

In [1]:

```
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
%matplotlib inline
```

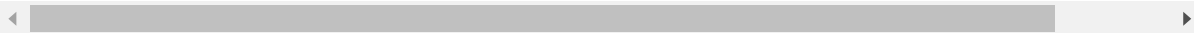
In [3]:

```
data=pd.read_csv("C:/Users/M.sailaja/Downloads/novel-corona-virus-2019-dataset/covid_19_data")
```

Out[3]:

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths
0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0
1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0
2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0
3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0
4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0
...
7921	7922	03/22/2020	NaN	Jersey	2020-03-17T18:33:03	0.0	0.0
7922	7923	03/22/2020	NaN	Puerto Rico	2020-03-22T22:43:02	0.0	1.0
7923	7924	03/22/2020	NaN	Republic of the Congo	2020-03-17T21:33:03	0.0	0.0
7924	7925	03/22/2020	NaN	The Bahamas	2020-03-19T12:13:38	0.0	0.0
7925	7926	03/22/2020	NaN	The Gambia	2020-03-18T14:13:56	0.0	0.0

7926 rows × 8 columns



In [4]:

```
c=pd.read_csv("C:/Users/M.sailaja/Downloads/novel-corona-virus-2019-dataset/time_series_cov
```

Out[4]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20
0	NaN	Thailand	15.0000	101.0000	2	3	5	7	
1	NaN	Japan	36.0000	138.0000	2	1	2	2	
2	NaN	Singapore	1.2833	103.8333	0	1	3	3	
3	NaN	Nepal	28.1667	84.2500	0	0	0	1	
4	NaN	Malaysia	2.5000	112.5000	0	0	0	3	
...
482	NaN	Dominica	15.4150	-61.3710	0	0	0	0	
483	NaN	Grenada	12.1165	-61.6790	0	0	0	0	
484	NaN	Mozambique	-18.6657	35.5296	0	0	0	0	
485	NaN	Syria	34.8021	38.9968	0	0	0	0	
486	NaN	Timor-Leste	-8.8742	125.7275	0	0	0	0	

487 rows × 65 columns



In [5]:

```
d=pd.read_csv("C:/Users/M.sailaja/Downloads/novel-corona-virus-2019-dataset/time_series_cov
```

Out[5]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20
0	NaN	Thailand	15.0000	101.0000	0	0	0	0	
1	NaN	Japan	36.0000	138.0000	0	0	0	0	
2	NaN	Singapore	1.2833	103.8333	0	0	0	0	
3	NaN	Nepal	28.1667	84.2500	0	0	0	0	
4	NaN	Malaysia	2.5000	112.5000	0	0	0	0	
...
482	NaN	Dominica	15.4150	-61.3710	0	0	0	0	
483	NaN	Grenada	12.1165	-61.6790	0	0	0	0	
484	NaN	Mozambique	-18.6657	35.5296	0	0	0	0	
485	NaN	Syria	34.8021	38.9968	0	0	0	0	
486	NaN	Timor-Leste	-8.8742	125.7275	0	0	0	0	

487 rows × 65 columns



In [18]:

```
re=pd.read_csv("C:/Users/M.sailaja/Downloads/novel-corona-virus-2019-dataset/time_series_co
re
```

Out[18]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20
0	NaN	Thailand	15.0000	101.0000	0	0	0	0	0
1	NaN	Japan	36.0000	138.0000	0	0	0	0	0
2	NaN	Singapore	1.2833	103.8333	0	0	0	0	0
3	NaN	Nepal	28.1667	84.2500	0	0	0	0	0
4	NaN	Malaysia	2.5000	112.5000	0	0	0	0	0
...
482	NaN	Dominica	15.4150	-61.3710	0	0	0	0	0
483	NaN	Grenada	12.1165	-61.6790	0	0	0	0	0
484	NaN	Mozambique	-18.6657	35.5296	0	0	0	0	0
485	NaN	Syria	34.8021	38.9968	0	0	0	0	0
486	NaN	Timor-Leste	-8.8742	125.7275	0	0	0	0	0

