In [1]:
```
import numpy as np
import pandas as pd
import folium
from folium import plugins
import warnings
warnings.filterwarnings(action='ignore')
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
PatientInfo = pd.read_csv('./PatientInfo.csv',parse_dates=["confirmed_date"])
region = pd.read_csv('./Region.csv')
PatientInfo.head()
```

Out[2]:

	patient_id	global_num	sex	birth_year	age	country	province	city	disease	infectic
0	1000000001	2.0	male	1964.0	50s	Korea	Seoul	Gangseo-gu	NaN	o
1	1000000002	5.0	male	1987.0	30s	Korea	Seoul	Jungnang-gu	NaN	o
2	1000000003	6.0	male	1964.0	50s	Korea	Seoul	Jongno-gu	NaN	conta
3	1000000004	7.0	male	1991.0	20s	Korea	Seoul	Mapo-gu	NaN	o
4	1000000005	9.0	female	1992.0	20s	Korea	Seoul	Seongbuk-gu	NaN	conta

In [3]:
```
region = pd.DataFrame(region)
```

In [4]:
```
Output = pd.merge(region, PatientInfo, on=['city','province'])
Output.describe()
```

Out[4]:

	code	latitude	longitude	elementary_school_count	kindergarten_count	univer
count	3253.000000	3253.000000	3253.000000	3253.000000	3253.000000	32
mean	35915.533354	36.628161	127.847826	38.721180	64.296034	
std	21892.690317	0.852660	0.887869	23.596678	43.574031	
min	10010.000000	33.488936	126.297950	4.000000	5.000000	
25%	14080.000000	35.825056	127.032693	22.000000	33.000000	
50%	20310.000000	36.789844	127.487396	31.000000	61.000000	
75%	60010.000000	37.483804	128.741544	48.000000	78.000000	
max	70000.000000	38.207022	129.416575	113.000000	195.000000	

In [5]:
```
subset = pd.DataFrame(Output[['province','city', 'latitude', 'longitude','infection_ca
subset.head()
```

Out[5]:

	province	city	latitude	longitude	infection_case
0	Seoul	Gangnam-gu	37.518421	127.047222	Shincheonji Church
1	Seoul	Gangnam-gu	37.518421	127.047222	etc

	province	city	latitude	longitude	infection_case
2	Seoul	Gangnam-gu	37.518421	127.047222	contact with patient
3	Seoul	Gangnam-gu	37.518421	127.047222	contact with patient
4	Seoul	Gangnam-gu	37.518421	127.047222	etc

In [6]:
```
RgnCnt = Output[['latitude','longitude']].dropna()
RgnCnt.head()
```

Out[6]:

	latitude	longitude
0	37.518421	127.047222
1	37.518421	127.047222
2	37.518421	127.047222
3	37.518421	127.047222
4	37.518421	127.047222

토탈 확진자 지도로 표시

In [7]:
```
ma = folium.Map([36.4, 128], zoom_start=7)
plugins.MousePosition().add_to(ma)
plugins.MarkerCluster(RgnCnt).add_to(ma)
ma
```

Out[7]: Make this Notebook Trusted to load map: File -> Trust Notebook

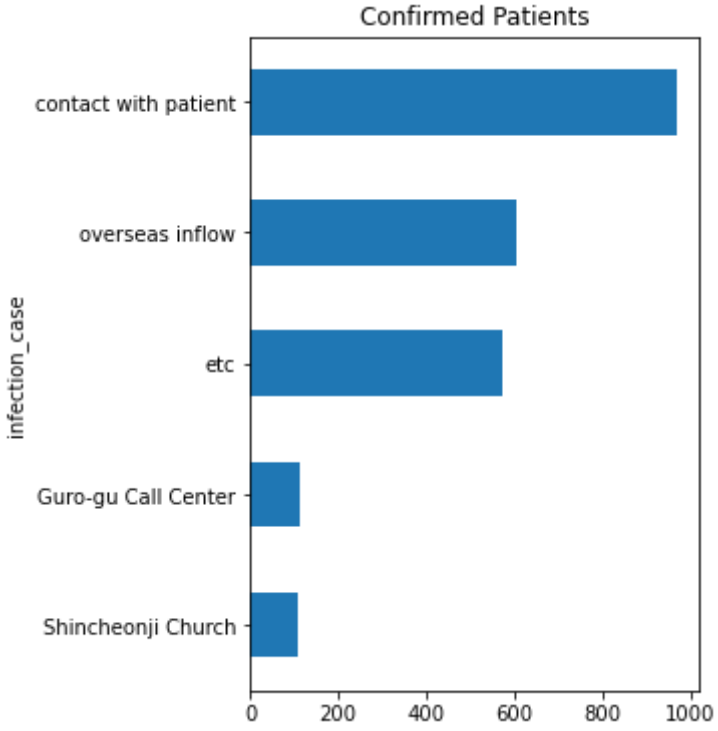
In [ ]:
```

```

감염경로 중 해외 유입이 604건으로 이에대해 조사해본다.

