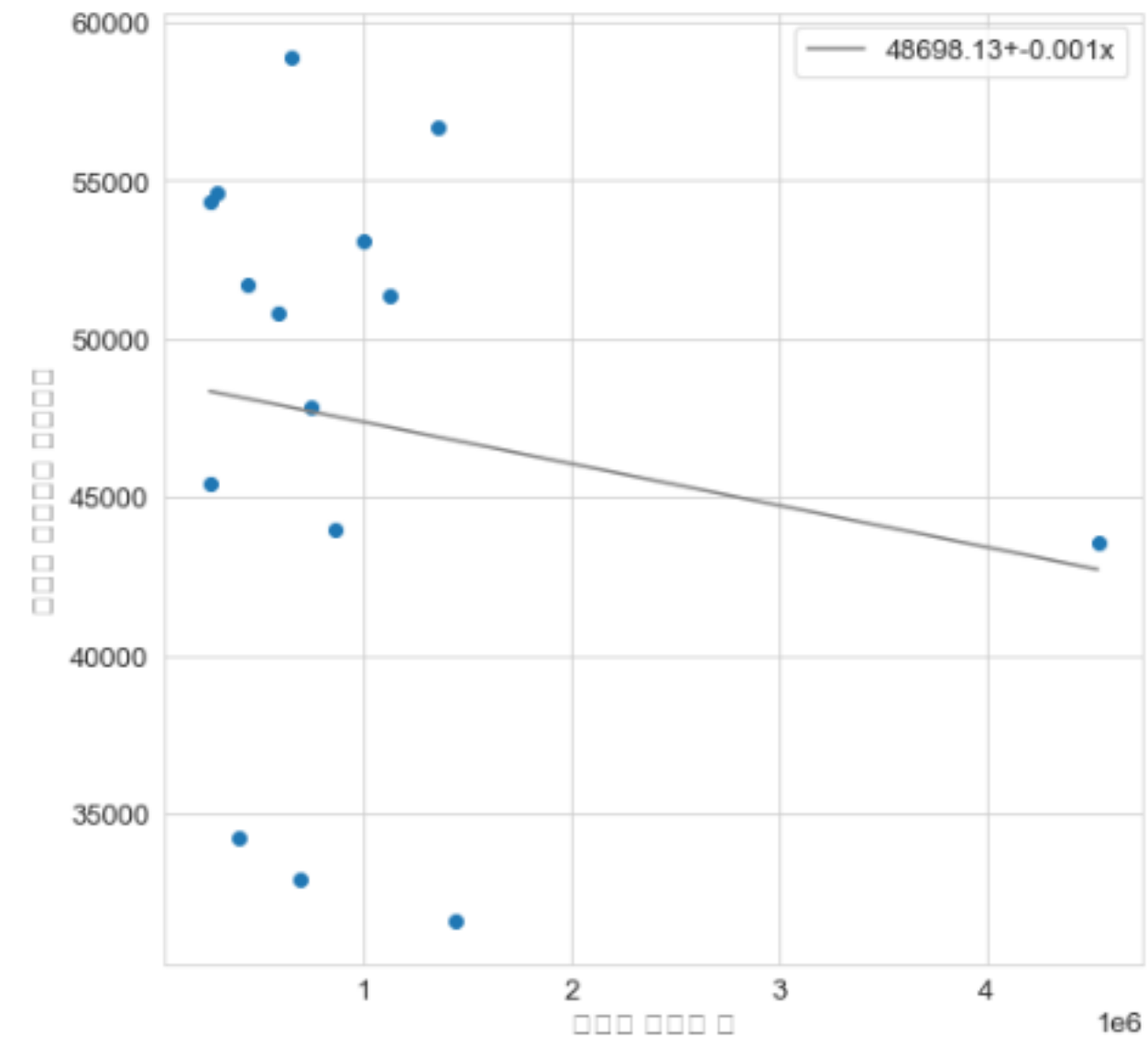


2) 확진자 수와 컨테이너 수출실적의 연단위 시계열 그래프와 상관분석



```
# 피어슨 상관계수  
from scipy.stats import *  
print(pearsonr(fin_cont_rus2020['확진자'], fin_cont_rus2020['러시아 연방']))
```

(0.8060784285103044, 0.0015438379665696006)



```
# 피어슨 상관계수  
from scipy.stats import *  
print(pearsonr(fin_cont_rus2021['확진자'], fin_cont_rus2021['러시아 연방']))
```

(-0.15954809828968208, 0.5700496160766833)

분석결과

- 해상운임 분석에서의 p-value가 0.05보다 작은 나라들의 물동량 분석에서 p-value를 살펴봤으나 대부분 0.05를 훨씬 초과했다.
- 기간을 길게 잡았을 때엔 어느정도 선형 상관성이 보장되는 것 같아 보이지만 연단위로 끊으면 상관관계를 알기 어려워진다.
- 연단위로 끊었을 때 물동량의 코로나 전/후의 시계열 그래프 개형을 살펴보면 전과 후에 차이가 거의 없다.
- 가끔 한 구간씩 확진자 추이와 상관성이 있는 것으로 나오지만 전후 비교를 해보면 다른 요인에 의해 우연히 그렇게 되었다고 판단하는 것이 더 타당해 보인다.