487 rows × 65 columns

In [7]:

```
c=c.melt(id_vars=['Province/State','Country/Region','Lat','Long'])
c.head()
```

Out[7]:

	Province/State	Country/Region	Lat	Long	variable	value
0	NaN	Thailand	15.0000	101.0000	1/22/20	2
1	NaN	Japan	36.0000	138.0000	1/22/20	2
2	NaN	Singapore	1.2833	103.8333	1/22/20	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0

In [8]:

```
c=c.rename({'variable': 'Date', 'value': 'Confirmed'}, axis='columns')
c.head()
```

Out[8]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed
0	NaN	Thailand	15.0000	101.0000	1/22/20	2
1	NaN	Japan	36.0000	138.0000	1/22/20	2
2	NaN	Singapore	1.2833	103.8333	1/22/20	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0

In [9]:

```
d=d.melt(id_vars=['Province/State', 'Country/Region', 'Lat', 'Long'])
d=d.rename({'variable': 'Date', 'value': 'Death'}, axis='columns')
d.head()
```

Out[9]:

	Province/State	Country/Region	Lat	Long	Date	Death
0	NaN	Thailand	15.0000	101.0000	1/22/20	0
1	NaN	Japan	36.0000	138.0000	1/22/20	0
2	NaN	Singapore	1.2833	103.8333	1/22/20	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0

In [19]:

```
re=re.melt(id_vars=['Province/State', 'Country/Region', 'Lat', 'Long'])
re=re.rename({'variable': 'Date', 'value': 'Recovered'}, axis='columns')
```

In [21]:

```
re.head()
```

Out[21]:

	Province/State	Country/Region	Lat	Long	Date	Recovered
0	NaN	Thailand	15.0000	101.0000	1/22/20	0
1	NaN	Japan	36.0000	138.0000	1/22/20	0
2	NaN	Singapore	1.2833	103.8333	1/22/20	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0

In [22]:

```
combo = [c, d, re]
combo
```

Out[22]:

	Province/State	Country/Region	Lat	Long	Date	Confirme
0	NaN	Thailand	15.0000	101.0000	1/22/20	
2						
1	NaN	Japan	36.0000	138.0000	1/22/20	
2						
2	NaN	Singapore	1.2833	103.8333	1/22/20	
0						
3	NaN	Nepal	28.1667	84.2500	1/22/20	
0						
4	NaN	Malaysia	2.5000	112.5000	1/22/20	
0						
...	
...						
29702	NaN	Dominica	15.4150	-61.3710	3/22/20	
1						
29703	NaN	Grenada	12.1165	-61.6790	3/22/20	
1						
29704	NaN	Mozambique	-18.6657	35.5296	3/22/20	
1						
29705	NaN	Syria	34.8021	38.9968	3/22/20	
1						
29706	NaN	Timor-Leste	-8.8742	125.7275	3/22/20	
1						

[29707 rows x 6 columns],

	Province/State	Country/Region	Lat	Long	Date	Death
0	NaN	Thailand	15.0000	101.0000	1/22/20	0
1	NaN	Japan	36.0000	138.0000	1/22/20	0
2	NaN	Singapore	1.2833	103.8333	1/22/20	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0
...
29702	NaN	Dominica	15.4150	-61.3710	3/22/20	0
29703	NaN	Grenada	12.1165	-61.6790	3/22/20	0
29704	NaN	Mozambique	-18.6657	35.5296	3/22/20	0
29705	NaN	Syria	34.8021	38.9968	3/22/20	0
29706	NaN	Timor-Leste	-8.8742	125.7275	3/22/20	0