In [8]:
```
ContactCnt=PatientInfo["infection_case"].value_counts()
ContactCntTop=ContactCnt.sort_values().tail(5)
ContactCntTop.plot.barh(figsize=(4,6))
plt.title("Confirmed Patients")
plt.ylabel('infection_case')
```

Out[8]: Text(0, 0.5, 'infection_case')



In [9]:
```
PatientInfo["OverSea"] = PatientInfo["infection_case"]
PatientInfo["OverSea"] = PatientInfo["OverSea"].str.strip()
PatientInfo.head()
```

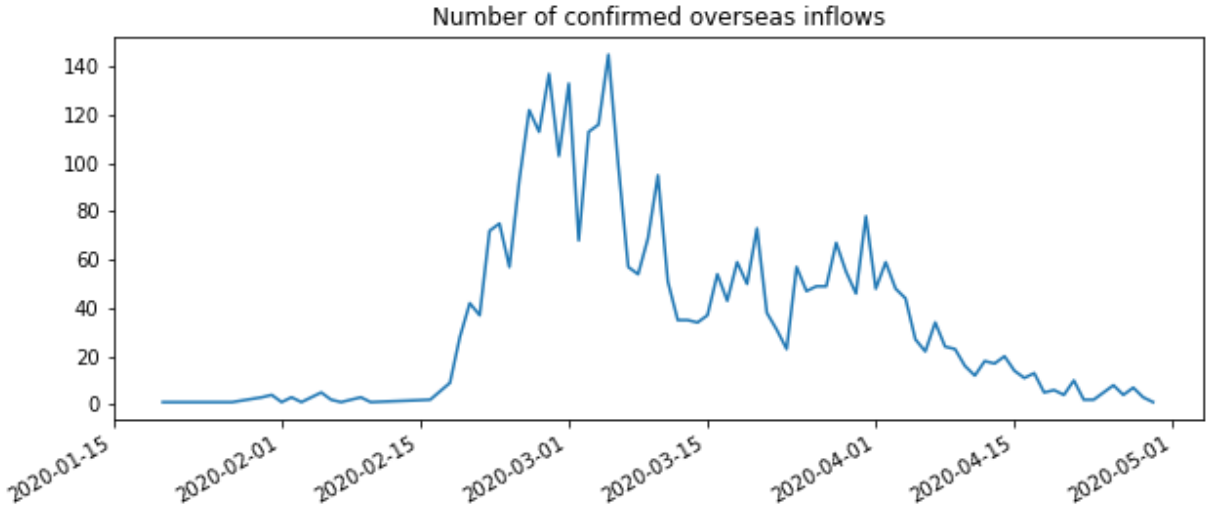
Out[9]:

	patient_id	global_num	sex	birth_year	age	country	province	city	disease	infectio
0	1000000001	2.0	male	1964.0	50s	Korea	Seoul	Gangseo-gu	NaN	o
1	1000000002	5.0	male	1987.0	30s	Korea	Seoul	Jungnang-gu	NaN	o
2	1000000003	6.0	male	1964.0	50s	Korea	Seoul	Jongno-gu	NaN	conta
3	1000000004	7.0	male	1991.0	20s	Korea	Seoul	Mapo-gu	NaN	o
4	1000000005	9.0	female	1992.0	20s	Korea	Seoul	Seongbuk-gu	NaN	conta

In [10]:
```
PatientInfo["Month"]=PatientInfo["confirmed_date"].dt.month
```

In [11]:
```
Overseas = PatientInfo[(PatientInfo["infection_case"]=='overseas inflow')].copy()
PatientInfo["confirmed_date"].value_counts().sort_index().plot(figsize=(10,4))
plt.title("Number of confirmed overseas inflows")
```

Out[11]:
```
Text(0.5, 1.0, 'Number of confirmed overseas inflows')
```



In [12]:
```
OverseasCntRgn = Overseas["province"].value_counts()
OverseasCntRgn.head()
```

Out[12]:
```
Seoul              257
Gyeonggi-do        157
Incheon             38
Busan               29
Gyeongsangnam-do    15
Name: province, dtype: int64
```