[29707 rows x 6 columns],

	Province/State	Country/Region	Lat	Long	Date	Recovere
d						
0	NaN	Thailand	15.0000	101.0000	1/22/20	
0						
1	NaN	Japan	36.0000	138.0000	1/22/20	
0						
2	NaN	Singapore	1.2833	103.8333	1/22/20	
0						
3	NaN	Nepal	28.1667	84.2500	1/22/20	
0						
4	NaN	Malaysia	2.5000	112.5000	1/22/20	
0						
...	

```
...
29702      NaN      Dominica  15.4150  -61.3710  3/22/20
0
29703      NaN      Grenada  12.1165  -61.6790  3/22/20
0
29704      NaN      Mozambique -18.6657   35.5296  3/22/20
0
29705      NaN      Syria    34.8021   38.9968  3/22/20
0
29706      NaN      Timor-Leste -8.8742  125.7275  3/22/20
0

[29707 rows x 6 columns]]
```

In [23]:

```
combo = [c, d, re]
combo = [df.set_index(['Province/State', 'Country/Region', 'Lat', 'Long', 'Date']) for df in co
combo=combo[0].join(combo[1:])
```

In [24]:

```
combo.head()
```

Out[24]:

Province/State	Country/Region	Lat	Long	Date	Confirmed	Death	Recovered
NaN	Thailand	15.0000	101.0000	1/22/20	2	0	0
	Japan	36.0000	138.0000	1/22/20	2	0	0
	Singapore	1.2833	103.8333	1/22/20	0	0	0
	Nepal	28.1667	84.2500	1/22/20	0	0	0
	Malaysia	2.5000	112.5000	1/22/20	0	0	0

In [25]:

```
combo=combo.reset_index()
combo.head()
```

Out[25]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Death	Recovered
0	NaN	Thailand	15.0000	101.0000	1/22/20	2	0	0
1	NaN	Japan	36.0000	138.0000	1/22/20	2	0	0
2	NaN	Singapore	1.2833	103.8333	1/22/20	0	0	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0	0	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0	0	0

In [26]:

```
combo[['Lat', 'Long', 'Confirmed', 'Death', 'Recovered']] = combo[['Lat', 'Long', 'Confirmed', 'Death', 'Recovered']]
```

In [27]:

```
combo.head()
```

Out[27]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Death	Recovered
0	NaN	Thailand	15.0000	101.0000	1/22/20	2	0	0
1	NaN	Japan	36.0000	138.0000	1/22/20	2	0	0
2	NaN	Singapore	1.2833	103.8333	1/22/20	0	0	0
3	NaN	Nepal	28.1667	84.2500	1/22/20	0	0	0
4	NaN	Malaysia	2.5000	112.5000	1/22/20	0	0	0

In [29]:

```
combo[['Date']] = combo[['Date']].apply(pd.to_datetime)
```

In [30]:

```
combo.head()
```

Out[30]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Death	Recovered
0	NaN	Thailand	15.0000	101.0000	2020-01-22	2	0	0
1	NaN	Japan	36.0000	138.0000	2020-01-22	2	0	0
2	NaN	Singapore	1.2833	103.8333	2020-01-22	0	0	0
3	NaN	Nepal	28.1667	84.2500	2020-01-22	0	0	0
4	NaN	Malaysia	2.5000	112.5000	2020-01-22	0	0	0

In [31]:

```
combo.dtypes
```

Out[31]:

```
Province/State    object
Country/Region    object
Lat              float64
Long             float64
Date             datetime64[ns]
Confirmed         int64
Death            int64
Recovered        int64
dtype: object
```

In [32]:

```
combo=combo.rename({'Province/State':'State','Country/Region': 'Country'}, axis='columns')
```

In [33]:

```
combo.head()
```

Out[33]:

	State	Country	Lat	Long	Date	Confirmed	Death	Recovered
0	NaN	Thailand	15.0000	101.0000	2020-01-22	2	0	0
1	NaN	Japan	36.0000	138.0000	2020-01-22	2	0	0
2	NaN	Singapore	1.2833	103.8333	2020-01-22	0	0	0
3	NaN	Nepal	28.1667	84.2500	2020-01-22	0	0	0
4	NaN	Malaysia	2.5000	112.5000	2020-01-22	0	0	0

In [34]:

```
combo.query("Country=='US' & State=='Washington'")
```

Out[34]:

	State	Country	Lat	Long	Date	Confirmed	Death	Recovered
98	Washington	US	47.4009	-121.4905	2020-01-22	0	0	0
585	Washington	US	47.4009	-121.4905	2020-01-23	0	0	0
1072	Washington	US	47.4009	-121.4905	2020-01-24	0	0	0
1559	Washington	US	47.4009	-121.4905	2020-01-25	0	0	0
2046	Washington	US	47.4009	-121.4905	2020-01-26	0	0	0
...
27370	Washington	US	47.4009	-121.4905	2020-03-18	1014	55	0
27857	Washington	US	47.4009	-121.4905	2020-03-19	1376	74	0
28344	Washington	US	47.4009	-121.4905	2020-03-20	1524	83	0
28831	Washington	US	47.4009	-121.4905	2020-03-21	1793	94	0
29318	Washington	US	47.4009	-121.4905	2020-03-22	1996	95	0

61 rows × 8 columns

In [36]:

```
combo.query("Country=='China' & State=='Hubei'")
```

Out[36]:

	State	Country	Lat	Long	Date	Confirmed	Death	Recovered
154	Hubei	China	30.9756	112.2707	2020-01-22	444	17	28
641	Hubei	China	30.9756	112.2707	2020-01-23	444	17	28
1128	Hubei	China	30.9756	112.2707	2020-01-24	549	24	31
1615	Hubei	China	30.9756	112.2707	2020-01-25	761	40	32
2102	Hubei	China	30.9756	112.2707	2020-01-26	1058	52	42
...
27426	Hubei	China	30.9756	112.2707	2020-03-18	67800	3122	56927
27913	Hubei	China	30.9756	112.2707	2020-03-19	67800	3130	57682
28400	Hubei	China	30.9756	112.2707	2020-03-20	67800	3133	58382
28887	Hubei	China	30.9756	112.2707	2020-03-21	67800	3139	58946
29374	Hubei	China	30.9756	112.2707	2020-03-22	67800	3144	59433

61 rows × 8 columns

In [37]:

```
combo['Active']=combo['Confirmed'] - combo['Death'] - combo['Recovered']
```

In [38]:

```
combo.query("Country=='China' & State=='Hubei'")
```

Out[38]:

	State	Country	Lat	Long	Date	Confirmed	Death	Recovered	Active
154	Hubei	China	30.9756	112.2707	2020-01-22	444	17	28	399
641	Hubei	China	30.9756	112.2707	2020-01-23	444	17	28	399
1128	Hubei	China	30.9756	112.2707	2020-01-24	549	24	31	494
1615	Hubei	China	30.9756	112.2707	2020-01-25	761	40	32	689
2102	Hubei	China	30.9756	112.2707	2020-01-26	1058	52	42	964
...
27426	Hubei	China	30.9756	112.2707	2020-03-18	67800	3122	56927	7751
27913	Hubei	China	30.9756	112.2707	2020-03-19	67800	3130	57682	6988
28400	Hubei	China	30.9756	112.2707	2020-03-20	67800	3133	58382	6285
28887	Hubei	China	30.9756	112.2707	2020-03-21	67800	3139	58946	5715
29374	Hubei	China	30.9756	112.2707	2020-03-22	67800	3144	59433	5223

61 rows × 9 columns

In [39]:

```
maxi=combo[combo.Date==combo['Date'].max()]
```

In [40]:

```
maxi[['Country','Active']].sort_values('Active',ascending=False).head(10)
```

Out[40]:

	Country	Active
29236	Italy	46638
29231	Germany	24513
29238	Spain	24421
29319	US	15676
29377	France	13144
29375	Iran	12022
29251	Switzerland	7016
29376	Korea, South	5884
29623	United Kingdom	5337
29374	China	5223

In [41]:

```
maxi[['Country', 'Confirmed']].groupby(['Country']).sum().sort_values('Confirmed', ascending=
```

Out[41]:

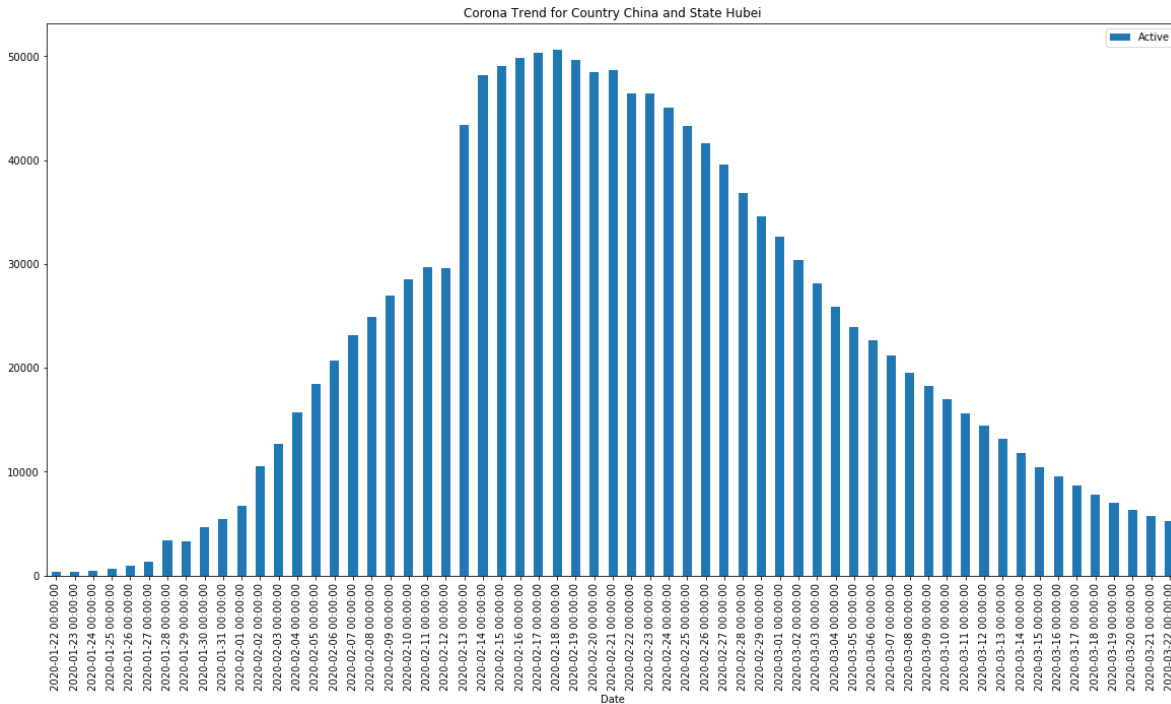
	Confirmed
Country	
China	81397
Italy	59138
US	33272
Spain	28768
Germany	24873
Iran	21638
France	16176
Korea, South	8897
Switzerland	7245
United Kingdom	5741

In [43]:

```
%matplotlib inline
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
combo[['Date', 'Country', 'State', 'Active']].query("Country=='China' & State=='Hubei']").plot()
```