In [13]:
```
TotalCntRgn = PatientInfo["province"].value_counts()
```

In [14]:
```
df_Total_Rgn = TotalCntRgn.to_frame()
df_Total_Rgn.columns = ["Total patients"]
df_Total_Rgn.head(20)
```

Out[14]:

	Total patients
Gyeongsangbuk-do	1230
Gyeonggi-do	677
Seoul	633
Chungcheongnam-do	143
Busan	134
Gyeongsangnam-do	116
Incheon	92
Daegu	63
Gangwon-do	51
Sejong	46
Chungcheongbuk-do	45
Ulsan	43
Daejeon	40
Gwangju	30
Jeollabuk-do	17
Jeollanam-do	15
Jeju-do	13

In [15]:
```
Overseas_Rgn = pd.DataFrame({"overseas inflow patients" : OverseasCntRgn})
Overseas_Rgn.head()
```

Out[15]:

	overseas inflow patients
Seoul	257
Gyeonggi-do	157
Incheon	38
Busan	29
Gyeongsangnam-do	15

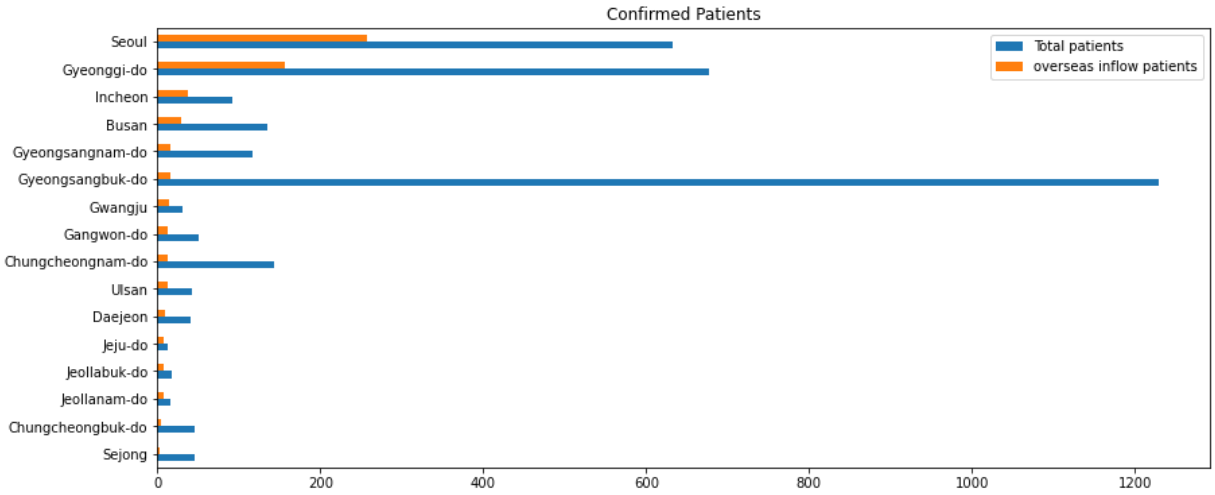
In [16]:
```
dfTotal_Overseas_case = df_Total_Rgn.merge(Overseas_Rgn, left_index=True, right_index
dfTotal_Overseas_case.head(20)
```

Out[16]:

	Total patients	overseas inflow patients
Gyeongsangbuk-do	1230	15
Gyeonggi-do	677	157
Seoul	633	257
Chungcheongnam-do	143	13
Busan	134	29
Gyeongsangnam-do	116	15
Incheon	92	38
Gangwon-do	51	13
Sejong	46	3
Chungcheongbuk-do	45	5
Ulsan	43	12
Daejeon	40	10
Gwangju	30	14
Jeollabuk-do	17	8
Jeollanam-do	15	7
Jeju-do	13	8

In [17]:
```
dfTotal_Overseas_case.sort_values(by="overseas inflow patients").plot.barh(figsize=(1
plt.title("Confirmed Patients")
```

Out[17]:
```
Text(0.5, 1.0, 'Confirmed Patients')
```


Confirmed Patients
Total patients
overseas inflow patients
Seoul
Gyeonggi-do
Incheon
Busan
Gyeongsangnam-do
Gyeongsangbuk-do
Gwangju
Gangwon-do
Chungcheongnam-do
Ulsan
Daejeon
Jeju-do
Jeollabuk-do
Jeollanam-do
Chungcheongbuk-do
Sejong
0
200
400
600
800
1000
1200