Out[43]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a3f98d7bb0>

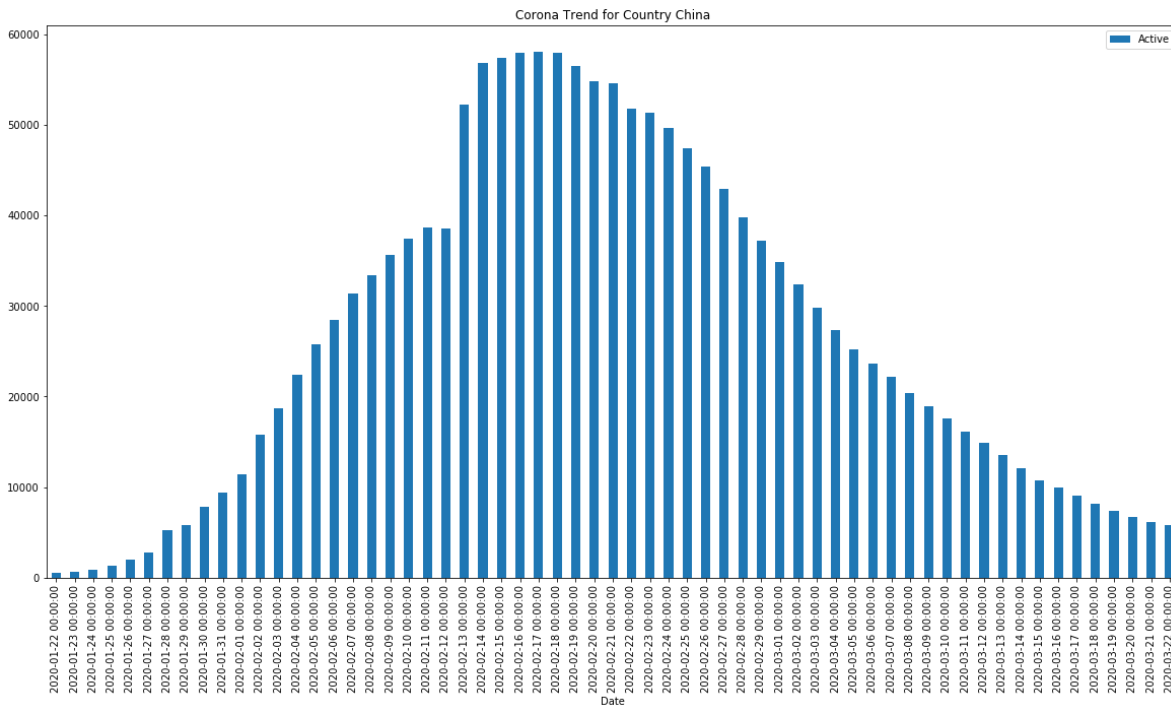


In [44]:

```
%matplotlib inline
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
combo[['Date', 'Country', 'State', 'Active']].groupby(['Date', 'Country']).sum().query("Country
```

Out[44]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a3fb29e970>

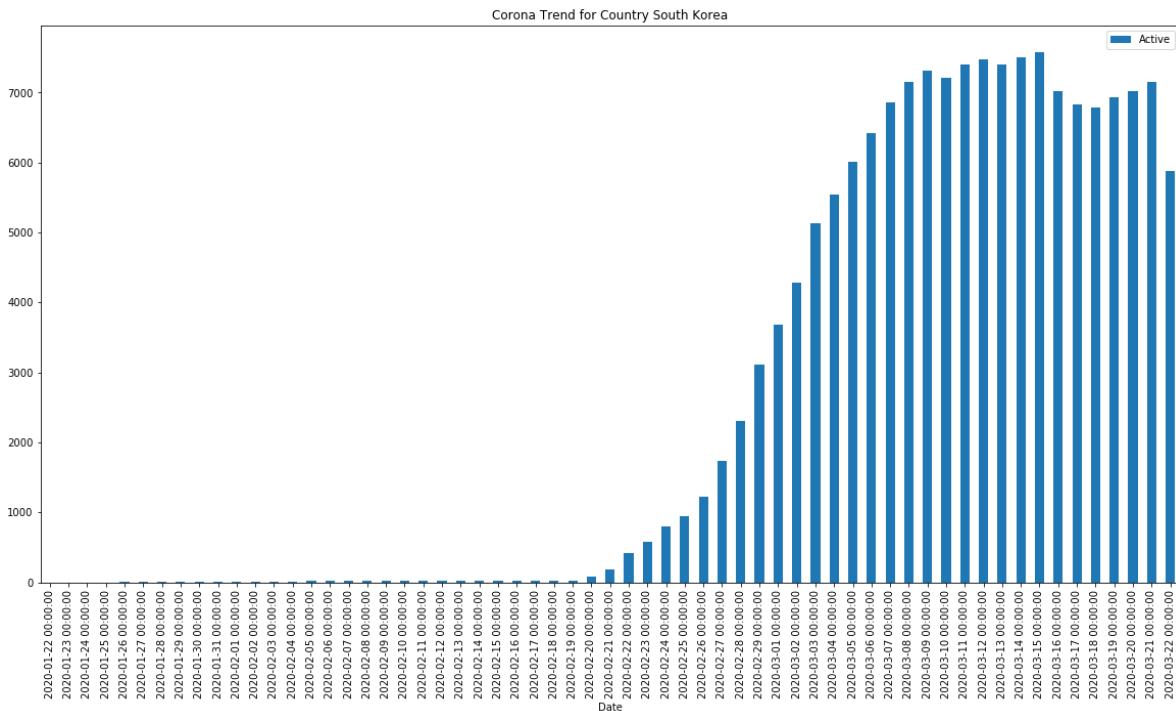


In [45]:

```
%matplotlib inline
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
combo[['Date', 'Country', 'State', 'Active']].query("Country=='Korea, South']").plot(x='Date', y
```

Out[45]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a3f9837a30>

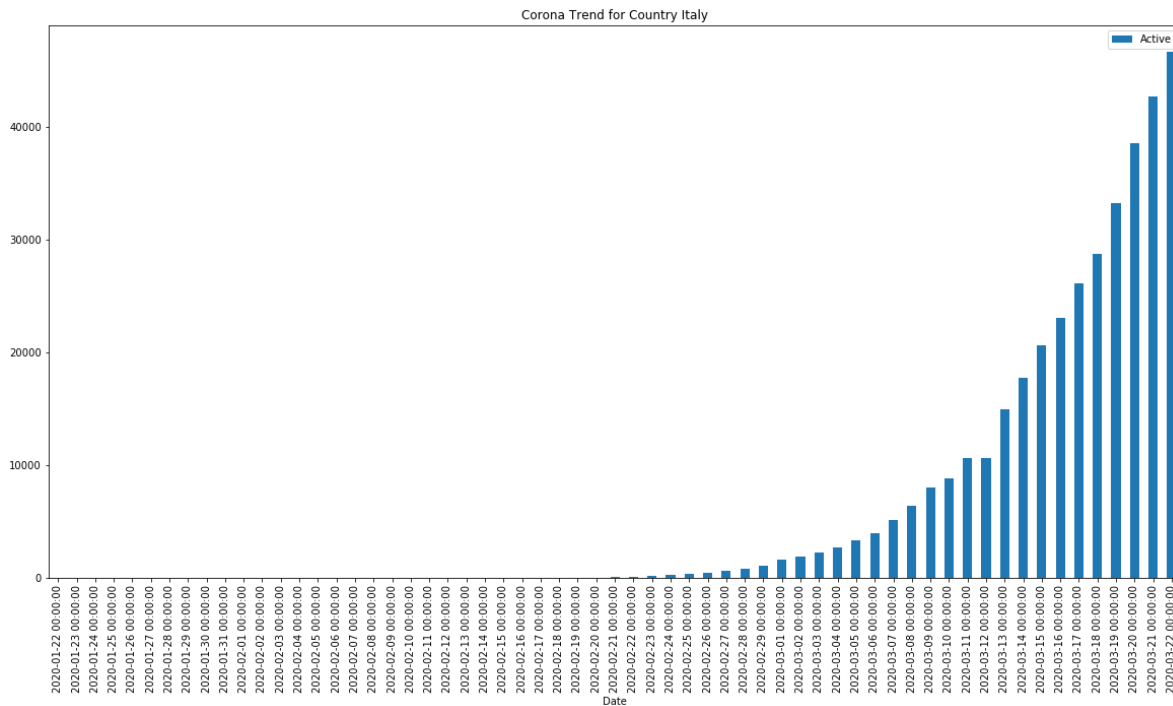


In [46]:

```
%matplotlib inline
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
combo[['Date', 'Country', 'State', 'Active']].query("Country=='Italy']").plot(x='Date', y='Active')
```

Out[46]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a3f98409d0>

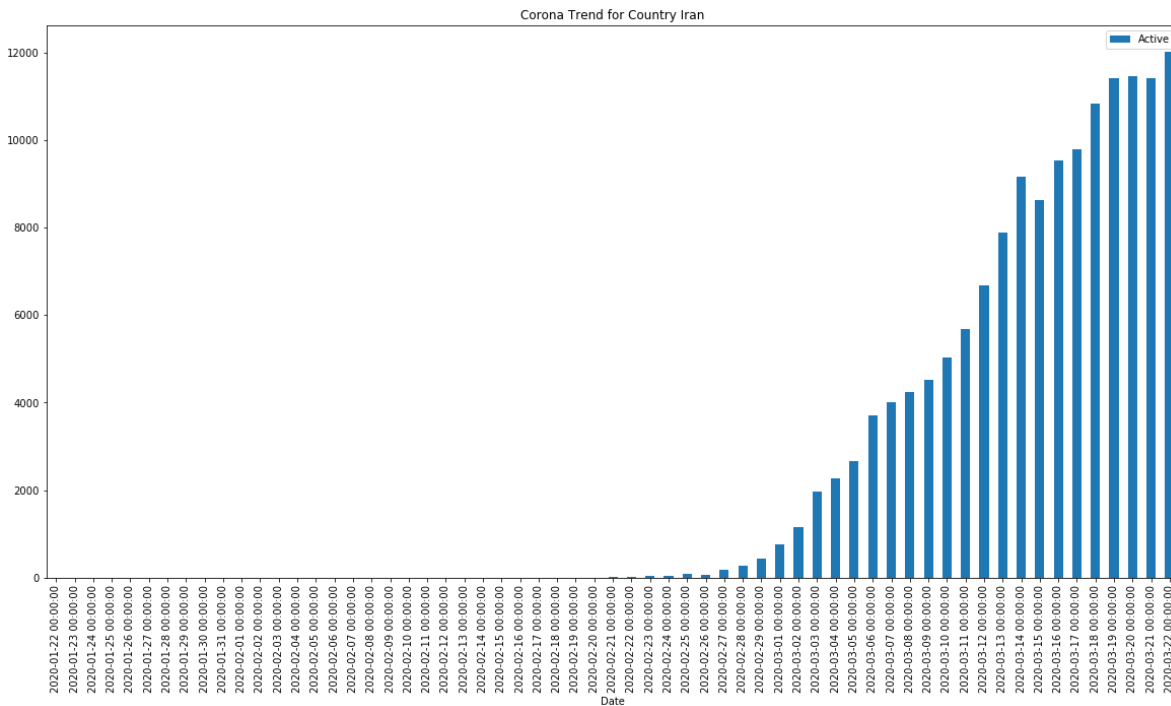


In [47]:

```
%matplotlib inline
mpl.rcParams['figure.figsize'] = (20, 10)
mpl.rcParams['axes.grid'] = False
combo[['Date', 'Country', 'State', 'Active']].query("Country=='Iran']").plot(x='Date', y='Active')
```

Out[47]:

<matplotlib.axes._subplots.AxesSubplot at 0x1a3f9d96790>



In []: